ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations

Ph. Grohs and F. Hornung and A. Jentzen and Ph. von Wurstemberger

Research Report No. 2018-32
September 2018

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations

Philipp Grohs[1], Fabian Hornung[2,3], Arnulf Jentzen[4], and Philippe von Wurstemberger[5]

[1]Faculty of Mathematics and Research Platform Data Science, University of Vienna, Austria, e-mail: philipp.grohs@univie.ac.at

[2]Institute for Analysis, Karlsruhe Institute of Technology, Germany

[3]Seminar for Applied Mathematics, ETH Zurich, Switzerland, e-mail: fabian.hornung@sam.math.ethz.ch

[4]Seminar for Applied Mathematics, ETH Zurich, Switzerland, e-mail: arnulf.jentzen@sam.math.ethz.ch

[5]Seminar for Applied Mathematics, ETH Zurich, Switzerland, e-mail: wurstep@student.ethz.ch

September 11, 2018

Abstract

Artificial neural networks (ANNs) have very successfully been used in numerical simulations for a series of computational problems ranging from image classification/image recognition, speech recognition, time series analysis, game intelligence, and computational advertising to numerical approximations of partial differential equations (PDEs). Such numerical simulations suggest that ANNs have the capacity to very efficiently approximate high-dimensional functions and, especially, such numerical simulations indicate that ANNs seem to admit the fundamental power to overcome the curse of dimensionality when approximating the high-dimensional functions appearing in the above named computational problems. There are also a series of rigorous mathematical approximation results for ANNs in the scientific literature. Some of these mathematical results prove convergence without convergence rates and some of these mathematical results even rigorously establish convergence rates but there are only a few special cases where mathematical results can rigorously explain the empirical success of ANNs when approximating high-dimensional functions. The key contribution of this article is to disclose that ANNs can efficiently approximate high-dimensional functions in the case of numerical approximations of Black-Scholes PDEs. More precisely, this work reveals that the number of required parameters of an ANN to approximate the solution of the Black-Scholes PDE grows at most polynomially in both the reciprocal of the prescribed approximation accuracy $\varepsilon > 0$ and the PDE dimension $d \in \mathbb{N}$ and we thereby prove, for the first time, that ANNs do indeed overcome the curse of dimensionality in the numerical approximation of Black-Scholes PDEs.

Contents

1 Introduction

Artificial neural networks (ANNs) (cf., e.g., Goodfellow et al. [29], McCulloch & Pitts [50], Priddy & Keller [63], Schmidhuber [66]) have very successfully been used in numerical simulations for a series of computational problems ranging from image classification/image recognition (cf., e.g., Huang et al. [40], Krizhevsky et al. [47], Simonyan & Zisserman [71]), speech recognition (cf., e.g., Dahl et al. [20], Hinton et al. [35], Graves et al. [30], Wu et al. [73]), time series analysis (cf., e.g., Goodfellow et al. [29], LeCun et al. [48]), game intelligence (cf., e.g., Silver et al. [69, 70]), and computational advertising to numerical approximations of partial differential equations (PDEs) (cf., e.g., [4, 5, 6, 22, 23, 26, 32, 34, 45, 54, 56, 64, 72]). Such numerical simulations suggest that ANNs have the capacity to very efficiently approximate high-dimensional functions. Particularly, such numerical simulations indicate that ANNs seem to admit the fundamental power to resolve the curse of dimensionality (cf., e.g., Bellman [7]) in the sense that the number of parameters of an ANN to approximate the high-dimensional functions appearing in the above named computational problems grows at most polynomially in both the reciprocal of the prescibed accuracy $\varepsilon > 0$ and the dimension $d \in \mathbb{N}$. There are also a series of rigorous mathematical approximation results for ANNs in the scientific literature (cf., e.g., [1, 2, 3, 8, 9, 10, 11, 12, 13, 18, 21, 24, 25, 27, 33, 36, 37, 38, 39, 49, 51, 52, 53, 57, 58, 59, 60, 61, 62, 67, 68, 72, 74, 75] and the references mentioned therein). Some of these mathematical results prove convergence without convergence rates and some of these mathematical results even rigorously establish convergence rates but there are only a few special cases where mathematical results can rigorously explain the empirical success of ANNs when approximating high-dimensional functions.

The key contribution of this article is to disclose that ANNs can efficiently approximate high-dimensional functions in the case of numerical approximations of Black-Scholes PDEs. More accurately, Theorem 3.14 below, which is the main result of this paper, reveals that the number of required parameters of an ANN to approximate the solution of the Black-Scholes PDE grows at most polynomially in both the reciprocal of the prescribed approximation accuracy $\varepsilon > 0$ and the PDE dimension $d \in \mathbb{N}$ and we thereby prove, for the first time, that ANNs do indeed resolve the curse of dimensionality in the numerical approximation of Black-Scholes PDEs. To illustrate the main result of this article (Theorem 3.14 in Subsection 3.6 below), we now present in the following theorem a special case of Theorem 3.14 below.

Theorem 1.1. *Let $T, \mathfrak{c} \in (0, \infty)$, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0, \infty)$ be the d-dimensional Euclidean norm and let $|||\cdot|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \colon \mathbb{R}^{d\times d} \to [0, \infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times d}$, let $\mathbf{A}_d \in C(\mathbb{R}^d, \mathbb{R}^d)$, $d \in \mathbb{N}$, and $\mathbf{a} \in C(\mathbb{R}, \mathbb{R})$ be functions which satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $\mathbf{A}_d(x) = (\mathbf{a}(x_1), \mathbf{a}(x_2), \ldots, \mathbf{a}(x_d))$, let $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, be continuous functions, let $\mu_d \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, and $\sigma_d \colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, be functions which satisfy for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that*

$$\mu_d(\lambda x + y) + \lambda\mu_d(0) = \lambda\mu_d(x) + \mu_d(y), \tag{1}$$

$$\sigma_d(\lambda x + y) + \lambda\sigma_d(0) = \lambda\sigma_d(x) + \sigma_d(y), \tag{2}$$

and $\|\mu_d(x)\|_{\mathbb{R}^d} + |||\sigma_d(x)|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \leq \mathfrak{c}(1 + \|x\|_{\mathbb{R}^d})$, let

$$\mathcal{N} = \cup_{\mathcal{L}\in\{2,3,\ldots\}} \cup_{(l_0,l_1,\ldots,l_\mathcal{L})\in((\mathbb{N}^\mathcal{L})\times\{1\})} \left(\times_{k=1}^{\mathcal{L}}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\right), \tag{3}$$

let $\mathcal{P} \colon \mathcal{N} \to \mathbb{N}$ and $\mathcal{R} \colon \mathcal{N} \to \cup_{d=1}^{\infty} C(\mathbb{R}^d, \mathbb{R})$ be the functions which satisfy for all $\mathcal{L} \in \{2, 3, \ldots\}$, $(l_0, l_1, \ldots, l_\mathcal{L}) \in ((\mathbb{N}^\mathcal{L}) \times \{1\})$, $\Phi = ((W_1, B_1), \ldots, (W_\mathcal{L}, B_\mathcal{L})) = ((W_k^{(i,j)})_{i\in\{1,2,\ldots,l_k\},j\in\{1,2,\ldots,l_{k-1}\}}, (B_k^{(i)})_{i\in\{1,2,\ldots,l_k\}})_{k\in\{1,2,\ldots,\mathcal{L}\}} \in (\times_{k=1}^{\mathcal{L}} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $x_0 \in \mathbb{R}^{l_0}, x_1 \in \mathbb{R}^{l_1}, \ldots, x_{\mathcal{L}-1} \in \mathbb{R}^{l_{\mathcal{L}-1}}$ with $\forall\, k \in \mathbb{N} \cap (0, \mathcal{L}) \colon x_k = \mathbf{A}_{l_k}(W_k x_{k-1} + B_k)$ that $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R})$, $(\mathcal{R}(\Phi))(x_0) = W_\mathcal{L} x_{\mathcal{L}-1} + B_\mathcal{L}$, and $\mathcal{P}(\Phi) = \sum_{k=1}^{\mathcal{L}} l_k(l_{k-1} + 1)$, and let $(\phi_{d,\delta})_{d\in\mathbb{N},\, \delta\in(0,1]} \subseteq \mathcal{N}$ satisfy for all $d \in \mathbb{N}$, $\delta \in (0, 1]$, $x \in \mathbb{R}^d$ that $\mathcal{P}(\phi_{d,\delta}) \leq \mathfrak{c}\, d^\mathfrak{c} \delta^{-\mathfrak{c}}$, $\mathcal{R}(\phi_{d,\delta}) \in C(\mathbb{R}^d, \mathbb{R})$, $|(\mathcal{R}(\phi_{d,\delta}))(x)| \leq \mathfrak{c}\, d^\mathfrak{c}(1 + \|x\|_{\mathbb{R}^d}^\mathfrak{c})$, and

$$|\varphi_d(x) - (\mathcal{R}(\phi_{d,\delta}))(x)| \leq \mathfrak{c}\, d^\mathfrak{c}\, \delta\, (1 + \|x\|_{\mathbb{R}^d}^\mathfrak{c}). \tag{4}$$

Then

(i) *there exist unique continuous functions* $u_d\colon [0,T]\times\mathbb{R}^d\to\mathbb{R}$*,* $d\in\mathbb{N}$*, which satisfy for all* $d\in\mathbb{N}$*,* $x\in\mathbb{R}^d$ *that* $u_d(0,x)=\varphi_d(x)$*, which satisfy for all* $d\in\mathbb{N}$ *that* $\inf_{q\in(0,\infty)}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\frac{|u_d(t,x)|}{1+\|x\|^q_{\mathbb{R}^d}}<\infty$*, and which satisfy for all* $d\in\mathbb{N}$ *that* $u_d|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$\begin{aligned}(\tfrac{\partial}{\partial t}u_d)(t,x) &= \operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t,x)\big)\\ &\quad + (\tfrac{\partial}{\partial x}u_d)(t,x)\,\mu_d(x)\end{aligned} \tag{5}$$

for $(t,x)\in(0,T)\times\mathbb{R}^d$ *and*

(ii) *for every* $p\in(0,\infty)$ *there exist* $\mathfrak{C}\in(0,\infty)$*,* $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\,\varepsilon\in(0,1]}\subseteq\mathcal{N}$ *such that for all* $d\in\mathbb{N}$*,* $\varepsilon\in(0,1]$ *it holds that* $\mathcal{P}(\psi_{d,\varepsilon})\leq\mathfrak{C}\,d^{\mathfrak{C}}\,\varepsilon^{-\mathfrak{C}}$*,* $\mathcal{R}(\psi_{d,\varepsilon})\in C(\mathbb{R}^d,\mathbb{R})$*, and*

$$\left[\int_{[0,1]^d}|u_d(T,x)-(\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p\,dx\right]^{1/p}\leq\varepsilon. \tag{6}$$

Theorem 1.1 is an immediate consequence of Corollary 3.16 in Subsection 3.6 below (with $r=1$ in the notation of Corollary 3.16). Corollary 3.16, in turn, follows from Theorem 3.14 (see Subsection 3.6 below). Note that in Theorem 1.1 above the set $\mathcal{N}$ in (3) corresponds to the set of all fully-connected artifical neural networks (with $\mathcal{L}+1$ layers, $\mathcal{L}-1$ hidden layers, l_0 neurons on the input layer, l_1 neurons on the first hidden layer, l_2 neurons on the second hidden layer, ..., $l_{\mathcal{L}-1}$ neurons on the $(\mathcal{L}-1)$-th hidden layer, and $l_{\mathcal{L}}=1$ neurons on the output layer). Moreover, observe that the function $\mathbf{a}$ in Theorem 1.1 is the activation function which is used in the employed artifical neural networks in Theorem 1.1 and observe that the functions $\mathbf{A}_d$, $d\in\mathbb{N}$, are the multidimensional versions associated to the activation function $\mathbf{a}$. In addition, note that for every artifical neural network $\Phi\in\mathcal{N}$ in Theorem 1.1 it holds that $\mathcal{P}(\Phi)$ is the number of parameters used in the artifical neural network Φ and note that for every artifical neural network $\Phi\in\mathcal{N}$ in Theorem 1.1 it holds that $\mathcal{R}(\Phi)$ is the mathematical function associated to the artifical neural network Φ (the realization associated to the artifical neural network Φ). In Section 4 below we apply Theorem 1.1 above and Theorem 3.14 below, respectively, to the Black-Scholes derivative pricing PDE with different payoff functions. More specifically, in Subsection 4.3 we apply Theorem 3.14 in the case of basket call options (cf. Proposition 4.7), in Subsection 4.4 we apply Theorem 3.14 in the case of basket put options (cf. Proposition 4.9), in Subsection 4.5 we apply Theorem 3.14 in the case of call on max options (cf. Proposition 4.13), and in Subsection 4.6 we apply Theorem 3.14 in the case of call on min options (cf. Proposition 4.13). Our proofs of Theorem 1.1 and Theorem 3.14, respectively, are based on probabilistic arguments. More formally, our proofs of Theorem 1.1 and Theorem 3.14, respectively, employ – besides other arguments – the Feynman-Kac formula for viscosity solutions of Kolmogorov PDEs (cf. Proposition 2.22 in Subsection 2.5 below, (217) in the proof of Proposition 3.4 in Subsection 3.2 below, and, e.g., Hairer et al. [31, Corollary 4.17]), Monte-Carlo approximations for the expected value in the Feynman-Kac formula (cf. Corollary 2.5 in Subsection 2.1 below and (226) in the proof of Proposition 3.4 in Subsection 3.2 below), the fact that the solution of the associated stochastic differential equation (SDE) depends affine linearly on the initial value (cf. Proposition 2.20 in Subsection 2.4 below and (213) in the proof of Proposition 3.4 in Subsection 3.2 below) since the considered SDE is affine linear, and an argument to prove the existence of a random realization with the desired approximation properties (cf. Proposition 3.3 in Subsection 3.1 below and (230) in the proof of Proposition 3.4 in Subsection 3.2 below) on the artifical probability space which we employ in our proof of Theorem 1.1 and Theorem 3.14, respectively.

The remainder of this article is organized as follows. In Section 2 we supply several auxiliary results on Monte Carlo approximations (Subsection 2.1), affine functions (Subsection 2.2), SDEs (Subsections 2.3–2.4), and viscosity solutions for PDEs (Subsection 2.5). These auxiliary results are then used in Section 3 to establish the approximation result for ANNs illustrated in Theorem 1.1 above. In particular, we prove in Theorem 3.14 in Section 3 the main result approximation result of this article. In Section 4 we illustrate the application of Theorem 3.14 in the case of the Black-Scholes PDE with different payoff functions.

2 Probabilistic and analytic preliminaries

In this section we provide several basic and in parts well-known auxiliary results on Monte Carlo approximations (Subsection 2.1), affine functions (Subsection 2.2), SDEs (Subsections 2.3–2.4), and viscosity solutions for PDEs (Subsection 2.5).

2.1 Monte Carlo approximations

In this subsection we employ Kahane-Khintchine-type estimates from the literature (cf., e.g., Hytönen et al. [41, Theorem 6.2.4 in Subsection 6.2b]) to present the known L^p-Monte Carlo estimate in Corollary 2.5 below. Corollary 2.5 is an immediate consequence of Lemma 2.2 and Proposition 2.4 below. Lemma 2.2, in turn, follows from Hytönen et al. [41, Theorem 6.2.4 in Subsection 6.2b] and Proposition 2.4 is, e.g., proved as Corollary 5.12 in Cox et al. [15]. To simplify the accessibility of Proposition 2.4 and Corollary 2.5 below, we include in this subsection also the statement and the proof of the well-known L^2-Monte Carlo error analysis in Lemma 2.3 below.

Definition 2.1. *Let $p, q \in (0, \infty)$. Then we denote by $\mathfrak{K}_{p,q} \in [0, \infty]$ the extended real number given by*

$$\mathfrak{K}_{p,q} = \sup\left\{ c \in [0,\infty) \colon \begin{bmatrix} \exists\, \mathbb{R}\text{-}Banach\ space\ (E, \|\cdot\|_E)\colon \\ \exists\ probability\ space\ (\Omega, \mathcal{F}, \mathbb{P})\colon \\ \exists\, \mathbb{P}\text{-}Rademacher\ family\ r_j \colon \Omega \to \{-1,1\},\, j \in \mathbb{N}\colon \\ \exists\, k \in \mathbb{N} \colon \exists\, x_1, x_2, \ldots, x_k \in E\backslash\{0\}\colon \\ \left(\mathbb{E}\big[\| \sum_{j=1}^k r_j x_j \|_E^p\big]\right)^{1/p} = c \left(\mathbb{E}\big[\| \sum_{j=1}^k r_j x_j \|_E^q\big]\right)^{1/q} \end{bmatrix} \right\} \tag{7}$$

and we call $\mathfrak{K}_{p,q}$ the (p,q)-Kahane-Khintchine constant.

Lemma 2.2. *For every $p \in [2, \infty)$ let $\mathfrak{K}_{p,2}$ be the $(p,2)$-Kahane-Khintchine constant (cf. Definition 2.1). Then it holds for all $p \in [2, \infty)$ that*

$$\mathfrak{K}_{p,2} \leq \sqrt{p-1}. \tag{8}$$

Proof of Lemma 2.2. Throughout this proof let $(E, \|\cdot\|_E)$ be a $\mathbb{R}$-Banach space, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $r_j \colon \Omega \to \{-1, 1\}$, $j \in \mathbb{N}$, be independent random variables which satisfy for all $j \in \mathbb{N}$ that

$$\mathbb{P}(r_j = -1) = \mathbb{P}(r_j = 1) = \frac{1}{2}, \tag{9}$$

and let $k \in \mathbb{N}$, $x_1, x_2, \ldots, x_k \in E\backslash\{0\}$. Note that Hytönen et al. [41, Theorem 6.2.4 in Subsection 6.2b] (with $X = E$, $q = p$, $p = 2$ for $p \in [2, \infty)$) in the notation of [41, Theorem 6.2.4]) implies that for all $p \in [2, \infty)$ it holds that

$$\Big(\mathbb{E}\big[\| \textstyle\sum_{j=1}^{k} r_j x_j \|_E^p\big]\Big)^{1/p} \leq (p-1)^{1/2} \Big(\mathbb{E}\big[\| \textstyle\sum_{j=1}^{k} r_j x_j \|_E^2\big]\Big)^{1/2}. \tag{10}$$

Combining this with (7) and the fact that $\mathbb{E}\big[\| \sum_{j=1}^{k} r_j x_j \|_E^2\big] > 0$ ensures that for all $p \in [2, \infty)$ it holds that $\mathfrak{K}_{p,2} \leq (p-1)^{1/2}$. The proof of Lemma 2.2 is thus completed. □

Lemma 2.3. *Let $n \in \mathbb{N}$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X_i \colon \Omega \to \mathbb{R}$, $i \in \{1, 2, \ldots, n\}$, be i.i.d. random variables with $\mathbb{E}\big[|X_1|\big] < \infty$. Then it holds that*

$$\big(\mathbb{E}\big[|\mathbb{E}[X_1] - \tfrac{1}{n}\big(\textstyle\sum_{i=1}^{n} X_i\big)|^2\big]\big)^{1/2} = n^{-1/2} \big(\mathbb{E}\big[|X_1 - \mathbb{E}[X_1]|^2\,\big]\big)^{1/2}. \tag{11}$$

Proof of Lemma 2.3. Note that the fact that for all independent random variables $Y, Z \colon \Omega \to \mathbb{R}$ with $\mathbb{E}[|Y| + |Z|] < \infty$ it holds that $\mathbb{E}[|YZ|] < \infty$ and $\mathbb{E}[YZ] = \mathbb{E}[Y]\,\mathbb{E}[Z]$ (cf., e.g., Klenke [46, Theorem 5.4]) and the hypothesis that $X_i \colon \Omega \to \mathbb{R}$, $i \in \{1, 2, \ldots, n\}$, are i.i.d. random variables assure that

$$\begin{aligned}
&\mathbb{E}\Big[\big|\mathbb{E}[X_1] - \tfrac{1}{n}\big(\textstyle\sum_{i=1}^{n} X_i\big)\big|^2\Big] \\
&= \mathbb{E}\Big[\big|\tfrac{1}{n}\big(\textstyle\sum_{i=1}^{n} \mathbb{E}[X_1] - X_i\big)\big|^2\Big] \\
&= \tfrac{1}{n^2}\mathbb{E}\Big[\big|\textstyle\sum_{i=1}^{n} \mathbb{E}[X_i] - X_i\big|^2\Big] \\
&= \tfrac{1}{n^2}\Big[\textstyle\sum_{i,j=1}^{n} \mathbb{E}[(\mathbb{E}[X_i] - X_i)(\mathbb{E}[X_j] - X_j)]\Big] \\
&= \tfrac{1}{n^2}\Big[\sum_{i=1}^{n} \mathbb{E}\big[|\mathbb{E}[X_i] - X_i|^2\big]\Big] \\
&\qquad + \tfrac{1}{n^2}\Bigg[\sum_{i,j=1, i\neq j}^{n} \mathbb{E}\big[(\mathbb{E}[X_i] - X_i)(\mathbb{E}[X_j] - X_j)\big]\Bigg] \\
&= \tfrac{1}{n^2}\big(n\,\mathbb{E}\big[|\mathbb{E}[X_1] - X_1|^2\big]\big) = n^{-1}\mathbb{E}\big[|X_1 - \mathbb{E}[X_1]|^2\big].
\end{aligned} \tag{12}$$

The proof of Lemma 2.3 is thus completed. □

Proposition 2.4. *Let $p \in [2, \infty)$, $d, n \in \mathbb{N}$, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the d-dimensional Euclidean norm, let $\mathfrak{K}_{p,2} \in (0, \infty)$ be the $(p, 2)$-Kahane-Khintchine constant (cf. Definition 2.1), let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space,* *and let $X_i\colon \Omega \to \mathbb{R}^d$, $i \in \{1, 2, \ldots, n\}$, be i.i.d. random variables with $\mathbb{E}\big[\|X_1\|\big] < \infty$. Then it holds that*

$$\big(\mathbb{E}\big[\|\mathbb{E}[X_1] - \tfrac{1}{n}\big(\textstyle\sum_{i=1}^n X_i\big)\|^p\big]\big)^{1/p} \leq \frac{2\,\mathfrak{K}_{p,2}}{\sqrt{n}} \big(\mathbb{E}\big[\|X_1 - \mathbb{E}[X_1]\|^p\,\big]\big)^{1/p}. \tag{13}$$

Corollary 2.5. *Let $p \in [2, \infty)$, $d, n \in \mathbb{N}$, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the d-dimensional Euclidean norm, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X_i\colon \Omega \to \mathbb{R}^d$, $i \in \{1, 2, \ldots, n\}$, be i.i.d. random variables with $\mathbb{E}\big[\|X_1\|\big] < \infty$. Then it holds that*

$$\big(\mathbb{E}\big[\|\mathbb{E}[X_1] - \tfrac{1}{n}\big(\textstyle\sum_{i=1}^n X_i\big)\|^p\big]\big)^{1/p} \leq 2\left[\frac{p-1}{n}\right]^{1/2} \big(\mathbb{E}\big[\|X_1 - \mathbb{E}[X_1]\|^p\,\big]\big)^{1/p}. \tag{14}$$

Proof of Corollary 2.5. Note that Proposition 2.4 and Lemma 2.2 demonstrate that

$$\begin{aligned}\big(\mathbb{E}\big[\|\mathbb{E}[X_1] - \tfrac{1}{n}\big(\textstyle\sum_{i=1}^n X_i\big)\|^p\big]\big)^{1/p} &\leq \frac{2\,\mathfrak{K}_{p,2}}{\sqrt{n}} \big(\mathbb{E}\big[\|X_1 - \mathbb{E}[X_1]\|^p\,\big]\big)^{1/p} \\ &\leq \frac{2\sqrt{p-1}}{\sqrt{n}} \big(\mathbb{E}\big[\|X_1 - \mathbb{E}[X_1]\|^p\,\big]\big)^{1/p} \\ &= 2\left[\frac{p-1}{n}\right]^{1/2} \big(\mathbb{E}\big[\|X_1 - \mathbb{E}[X_1]\|^p\,\big]\big)^{1/p}.\end{aligned} \tag{15}$$

The proof of Corollary 2.5 is thus completed. ☐

2.2 Properties of affine functions

This subsection recalls in Lemmas 2.6–2.7 and Corollaries 2.8–2.10 a few well-known properties for affine functions. For the sake of completeness we include in this subsection also proofs for Lemmas 2.6–2.7 and Corollaries 2.8–2.10.

Lemma 2.6. *Let $d, m \in \mathbb{N}$, $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$ and let $\varphi\colon \mathbb{R}^d \to \mathbb{R}^m$ be the function which satisfies for all $x \in \mathbb{R}^d$ that*

$$\varphi(x) = Ax + b. \tag{16}$$

Then it holds for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that

$$\varphi(\lambda x + y) + \lambda\varphi(0) = \lambda\varphi(x) + \varphi(y). \tag{17}$$

Proof of Lemma 2.6. Observe that (16) assures that for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ it holds that

$$\begin{aligned} \varphi(\lambda x + y) + \lambda\varphi(0) &= A(\lambda x + y) + b + \lambda(A \cdot 0 + b) \\ &= \lambda(Ax + b) + Ay + b = \lambda\varphi(x) + \varphi(y). \end{aligned} \tag{18}$$

The proof of Lemma 2.6 is thus completed. □

Lemma 2.7. *Let $d, m \in \mathbb{N}$, $e_1, e_2, \ldots, e_d \in \mathbb{R}^d$ satisfy $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0), \ldots,$ $e_d = (0, \ldots, 0, 1)$, let $\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_m)\colon \mathbb{R}^d \to \mathbb{R}^m$ be a function which satisfies for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that*

$$\varphi(\lambda x + y) + \lambda\varphi(0) = \lambda\varphi(x) + \varphi(y), \tag{19}$$

and let $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$ satisfy $b = \varphi(0)$ and

$$\begin{aligned} A &= \begin{pmatrix} \varphi_1(e_1) - \varphi_1(0) & \varphi_1(e_2) - \varphi_1(0) & \ldots & \varphi_1(e_d) - \varphi_1(0) \\ \varphi_2(e_1) - \varphi_2(0) & \varphi_2(e_2) - \varphi_2(0) & \ldots & \varphi_2(e_d) - \varphi_2(0) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_m(e_1) - \varphi_m(0) & \varphi_m(e_2) - \varphi_m(0) & \ldots & \varphi_m(e_d) - \varphi_m(0) \end{pmatrix} \\ &= \Big(\varphi(e_1) - \varphi(0) \Big| \varphi(e_2) - \varphi(0) \Big| \cdots \Big| \varphi(e_d) - \varphi(0) \Big). \end{aligned} \tag{20}$$

Then it holds for all $x \in \mathbb{R}^d$ that

$$\varphi(x) = Ax + b. \tag{21}$$

Proof of Lemma 2.7. First, note that (19) implies that for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ it holds that

$$\varphi(\lambda x + y) = \lambda(\varphi(x) - \varphi(0)) + \varphi(y). \tag{22}$$

This, induction, and (20) assure that for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ it holds

that

$$
\begin{aligned}
\varphi(x) &= \varphi\left(\sum_{i=1}^{d} x_i e_i\right) = \varphi\left(x_1 e_1 + \sum_{i=2}^{d} x_i e_i\right) \\
&= x_1(\varphi(e_1) - \varphi(0)) + \varphi\left(\sum_{i=2}^{d} x_i e_i\right) \\
&= \left[\sum_{i=1}^{\min\{1,d\}} x_i(\varphi(e_i) - \varphi(0))\right] + \varphi\left(\sum_{i=\min\{1,d\}+1}^{d} x_i e_i\right) \\
&= \left[\sum_{i=1}^{\min\{2,d\}} x_i(\varphi(e_i) - \varphi(0))\right] + \varphi\left(\sum_{i=\min\{2,d\}+1}^{d} x_i e_i\right) \\
&= \ldots \\
&= \left[\sum_{i=1}^{\min\{d,d\}} x_i(\varphi(e_i) - \varphi(0))\right] + \varphi\left(\sum_{i=\min\{d,d\}+1}^{d} x_i e_i\right) \\
&= \left[\sum_{i=1}^{d} x_i(\varphi(e_i) - \varphi(0))\right] + \varphi(0) \\
&= Ax + b.
\end{aligned}
\tag{23}
$$

The proof of Lemma 2.7 is thus completed. □

Corollary 2.8. *Let $d, m \in \mathbb{N}$ and let $\varphi\colon \mathbb{R}^d \to \mathbb{R}^m$ be a function. Then the following two statements are equivalent:*

(i) There exist $A \in \mathbb{R}^{m\times d}$, $b \in \mathbb{R}^m$ such that for all $x \in \mathbb{R}^d$ it holds that

$$\varphi(x) = Ax + b. \tag{24}$$

(ii) It holds for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that

$$\varphi(\lambda x + y) + \lambda\varphi(0) = \lambda\varphi(x) + \varphi(y). \tag{25}$$

Proof of Corollary 2.8. Note that Lemma 2.6 establishes that ((i) ⇒ (ii)). In addition, observe that Lemma 2.7 demonstrates that ((ii) ⇒ (i)). The proof of Corollary 2.8 is thus completed. □

Corollary 2.9. *Let $d, m \in \mathbb{N}$, let $\varphi\colon \mathbb{R}^d \to \mathbb{R}^m$ be a function which satisfies for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that*

$$\varphi(\lambda x + y) + \lambda\varphi(0) = \lambda\varphi(x) + \varphi(y), \tag{26}$$

and for every $k \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^k} : \mathbb{R}^k \to [0, \infty)$ be the k-dimensional Euclidean norm. Then there exists $c \in [0, \infty)$ such that for all $x, y \in \mathbb{R}^d$ it holds that

$$\|\varphi(x)\|_{\mathbb{R}^m} \leq c(1 + \|x\|_{\mathbb{R}^d}) \quad \text{and} \quad \|\varphi(x) - \varphi(y)\|_{\mathbb{R}^m} \leq c \|x - y\|_{\mathbb{R}^d} . \tag{27}$$

Proof of Corollary 2.9. Throughout this proof let $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$ satisfy for all $x \in \mathbb{R}^d$ that

$$\varphi(x) = Ax + b \tag{28}$$

(cf. Corollary 2.8) and let $c \in [0, \infty)$ be given by

$$c = \max\left\{ \left[\sup_{v \in \mathbb{R}^d \setminus \{0\}} \tfrac{\|Av\|_{\mathbb{R}^m}}{\|v\|_{\mathbb{R}^d}}\right], \|b\|_{\mathbb{R}^m} \right\}. \tag{29}$$

Note that (28) and (29) assure that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \|\varphi(x)\|_{\mathbb{R}^m} &= \|Ax + b\|_{\mathbb{R}^m} \leq \|Ax\|_{\mathbb{R}^m} + \|b\|_{\mathbb{R}^m} \\ &\leq \left[\sup_{v \in \mathbb{R}^d \setminus \{0\}} \tfrac{\|Av\|_{\mathbb{R}^m}}{\|v\|_{\mathbb{R}^d}}\right] \|x\|_{\mathbb{R}^d} + \|b\|_{\mathbb{R}^m} \leq c(\|x\|_{\mathbb{R}^d} + 1). \end{aligned} \tag{30}$$

Furthermore, observe that (28) and (29) imply that for all $x, y \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \|\varphi(x) - \varphi(y)\|_{\mathbb{R}^m} &= \|(Ax + b) - (Ay + b)\|_{\mathbb{R}^m} = \|A(x - y)\|_{\mathbb{R}^m} \\ &\leq \left[\sup_{v \in \mathbb{R}^d \setminus \{0\}} \tfrac{\|Av\|_{\mathbb{R}^m}}{\|v\|_{\mathbb{R}^d}}\right] \|x - y\|_{\mathbb{R}^d} \\ &\leq c \|x - y\|_{\mathbb{R}^d} . \end{aligned} \tag{31}$$

Combining this and (30) establishes (27). The proof of Corollary 2.9 is thus completed. ◻

Corollary 2.10. *Let $d, k, m \in \mathbb{N}$, let $\sigma \colon \mathbb{R}^d \to \mathbb{R}^{k \times m}$ be a function which satisfies for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that*

$$\sigma(\lambda x + y) + \lambda \sigma(0) = \lambda \sigma(x) + \sigma(y), \tag{32}$$

let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the d-dimensional Euclidean norm, and let $|||\cdot||| \colon \mathbb{R}^{k \times m} \to [0, \infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{k \times m}$. Then there exists $c \in [0, \infty)$ such that for all $x, y \in \mathbb{R}^d$ it holds that

$$|||\sigma(x)||| \leq c(1 + \|x\|) \qquad \text{and} \qquad |||\sigma(x) - \sigma(y)||| \leq c \|x - y\| . \tag{33}$$

Proof of Corollary 2.10. Throughout this proof for every $\mathfrak{d} \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^{\mathfrak{d}}} : \mathbb{R}^{\mathfrak{d}} \to [0, \infty)$ be the $\mathfrak{d}$-dimensional Euclidean norm, let $e_1, e_2, \ldots, e_m \in \mathbb{R}^m$ satisfy $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0)$, …, $e_m = (0, \ldots, 0, 1)$, and let $\varphi \colon \mathbb{R}^d \to \mathbb{R}^{(mk)}$ be the function which satisfies for all $x \in \mathbb{R}^d$ that

$$\varphi(x) = \begin{pmatrix} (\sigma(x))e_1 \\ (\sigma(x))e_2 \\ \ldots \\ (\sigma(x))e_m \end{pmatrix}. \tag{34}$$

Note that (32) and (34) ensure that for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ it holds that

$$\begin{aligned} \varphi(\lambda x + y) + \lambda\varphi(0) &= \begin{pmatrix} (\sigma(\lambda x + y))e_1 \\ (\sigma(\lambda x + y))e_2 \\ \ldots \\ (\sigma(\lambda x + y))e_m \end{pmatrix} + \lambda \begin{pmatrix} (\sigma(0))e_1 \\ (\sigma(0))e_2 \\ \ldots \\ (\sigma(0))e_m \end{pmatrix} \\ &= \begin{pmatrix} (\sigma(\lambda x + y))e_1 + \lambda(\sigma(0))e_1 \\ (\sigma(\lambda x + y))e_2 + \lambda(\sigma(0))e_2 \\ \ldots \\ (\sigma(\lambda x + y))e_m + \lambda(\sigma(0))e_m \end{pmatrix} = \begin{pmatrix} [\sigma(\lambda x + y) + \lambda\sigma(0)]\, e_1 \\ [\sigma(\lambda x + y) + \lambda\sigma(0)]\, e_2 \\ \ldots \\ [\sigma(\lambda x + y) + \lambda\sigma(0)]\, e_m \end{pmatrix} \\ &= \begin{pmatrix} [\lambda\sigma(x) + \sigma(y)]\, e_1 \\ [\lambda\sigma(x) + \sigma(y)]\, e_2 \\ \ldots \\ [\lambda\sigma(x) + \sigma(y)]\, e_m \end{pmatrix} = \lambda \begin{pmatrix} (\sigma(x))e_1 \\ (\sigma(x))e_2 \\ \ldots \\ (\sigma(x))e_m \end{pmatrix} + \begin{pmatrix} (\sigma(y))e_1 \\ (\sigma(y))e_2 \\ \ldots \\ (\sigma(y))e_m \end{pmatrix} = \lambda\varphi(x) + \varphi(y). \end{aligned} \tag{35}$$

This and Corollary 2.9 (with $d = d$, $m = mk$, $\varphi = \varphi$ in the notation of Corollary 2.9) imply that there exists $c \in [0, \infty)$ such that for all $x, y \in \mathbb{R}^d$ it holds that

$$\|\varphi(x)\|_{\mathbb{R}^{(mk)}} \leq c(1 + \|x\|_{\mathbb{R}^d}) \quad \text{and} \quad \|\varphi(x) - \varphi(y)\|_{\mathbb{R}^{(mk)}} \leq c\, \|x - y\|_{\mathbb{R}^d}. \tag{36}$$

Furthermore, note that for all $x, y \in \mathbb{R}^d$ it holds that

$$|||\sigma(x)|||^2 = \sum_{j=1}^{m} \|[\sigma(x)]e_j\|_{\mathbb{R}^k}^2 = \|\varphi(x)\|_{\mathbb{R}^{(mk)}}^2 \tag{37}$$

and

$$|||\sigma(x) - \sigma(y)|||^2 = \sum_{j=1}^{m} \|[\sigma(x) - \sigma(y)]e_j\|_{\mathbb{R}^k}^2 = \|\varphi(x) - \varphi(y)\|_{\mathbb{R}^{(mk)}}^2. \tag{38}$$

Combining this with (36) ensures that for all $x, y \in \mathbb{R}^d$ it holds that

$$|||\sigma(x)||| = \|\varphi(x)\|_{\mathbb{R}^{(mk)}} \leq c(1 + \|x\|_{\mathbb{R}^d}) \tag{39}$$

and

$$|||\sigma(x) - \sigma(y)||| = \|\varphi(x) - \varphi(y)\|_{\mathbb{R}^{(mk)}} \leq c \,\|x - y\|_{\mathbb{R}^d}\,. \tag{40}$$

The proof of Corollary 2.10 is thus completed. □

2.3 A priori estimates for solutions of stochastic differential equations

In this subsection we establish in Proposition 2.14 below an elementary a priori estimate for solutions of SDEs with at most linearly growing coefficient functions (see (50) in Proposition 2.14 below for details). Our proof of Proposition 2.14 employs the Gronwall integral inequality (see Lemma 2.11 below), a special case of Minkowksi's integral inequality (see Lemma 2.12 below), and the Burkholder-Davis-Gundy type inequality in Da Prato & Zabczyk [19, Lemma 7.7] (see Lemma 2.13 below). For the sake of completeness we include in this subsection also the proof of Lemma 2.11. Lemma 2.12 follows, e.g., from Garling [28, Corollary 5.4.2] or Jentzen & Kloeden [42, Corollary A.1 in Appendix A]. Lemma 2.13 is, e.g., proved as Lemma 7.7 in Da Prato & Zabczyk [19].

Lemma 2.11. *Let $\alpha, \beta, T \in [0, \infty)$ and let $f\colon [0, T] \to \mathbb{R}$ be a $\mathcal{B}([0, T])/\mathcal{B}(\mathbb{R})$-measurable function which satisfies for all $t \in [0, T]$ that $\int_0^T |f(s)|\,ds < \infty$ and*

$$f(t) \leq \alpha + \beta \int_0^t f(s)\,ds. \tag{41}$$

Then it holds for all $t \in [0, T]$ that

$$f(t) \leq \alpha e^{\beta t}. \tag{42}$$

Proof of Lemma 2.11. Throughout this proof assume w.l.o.g. that $T > 0$ and let $u\colon [0, T] \to \mathbb{R}$ be the function which satisfies for all $t \in [0, T]$ that

$$u(t) = \alpha + \beta \int_0^t f(s)\,ds. \tag{43}$$

Observe that (41) and (43) imply that for all $t \in [0, T]$ it holds that

$$f(t) \leq u(t). \tag{44}$$

Next note that (43) and the assumption that $\int_0^T |f(s)| \, ds < \infty$ assure that u is absolutely continuous and that for Lebesgue-almost all $t \in [0, T]$ it holds that

$$u'(t) = \beta f(t) \tag{45}$$

(cf., e.g., Jones [43, Page 550 in Section E in Chapter 16]). This, the integration by parts formula for absolutely continuous functions (cf., e.g., Jones [43, Page 553 in Section F in Chapter 16]), (43), and (44) imply that for all $t \in [0, T]$ it holds that

$$\begin{aligned} u(t)e^{-\beta t} &= u(0)e^0 + \int_0^t \left[u'(s)e^{-\beta s} + u(s)\tfrac{d}{ds}(e^{-\beta s})\right] ds \\ &= \alpha + \int_0^t \left[\beta f(s)e^{-\beta s} + u(s)(-\beta)e^{-\beta s}\right] ds \\ &= \alpha + \int_0^t \beta e^{-\beta s} \left[f(s) - u(s)\right] ds \leq \alpha. \end{aligned} \tag{46}$$

Combining this and (44) assures that for all $t \in [0, T]$ it holds that

$$f(t) \leq u(t) = u(t)e^{-\beta t}e^{\beta t} \leq \alpha e^{\beta t}. \tag{47}$$

The proof of Lemma 2.11 is thus completed. □

Lemma 2.12 (Moments of pathwise integrals)**.** *Let $T \in (0, \infty)$, $p \in [1, \infty)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X \colon [0, T] \times \Omega \to [0, \infty)$ be a $(\mathcal{B}([0, T]) \otimes \mathcal{F})/\mathcal{B}([0, \infty))$-measurable function. Then it holds for all $t \in [0, T]$ that*

$$\left(\mathbb{E}\left[\left|\int_0^t X_s \, ds\right|^p\right]\right)^{1/p} \leq \int_0^t \left(\mathbb{E}\left[|X_s|^p\right]\right)^{1/p} ds. \tag{48}$$

Lemma 2.13. *Let $d, m \in \mathbb{N}$, $p \in [2, \infty)$, $T \in (0, \infty)$, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the d-dimensional Euclidean norm, let $|||\cdot||| \colon \mathbb{R}^{d \times m} \to [0, \infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d \times m}$, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t \in [0,T]})$ be a filtered probability space which fulfils the usual conditions, let $W \colon [0, T] \times \Omega \to \mathbb{R}^m$ be a standard*

$(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion, and let $X\colon[0,T]\times\Omega\to\mathbb{R}^{d\times m}$ be an $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic process which satisfies $\mathbb{P}\big(\int_0^T |||X_s|||^2\,ds<\infty\big)=1$. Then it holds for all $t\in[0,T]$, $s\in[0,t]$ that

$$\left(\mathbb{E}\left[\left\|\int_s^t X_r\,dW_r\right\|^p\right]\right)^{1/p}\leq\left[\frac{p(p-1)}{2}\right]^{1/2}\left[\int_s^t\big(\mathbb{E}[|||X_r|||^p]\big)^{2/p}\,dr\right]^{1/2}. \quad (49)$$

Proposition 2.14. *Let $d,m\in\mathbb{N}$, $p\in[2,\infty)$, $T,\mathfrak{m}_1,\mathfrak{m}_2,\mathfrak{s}_1,\mathfrak{s}_2\in[0,\infty)$, $\xi\in\mathbb{R}^d$, let $\|\cdot\|:\mathbb{R}^d\to[0,\infty)$ be the d-dimensional Euclidean norm, let $|||\cdot|||\colon\mathbb{R}^{d\times m}\to[0,\infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times m}$, let $(\Omega,\mathcal{F},\mathbb{P},(\mathbb{F}_t)_{t\in[0,T]})$ be a filtered probability space which fulfils the usual conditions, let $W\colon[0,T]\times\Omega\to\mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion, let $\mu\colon\mathbb{R}^d\to\mathbb{R}^d$ be $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable, let $\sigma\colon\mathbb{R}^d\to\mathbb{R}^{d\times m}$ be $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^{d\times m})$-measurable, assume for all $x\in\mathbb{R}^d$ that*

$$\|\mu(x)\|\leq\mathfrak{m}_1+\mathfrak{m}_2\|x\| \qquad \text{and} \qquad |||\sigma(x)|||\leq\mathfrak{s}_1+\mathfrak{s}_2\|x\|, \quad (50)$$

and let $X\colon[0,T]\times\Omega\to\mathbb{R}^d$ be an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths which satisfies that for all $t\in[0,T]$ it holds $\mathbb{P}$-a.s. that

$$X_t=\xi+\int_0^t\mu(X_s)\,ds+\int_0^t\sigma(X_s)\,dW_s. \quad (51)$$

Then it holds for all $t\in[0,T]$ that

$$\begin{aligned}&\big(\mathbb{E}\big[\|X_t\|^p\big]\big)^{1/p}\\&\leq\sqrt{2}\Big(\|\xi\|+\mathfrak{m}_1T+\mathfrak{s}_1\sqrt{\tfrac{p(p-1)T}{2}}\Big)\exp\left(\left[\mathfrak{m}_2\sqrt{T}+\mathfrak{s}_2\sqrt{\tfrac{p(p-1)}{2}}\right]^2t\right)\\&\leq\sqrt{2}\Big(\|\xi\|+\mathfrak{m}_1T+\mathfrak{s}_1p\sqrt{T}\Big)\exp\left(\left[\mathfrak{m}_2\sqrt{T}+\mathfrak{s}_2p\right]^2t\right).\end{aligned} \quad (52)$$

Proof of Proposition 2.14. Throughout this proof assume w.l.o.g. that $T>0$ and let $\tau_n\colon\Omega\to[0,T]$, $n\in\mathbb{N}$, be the functions which satisfy for every $n\in\mathbb{N}$ that

$$\tau_n=\inf(\{t\in[0,T]\colon\ \|X_t\|>n\}\cup\{T\}). \quad (53)$$

Note that the hypothesis that $X\colon[0,T]\times\Omega\to\mathbb{R}^d$ is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths ensures that for all $t\in(0,T]$,

$n \in \mathbb{N}$ it holds that

$$\begin{aligned} \{\tau_n < t\} &= \{\exists\, s \in [0,t)\colon\ \|X_s\| > n\} \\ &= \{\exists\, s \in [0,t) \cap \mathbb{Q}\colon\ \|X_s\| > n\} \\ &= \big(\cup_{s\in[0,t)\cap\mathbb{Q}}\{\|X_s\| > n\}\big) \in \mathbb{F}_t. \end{aligned} \tag{54}$$

This demonstrates that for all $t \in [0,T)$, $r \in (t,T]$, $n \in \mathbb{N}$ it holds that

$$\{\tau_n \le t\} = \big(\cap_{k\in\mathbb{N}}\{\tau_n < t + \tfrac{1}{k}\}\big) = \big(\cap_{k\in\mathbb{N},\, t+1/k\le r}\{\tau_n < t + \tfrac{1}{k}\}\big) \in \mathbb{F}_r. \tag{55}$$

The hypothesis that $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ fulfils the usual conditions hence ensures that for all $t \in [0,T)$, $n \in \mathbb{N}$ it holds that $\{\tau_n \le t\} \in \mathbb{F}_t^+ = \mathbb{F}_t$. Therefore, we obtain that for all $n \in \mathbb{N}$ it holds that τ_n is an $(\mathbb{F}_t)_{t\in[0,T]}$-stopping time. Moreover, observe that (51) and the triangle inequality assure that for all $t \in [0,T]$, $n \in \mathbb{N}$ it holds that

$$\begin{aligned} \big(\mathbb{E}\big[\|X_{\min\{t,\tau_n\}}\|^p\big]\big)^{1/p} \le \|\xi\| &+ \left(\mathbb{E}\left[\left\|\int_0^{\min\{t,\tau_n\}} \mu(X_s)\, ds\right\|^p\right]\right)^{1/p} \\ &+ \left(\mathbb{E}\left[\left\|\int_0^{\min\{t,\tau_n\}} \sigma(X_s)\, dW_s\right\|^p\right]\right)^{1/p}. \end{aligned} \tag{56}$$

Next note that Lemma 2.12, (50), and the triangle inequality demonstrate that for all $t \in [0,T]$, $n \in \mathbb{N}$ it holds that

$$\begin{aligned} \left(\mathbb{E}\left[\left\|\int_0^{\min\{t,\tau_n\}} \mu(X_s)\, ds\right\|^p\right]\right)^{1/p} &\le \left(\mathbb{E}\left[\left|\int_0^{\min\{t,\tau_n\}} \|\mu(X_s)\|\ ds\right|^p\right]\right)^{1/p} \\ &\le \int_0^t \big(\mathbb{E}\big[\|\mu(X_s)\|^p \mathbb{1}_{\{s\le\tau_n\}}\big]\big)^{1/p}\, ds \\ &\le \int_0^t \big(\mathbb{E}\big[\|\mu(X_{\min\{s,\tau_n\}})\|^p\big]\big)^{1/p}\, ds \\ &\le \int_0^t \big(\mathbb{E}\big[\big(\mathfrak{m}_1 + \mathfrak{m}_2\|X_{\min\{s,\tau_n\}}\|\big)^p\big]\big)^{1/p}\, ds \\ &\le \int_0^t \Big[\mathfrak{m}_1 + \mathfrak{m}_2\,\big(\mathbb{E}\big[\|X_{\min\{s,\tau_n\}}\|^p\big]\big)^{1/p}\Big]\, ds \\ &= \mathfrak{m}_1 t + \mathfrak{m}_2\left(\int_0^t \big(\mathbb{E}\big[\|X_{\min\{s,\tau_n\}}\|^p\big]\big)^{1/p}\, ds\right). \end{aligned} \tag{57}$$

The Cauchy-Schwarz inequality hence proves that for all $t \in [0,T]$, $n \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left\|\int_0^{\min\{t,\tau_n\}} \mu(X_s)\, ds\right\|^p\right]\right)^{1/p} \\
&\leq \mathfrak{m}_1 T + \mathfrak{m}_2 \left[\int_0^t 1^2\, ds\right]^{1/2} \left[\int_0^t \left(\mathbb{E}\left[\|X_{\min\{s,\tau_n\}}\|^p\right]\right)^{2/p} ds\right]^{1/2} \\
&\leq \mathfrak{m}_1 T + \mathfrak{m}_2 \sqrt{T} \left[\int_0^t \left(\mathbb{E}\left[\|X_{\min\{s,\tau_n\}}\|^p\right]\right)^{2/p} ds\right]^{1/2}.
\end{aligned} \tag{58}$$

Moreover, note that the hypothesis that $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths shows that $X\colon [0,T] \times \Omega \to \mathbb{R}^d$ is an $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic process. The fact that for every $n \in \mathbb{N}$ it holds that $([0,T]\times\Omega \ni (t,\omega) \mapsto \mathbb{1}_{\{t\leq\tau_n(\omega)\}} \in \{0,1\})$ is an $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic process (cf., e.g., Kallenberg [44, Lemma 22.1]) and the hypothesis that $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ is a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^{d\times m})$-measurable function hence ensure that for every $n\in\mathbb{N}$ it holds that

$$([0,T]\times\Omega \ni (t,\omega) \mapsto \sigma(X_t(\omega))\mathbb{1}_{\{t\leq\tau_n(\omega)\}} \in \mathbb{R}^{d\times m}) \tag{59}$$

is an $(\mathbb{F}_t)_{t\in[0,T]}$-predictable stochastic process. Combining this, (50), and (53) with the hypothesis that $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ has continuous sample paths demonstrates that for all $n \in \mathbb{N}\cap(\|\xi\|,\infty)$ it holds that

$$\begin{aligned}
\int_0^T \left|\!\left|\!\left|\sigma(X_s)\mathbb{1}_{\{s\leq\tau_n\}}\right|\!\right|\!\right|^2 ds &\leq T\left[\sup_{s\in[0,\tau_n]} |\!|\!|\sigma(X_s)|\!|\!|^2\right] \\
&\leq T\left[\sup_{s\in[0,\tau_n]} \left[(\mathfrak{s}_1 + \mathfrak{s}_2\|X_s\|)^2\right]\right] \\
&\leq T(\mathfrak{s}_1+\mathfrak{s}_2 n)^2 < \infty.
\end{aligned} \tag{60}$$

Lemma 2.13, (59), (50), and the triangle inequality therefore establish that for all $t \in [0,T]$, $n \in \mathbb{N} \cap (\|\xi\|, \infty)$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left\|\int_0^{\min\{t,\tau_n\}} \sigma(X_s)\, dW_s\right\|^p\right]\right)^{1/p} \\
&= \left(\mathbb{E}\left[\left\|\int_0^t \sigma(X_s) \mathbb{1}_{\{s \le \tau_n\}}\, dW_s\right\|^p\right]\right)^{1/p} \\
&\le \sqrt{\tfrac{p(p-1)}{2}} \left(\int_0^t \left(\mathbb{E}\left[|\!|\!|\sigma(X_s)|\!|\!|^p \mathbb{1}_{\{s \le \tau_n\}}\right]\right)^{2/p} ds\right)^{1/2} \\
&\le \sqrt{\tfrac{p(p-1)}{2}} \left(\int_0^t \left(\mathbb{E}\left[|\!|\!|\sigma(X_{\min\{s,\tau_n\}})|\!|\!|^p\right]\right)^{2/p} ds\right)^{1/2} \\
&\le \sqrt{\tfrac{p(p-1)}{2}} \left(\int_0^t \left(\mathbb{E}\left[\left(\mathfrak{s}_1 + \mathfrak{s}_2 \left\|X_{\min\{s,\tau_n\}}\right\|\right)^p\right]\right)^{2/p} ds\right)^{1/2} \\
&\le \sqrt{\tfrac{p(p-1)}{2}} \left(\int_0^t \left(\mathfrak{s}_1 + \mathfrak{s}_2 \left(\mathbb{E}\left[\|X_{\min\{s,\tau_n\}}\|^p\right]\right)^{1/p}\right)^2 ds\right)^{1/2} \\
&\le \sqrt{\tfrac{p(p-1)}{2}} \left(\mathfrak{s}_1 \left[\int_0^t 1^2\, ds\right]^{1/2} + \mathfrak{s}_2 \left[\int_0^t \left(\mathbb{E}\left[\|X_{\min\{s,\tau_n\}}\|^p\right]\right)^{2/p} ds\right]^{1/2}\right) \\
&\le \mathfrak{s}_1 \sqrt{\tfrac{p(p-1)T}{2}} + \mathfrak{s}_2 \sqrt{\tfrac{p(p-1)}{2}} \left[\int_0^t \left(\mathbb{E}\left[\|X_{\min\{s,\tau_n\}}\|^p\right]\right)^{2/p} ds\right]^{1/2}.
\end{aligned} \tag{61}$$

Combining this, (56), and (58) proves that for all $t \in [0,T]$, $n \in \mathbb{N} \cap (\|\xi\|, \infty)$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\|X_{\min\{t,\tau_n\}}\|^p\right]\right)^{1/p} \\
&\le \|\xi\| + \mathfrak{m}_1 T + \mathfrak{s}_1 \sqrt{\tfrac{p(p-1)T}{2}} \\
&\qquad + \left(\mathfrak{m}_2 \sqrt{T} + \mathfrak{s}_2 \sqrt{\tfrac{p(p-1)}{2}}\right) \left[\int_0^t \left(\mathbb{E}\left[\|X_{\min\{s,\tau_n\}}\|^p\right]\right)^{2/p} ds\right]^{1/2}.
\end{aligned} \tag{62}$$

The fact that for all $x, y \in \mathbb{R}$ it holds that $|x+y|^2 \le 2(x^2 + y^2)$ therefore demonstrates that for all $t \in [0,T]$, $n \in \mathbb{N} \cap (\|\xi\|, \infty)$ it holds that

$$
\begin{aligned}
&\big(\mathbb{E}\big[\|X_{\min\{t,\tau_n\}}\|^p\big]\big)^{2/p} \\
&\leq 2\left[\|\xi\| + \mathfrak{m}_1 T + \mathfrak{s}_1 \sqrt{\tfrac{p(p-1)T}{2}}\right]^2 \\
&\quad + 2\left[\mathfrak{m}_2 \sqrt{T} + \mathfrak{s}_2 \sqrt{\tfrac{p(p-1)}{2}}\right]^2 \left[\int_0^t \big(\mathbb{E}\big[\|X_{\min\{s,\tau_n\}}\|^p\big]\big)^{2/p} \, ds\right].
\end{aligned}
\tag{63}
$$

Next note that (53) ensures for all $n \in \mathbb{N} \cap (\|\xi\|, \infty)$ that

$$
\int_0^T \big(\mathbb{E}\big[\|X_{\min\{s,\tau_n\}}\|^p\big]\big)^{2/p} \, ds \leq \int_0^T \big(\mathbb{E}\big[n^p\big]\big)^{2/p} \, ds = Tn^2 < \infty. \tag{64}
$$

Combining this and (63) with Lemma 2.11 (with $\alpha = 2\big[\|\xi\| + \mathfrak{m}_1 T + \mathfrak{s}_1 \sqrt{p(p-1)T/2}\big]^2$, $\beta = 2\big[\mathfrak{m}_2\sqrt{T} + \mathfrak{s}_2\sqrt{p(p-1)/2}\big]^2$, $T = T$, $f = \big([0,T] \ni t \mapsto \big(\mathbb{E}\big[\|X_{\min\{t,\tau_n\}}\|^p\big]\big)^{2/p} \in \mathbb{R}\big)$ in the notation of Lemma 2.11) demonstrates that for all $t \in [0,T]$, $n \in \mathbb{N} \cap (\|\xi\|, \infty)$ it holds that

$$
\begin{aligned}
&\big(\mathbb{E}\big[\|X_{\min\{t,\tau_n\}}\|^p\big]\big)^{2/p} \\
&\leq 2\left[\|\xi\| + \mathfrak{m}_1 T + \mathfrak{s}_1 \sqrt{\tfrac{p(p-1)T}{2}}\right]^2 \exp\!\left(2\left[\mathfrak{m}_2\sqrt{T} + \mathfrak{s}_2\sqrt{\tfrac{p(p-1)}{2}}\right]^2 t\right).
\end{aligned}
\tag{65}
$$

Therefore, we obtain that for all $t \in [0,T]$, $n \in \mathbb{N} \cap (\|\xi\|, \infty)$ it holds that

$$
\begin{aligned}
&\big(\mathbb{E}\big[\|X_{\min\{t,\tau_n\}}\|^p\big]\big)^{1/p} \\
&\leq \sqrt{2}\left[\|\xi\| + \mathfrak{m}_1 T + \mathfrak{s}_1 \sqrt{\tfrac{p(p-1)}{2}}\sqrt{T}\right] \exp\!\left(\left[\mathfrak{m}_2\sqrt{T} + \mathfrak{s}_2\sqrt{\tfrac{p(p-1)}{2}}\right]^2 t\right).
\end{aligned}
\tag{66}
$$

Furthermore, observe that (53) and the fact that $X\colon [0,T] \times \Omega \to \mathbb{R}^d$ is a stochastic process with continuous sample paths ensure that for all $t \in [0,T]$ it holds that $\lim_{n\to\infty} \min\{t,\tau_n\} = t$. Therefore, we obtain that for all $t \in [0,T]$ it holds that

$$
\|X_t\| = \|X_{(\lim_{n\to\infty}\min\{t,\tau_n\})}\| = \|\lim_{n\to\infty} X_{\min\{t,\tau_n\}}\| = \lim_{n\to\infty} \|X_{\min\{t,\tau_n\}}\|. \tag{67}
$$

Fatou's Lemma and (66) hence imply for all $t \in [0,T]$ that

$$\begin{aligned}
&\big(\mathbb{E}\big[\|X_t\|^p\big]\big)^{1/p} = \Big(\mathbb{E}\big[\lim_{n\to\infty} \|X_{\min\{t,\tau_n\}}\|^p\big]\Big)^{1/p} \\
&\leq \Big(\liminf_{n\to\infty} \mathbb{E}\big[\|X_{\min\{t,\tau_n\}}\|^p\big]\Big)^{1/p} \leq \sup_{n\in\mathbb{N}\cap(\|\xi\|,\infty)} \big(\mathbb{E}\big[\|X_{\min\{t,\tau_n\}}\|^p\big]\big)^{1/p} \\
&\leq \sqrt{2}\left[\|\xi\| + \mathfrak{m}_1 T + \mathfrak{s}_1\sqrt{\tfrac{p(p-1)}{2}}\sqrt{T}\right] \exp\!\left(\left[\mathfrak{m}_2\sqrt{T} + \mathfrak{s}_2\sqrt{\tfrac{p(p-1)}{2}}\right]^2 t\right).
\end{aligned} \tag{68}$$

The fact that $\sqrt{\frac{p(p-1)}{2}} \leq \sqrt{p^2-p} \leq \sqrt{p^2} = p$ therefore establishes (52). The proof of Proposition 2.14 is thus completed. □

2.4 Stochastic differential equations with affine coefficient functions

In this subsection we establish in Proposition 2.20 elementary regularity properties for SDEs with affine coefficient functions. Our proof of Proposition 2.20, roughly speaking, employs the elementary results in Lemma 2.15 and Proposition 2.17 (which are, loosely speaking, alleviated versions of Proposition 2.20), the well-known fact that modifications of continuous stochastic processes are indistinguishable (cf. Lemma 2.16 below), the well-known fact that a modification of an adapted stochastic process is an adapted stochastic process (see Lemma 2.18 below for details), and a version of the Kolmogorov-Chentsov theorem (see Lemma 2.19 below for details). Results similar to Lemma 2.19 can, e.g., be found in Cox et al. [14, Theorem 3.5 in Subsection 3.1] and Mittmann & Steinwart [55, Theorem 2.1 in Section 2]. For the sake of completeness we include in this subsection also proofs for Lemmas 2.16 and 2.18.

Lemma 2.15. *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ be a filtered probability space which fulfils the usual conditions, let $W\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion, let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be functions which satisfy for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that*

$$\mu(\lambda x + y) + \lambda\mu(0) = \lambda\mu(x) + \mu(y) \tag{69}$$

and

$$\sigma(\lambda x + y) + \lambda\sigma(0) = \lambda\sigma(x) + \sigma(y), \tag{70}$$

and let $X^x\colon [0,T]\times\Omega\to\mathbb{R}^d$, $x\in\mathbb{R}^d$, *be* $(\mathbb{F}_t)_{t\in[0,T]}$*-adapted stochastic processes with continuous sample paths which satisfy that for all* $x\in\mathbb{R}^d$, $t\in[0,T]$ *it holds* $\mathbb{P}$*-a.s. that*

$$X_t^x = x + \int_0^t \mu(X_s^x)\,ds + \int_0^t \sigma(X_s^x)\,dW_s. \tag{71}$$

Then it holds for all $t\in[0,T]$, $x,y\in\mathbb{R}^d$, $\lambda\in\mathbb{R}$ *that*

$$\mathbb{P}\Big(X_t^{\lambda x+y} + \lambda X_t^0 = \lambda X_t^x + X_t^y\Big) = 1. \tag{72}$$

Proof of Lemma 2.15. Throughout this proof let $x,y\in\mathbb{R}^d$, $\lambda\in\mathbb{R}$ and let $Y\colon[0,T]\times\Omega\to\mathbb{R}^d$ be the stochastic process which satisfies for all $t\in[0,T]$ that

$$Y_t = \lambda(X_t^x - X_t^0) + X_t^y. \tag{73}$$

Note that the hypothesis that for all $z\in\mathbb{R}^d$ it holds that $X^z\colon[0,T]\times\Omega\to\mathbb{R}^d$ is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths assures that Y is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths. Moreover, observe that (71) and (73) ensure that for all $t\in[0,T]$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned}
Y_t &= \lambda(X_t^x - X_t^0) + X_t^y\\
&= \lambda\Big(\Big[x + \textstyle\int_0^t \mu(X_s^x)\,ds + \int_0^t \sigma(X_s^x)\,dW_s\Big]\\
&\quad - \Big[0 + \textstyle\int_0^t \mu(X_s^0)\,ds + \int_0^t \sigma(X_s^0)\,dW_s\Big]\Big)\\
&\quad + \Big[y + \textstyle\int_0^t \mu(X_s^y)\,ds + \int_0^t \sigma(X_s^y)\,dW_s\Big]\\
&= \lambda x + y + \textstyle\int_0^t \left[\lambda\left(\mu(X_s^x) - \mu(X_s^0)\right) + \mu(X_s^y)\right] ds\\
&\quad + \textstyle\int_0^t \left[\lambda\left(\sigma(X_s^x) - \sigma(X_s^0)\right) + \sigma(X_s^y)\right] dW_s.
\end{aligned} \tag{74}$$

In addition, note that (69) and (70) ensure that for all $\nu\in\{\mu,\sigma\}$, $a,b,c\in\mathbb{R}^d$, $\lambda\in\mathbb{R}$ it holds that

$$\begin{aligned}
\lambda\left(\nu(a) - \nu(b)\right) + \nu(c) &= \lambda\nu(a) + \nu(c) - \lambda\nu(b)\\
&= \nu(\lambda a + c) + \lambda\nu(0) - \lambda\nu(b)\\
&= (-\lambda)\nu(b) + \nu(\lambda a + c) + \lambda\nu(0)\\
&= \nu((-\lambda)b + \lambda a + c) + (-\lambda)\nu(0) + \lambda\nu(0)\\
&= \nu(\lambda(a-b) + c).
\end{aligned} \tag{75}$$

Combining this with (74) implies that for all $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} Y_t &= \lambda x + y + \int_0^t \mu\big(\lambda(X_s^x - X_s^0) + X_s^y\big)\, ds \\ &\quad + \int_0^t \sigma\big(\lambda(X_s^x - X_s^0) + X_s^y\big)\, dW_s \\ &= \lambda x + y + \int_0^t \mu(Y_s)\, ds + \int_0^t \sigma(Y_s)\, dW_s. \end{aligned} \tag{76}$$

The fact that for all $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$X_t^{\lambda x + y} = \lambda x + y + \int_0^t \mu(X_s^{\lambda x+y})\, ds + \int_0^t \sigma(X_s^{\lambda x+y})\, dW_s, \tag{77}$$

Corollary 2.9, Corollary 2.10, and, e.g., Da Prato & Zabczyk [19, Item (i) in Theorem 7.4] (cf., e.g., Klenke [46, Theorem 26.8]) hence demonstrate that for all $t \in [0, T]$ it holds that

$$\mathbb{P}\Big(X_t^{\lambda x+y} = Y_t\Big) = 1. \tag{78}$$

This and (73) imply that for all $t \in [0, T]$ it holds that

$$\begin{aligned} \mathbb{P}\Big(X_t^{\lambda x+y} + \lambda X_t^0 = \lambda X_t^x + X_t^y\Big) &= \mathbb{P}\Big(X_t^{\lambda x+y} = \lambda(X_t^x - X_t^0) + X_t^y\Big) \\ &= \mathbb{P}\Big(X_t^{\lambda x+y} = Y_t\Big) = 1. \end{aligned} \tag{79}$$

The proof of Lemma 2.15 is thus completed. □

Lemma 2.16 (Modifications of continuous random fields are indistinguishable)**.** *Let $d \in \mathbb{N}$, let (E, δ) be a separable metric space, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X, Y : E \times \Omega \to \mathbb{R}^d$ be random fields, assume for all $\omega \in \Omega$ that*

$$(E \ni e \mapsto X_e(\omega) \in \mathbb{R}^d),\ (E \ni e \mapsto Y_e(\omega) \in \mathbb{R}^d) \in C(E, \mathbb{R}^d), \tag{80}$$

and assume for all $e \in E$ that $\mathbb{P}(X_e = Y_e) = 1$. Then

(i) it holds that $\{\forall\, e \in E : X_e = Y_e\} \in \mathcal{F}$ and

(ii) it holds that $\mathbb{P}(\forall\, e \in E : X_e = Y_e) = 1$.

Proof of Lemma 2.16. Throughout this proof assume w.l.o.g. that $E \neq \emptyset$, let $(e_n)_{n \in \mathbb{N}} \subseteq E$ satisfy that

$$\overline{\{e_n \in E : n \in \mathbb{N}\}} = E, \tag{81}$$

and let $\mathcal{N} \subseteq \Omega$ satisfy that

$$\mathcal{N} = \cup_{n\in\mathbb{N}}\{X_{e_n} \neq Y_{e_n}\}. \tag{82}$$

Note that the fact that X and Y are random fields assures that for all $e \in E$ it holds that

$$\{X_e = Y_e\} = \{X_e - Y_e = 0\} \in \mathcal{F}. \tag{83}$$

Hence, we obtain that

$$(\cap_{n\in\mathbb{N}}\{X_{e_n} = Y_{e_n}\}) \in \mathcal{F}. \tag{84}$$

Combining this and (82) implies that

$$\mathcal{N} = [\Omega\backslash\,(\cap_{n\in\mathbb{N}}\{X_{e_n} = Y_{e_n}\})] \in \mathcal{F}. \tag{85}$$

Moreover, observe that the hypothesis that for all $e \in E$ it holds that $\mathbb{P}(X_e = Y_e) = 1$ ensures that for all $n \in \mathbb{N}$ it holds that $\mathbb{P}(X_{e_n} \neq Y_{e_n}) = 0$. Therefore, we obtain that

$$\mathbb{P}(\mathcal{N}) \leq \sum_{n=1}^{\infty} \mathbb{P}(X_{e_n} \neq Y_{e_n}) = 0. \tag{86}$$

Next note that (81) implies that for every $v \in E$ there exists a strictly increasing function $n_v \colon \mathbb{N} \to \mathbb{N}$ such that $\limsup_{k\to\infty} \delta(e_{n_v(k)}, v) = 0$. Combining this with (80) ensures that for every $v \in E$ there exists a strictly increasing function $n_v \colon \mathbb{N} \to \mathbb{N}$ such that for every $\omega \in \{\forall\, k \in \mathbb{N} \colon X_{e_k} = Y_{e_k}\}$ it holds that $\limsup_{k\to\infty} \delta(e_{n_v(k)}, v) = 0$ and

$$X_v(\omega) = \lim_{k\to\infty} X_{e_{n_v(k)}}(\omega) = \lim_{k\to\infty} Y_{e_{n_v(k)}}(\omega) = Y_v(\omega). \tag{87}$$

This and (84) demonstrate that

$$\begin{aligned} \{\forall\, e \in E \colon X_e = Y_e\} &= \{\forall\, n \in \mathbb{N} \colon X_{e_n} = Y_{e_n}\} \\ &= (\cap_{n\in\mathbb{N}}\{X_{e_n} = Y_{e_n}\}) \in \mathcal{F}. \end{aligned} \tag{88}$$

This proves item (i). Combining (85) and (86) hence implies that

$$\mathbb{P}(\forall\, e \in E \colon X_e = Y_e) = \mathbb{P}(\cap_{n\in\mathbb{N}}\{X_{e_n} = Y_{e_n}\}) = \mathbb{P}(\Omega\backslash\mathcal{N}) = 1 - \mathbb{P}(\mathcal{N}) = 1. \tag{89}$$

This establishes item (ii). The proof of Lemma 2.16 is thus completed. □

Proposition 2.17. *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ be a filtered probability space which fulfils the usual conditions, let $W\colon [0,T] \times \Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion, let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be functions which satisfy for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that*

$$\mu(\lambda x + y) + \lambda\mu(0) = \lambda\mu(x) + \mu(y) \tag{90}$$

and

$$\sigma(\lambda x + y) + \lambda\sigma(0) = \lambda\sigma(x) + \sigma(y), \tag{91}$$

let $X^x\colon [0,T] \times \Omega \to \mathbb{R}^d$, $x \in \mathbb{R}^d$, be $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic processes, assume for all $\omega \in \Omega$ that

$$\big(\mathbb{R}^d \times [0,T] \ni (x,t) \mapsto X_t^x(\omega) \in \mathbb{R}^d\big) \in C(\mathbb{R}^d \times [0,T], \mathbb{R}^d), \tag{92}$$

and assume that for all $x \in \mathbb{R}^d$, $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$X_t^x = x + \int_0^t \mu(X_s^x)\, ds + \int_0^t \sigma(X_s^x)\, dW_s. \tag{93}$$

Then

(i) it holds that

$$\Big\{\forall\, x, y \in \mathbb{R}^d, \lambda \in \mathbb{R}, t \in [0,T]\colon X_t^{\lambda x + y} + \lambda X_t^0 = \lambda X_t^x + X_t^y\Big\} \in \mathcal{F} \tag{94}$$

and

(ii) it holds that

$$\mathbb{P}\Big(\forall\, x, y \in \mathbb{R}^d, \lambda \in \mathbb{R}, t \in [0,T]\colon X_t^{\lambda x + y} + \lambda X_t^0 = \lambda X_t^x + X_t^y\Big) = 1. \tag{95}$$

Proof of Proposition 2.17. Throughout this proof let $Y, Z\colon (\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times [0,T]) \times \Omega \to \mathbb{R}^d$ be the random fields which satisfy for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$, $t \in [0,T]$ that

$$Y_{(x,y,\lambda,t)} = X_t^{\lambda x + y} + \lambda X_t^0 \qquad \text{and} \qquad Z_{(x,y,\lambda,t)} = \lambda X_t^x + X_t^y. \tag{96}$$

Observe that Lemma 2.15 assures that for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$, $t \in [0,T]$ it holds that

$$\mathbb{P}\big(Y_{(x,y,\lambda,t)} = Z_{(x,y,\lambda,t)}\big) = \mathbb{P}\big(X_t^{\lambda x + y} + \lambda X_t^0 = \lambda X_t^x + X_t^y\big) = 1. \tag{97}$$

Moreover, note that (92) and the fact that

$$\big(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times [0,T] \ni (x,y,\lambda,t) \mapsto (\lambda x + y, t) \in \mathbb{R}^d \times [0,T]\big) \\ \in C(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times [0,T], \mathbb{R}^d \times [0,T]) \tag{98}$$

demonstrate that for all $\omega \in \Omega$ it holds that

$$Y_{\cdot}(\omega), Z_{\cdot}(\omega) \in C(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times [0,T], \mathbb{R}^d). \tag{99}$$

Combining this, (97), and Lemma 2.16 (with $d = d$, $E = \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times [0,T]$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $X = Y$, $Y = Z$ in the notation of Lemma 2.16) proves that

$$\big\{\forall\, x, y \in \mathbb{R}^d, \lambda \in \mathbb{R}, t \in [0,T] \colon Y_{(x,y,\lambda,t)} = Z_{(x,y,\lambda,t)}\big\} \in \mathcal{F} \tag{100}$$

and

$$\mathbb{P}\big(\forall\, x, y \in \mathbb{R}^d, \lambda \in \mathbb{R}, t \in [0,T] \colon Y_{(x,y,\lambda,t)} = Z_{(x,y,\lambda,t)}\big) = 1. \tag{101}$$

This and (96) demonstrate that

$$\begin{split} &\Big\{\forall\, x, y \in \mathbb{R}^d, \lambda \in \mathbb{R}, t \in [0,T] \colon X_t^{\lambda x + y} + \lambda X_t^0 = \lambda X_t^x + X_t^y\Big\} \\ &= \big\{\forall\, x, y \in \mathbb{R}^d, \lambda \in \mathbb{R}, t \in [0,T] \colon Y_{(x,y,\lambda,t)} = Z_{(x,y,\lambda,t)}\big\} \in \mathcal{F} \end{split} \tag{102}$$

and

$$\begin{split} &\mathbb{P}\Big(\forall\, x, y \in \mathbb{R}^d, \lambda \in \mathbb{R}, t \in [0,T] \colon X_t^{\lambda x + y} + \lambda X_t^0 = \lambda X_t^x + X_t^y\Big) \\ &= \mathbb{P}\big(\forall\, x, y \in \mathbb{R}^d, \lambda \in \mathbb{R}, t \in [0,T] \colon Y_{(x,y,\lambda,t)} = Z_{(x,y,\lambda,t)}\big) = 1. \end{split} \tag{103}$$

This establishes items (i)–(ii). The proof of Proposition 2.17 is thus completed. □

Lemma 2.18 (Modifications of adapted processes are adapted)**.** *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ be a filtered probability space which fulfils the usual conditions, let $X, Y \colon [0,T] \times \Omega \to \mathbb{R}^d$ be stochastic processes, assume that X is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process, and assume for all $t \in [0,T]$ that $\mathbb{P}(X_t = Y_t) = 1$. Then it holds that Y is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process.*

Proof of Lemma 2.18. Throughout this proof let $t \in [0, T]$. Note that the hypothesis that $\mathbb{P}(X_t = Y_t) = 1$ ensures that

$$\mathbb{P}(X_t \neq Y_t) = 0. \tag{104}$$

This and the hypothesis that $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ is a filtered probability space which fulfils the usual conditions imply that $\{X_t \neq Y_t\} \in \mathbb{F}_0 \subseteq \mathbb{F}_t$. Hence, we obtain that

$$\{X_t = Y_t\} = \Omega \backslash \{X_t \neq Y_t\} \in \mathbb{F}_t. \tag{105}$$

Moreover, observe that (104) demonstrates that for all $B \in \mathcal{B}(\mathbb{R}^d)$ it holds that

$$\mathbb{P}(\{Y_t \in B\} \cap \{X_t \neq Y_t\}) \leq \mathbb{P}(X_t \neq Y_t) = 0. \tag{106}$$

The hypothesis that $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ is a filtered probability space which fulfils the usual conditions therefore implies that for all $B \in \mathcal{B}(\mathbb{R}^d)$ it holds that

$$(\{Y_t \in B\} \cap \{X_t \neq Y_t\}) \in \mathbb{F}_0 \subseteq \mathbb{F}_t. \tag{107}$$

Combining this with the hypothesis that X is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process and (105) demonstrates that for all $B \in \mathcal{B}(\mathbb{R}^d)$ it holds that

$$\begin{aligned} \{Y_t \in B\} &= (\{Y_t \in B\} \cap \{X_t = Y_t\}) \cup (\{Y_t \in B\} \cap \{X_t \neq Y_t\}) \\ &= (\{X_t \in B\} \cap \{X_t = Y_t\}) \cup (\{Y_t \in B\} \cap \{X_t \neq Y_t\}) \in \mathbb{F}_t. \end{aligned} \tag{108}$$

This establishes that Y is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process. The proof of Lemma 2.18 is thus completed. □

Lemma 2.19 (A version of the Kolmogorov-Chentsov theorem)**.** *Let $d, k \in \mathbb{N}$, $p \in (d, \infty)$, $\alpha \in ({}^{d}\!/_{p}, \infty)$, for every $\mathfrak{d} \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^\mathfrak{d}} : \mathbb{R}^\mathfrak{d} \to [0, \infty)$ be the $\mathfrak{d}$-dimensional Euclidean norm, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $D \subseteq \mathbb{R}^d$ be a non-empty set, and let $X\colon D \times \Omega \to \mathbb{R}^k$ be a random field which satisfies for all $n \in \mathbb{N}$ that*

$$\begin{aligned} \sup\Bigg(\Bigg\{ \frac{\big(\mathbb{E}\big[\|X_v - X_w\|_{\mathbb{R}^k}^p\big]\big)^{1/p}}{\|v - w\|_{\mathbb{R}^d}^\alpha} : v, w \in D \cap [-n, n]^d,\, v \neq w \Bigg\} \\ \cup \Big\{ \big(\mathbb{E}\big[\|X_v\|_{\mathbb{R}^k}^p\big]\big)^{1/p} : v \in D \cap [-n, n]^d \Big\} \cup \{0\} \Bigg) < \infty. \end{aligned} \tag{109}$$

Then there exists a random field $Y\colon D \times \Omega \to \mathbb{R}^k$ which satisfies *(i) that for all $\omega \in \Omega$ it holds that $(D \ni v \mapsto Y_v(\omega) \in \mathbb{R}^k) \in C(D, \mathbb{R}^k)$ and*

(ii) that for all $v \in D$ it holds that $\mathbb{P}(X_v = Y_v) = 1$.

Proof of Lemma 2.19. Throughout this proof let $g_n \colon D \cap [-n, n]^d \to L^p(\Omega; \mathbb{R}^k)$, $n \in \mathbb{N}$, be functions which satisfy that for all $n \in \mathbb{N}$, $v \in D \cap [-n, n]^d$ it holds $\mathbb{P}$-a.s. that $X_v = g_n(v)$ (cf. (109)), let $\mathfrak{c} \in [0, \infty)$ be a real number which satisfies that for all $n \in \mathbb{N}$, $v, w \in D \cap [-n, n]^d$ it holds that

$$\left(\mathbb{E}\big[\|X_v - X_w\|_{\mathbb{R}^k}^p\big]\right)^{1/p} \leq \mathfrak{c} \, \|v - w\|_{\mathbb{R}^d}^{\alpha} \tag{110}$$

(cf. (109)), let $\mathfrak{a} = \min\{\alpha, 1\}$, and let $(\cdot)^+ \colon \mathbb{R} \to [0, \infty)$ be the function which satisfies for all $q \in \mathbb{R}$ that $(q)^+ = \max\{q, 0\}$. Note that for all $n \in \mathbb{N}$, $v, w \in D \cap [-n, n]^d$ it holds that

$$\begin{aligned} \|v - w\|_{\mathbb{R}^d}^{(\alpha-1)^+} &\leq (\|v\|_{\mathbb{R}^d} + \|w\|_{\mathbb{R}^d})^{(\alpha-1)^+} \leq \left(n\sqrt{d} + n\sqrt{d}\right)^{(\alpha-1)^+} \\ &\leq \left(2n\sqrt{d}\right)^{(\alpha-1)^+}. \end{aligned} \tag{111}$$

Combining this and (110) with the fact that $\alpha - \mathfrak{a} = (\alpha - 1)^+$ ensures that for all $n \in \mathbb{N}$, $v, w \in D \cap [-n, n]^d$ it holds that

$$\begin{aligned} \left(\mathbb{E}\big[\|X_v - X_w\|_{\mathbb{R}^k}^p\big]\right)^{1/p} &\leq \mathfrak{c} \, \|v - w\|_{\mathbb{R}^d}^{\alpha} \leq \mathfrak{c} \, \|v - w\|_{\mathbb{R}^d}^{\mathfrak{a}} \, \|v - w\|_{\mathbb{R}^d}^{\alpha - \mathfrak{a}} \\ &= \mathfrak{c} \, \|v - w\|_{\mathbb{R}^d}^{\mathfrak{a}} \, \|v - w\|_{\mathbb{R}^d}^{(\alpha-1)^+} \\ &\leq \mathfrak{c} \, \|v - w\|_{\mathbb{R}^d}^{\mathfrak{a}} \left(2n\sqrt{d}\right)^{(\alpha-1)^+}. \end{aligned} \tag{112}$$

This and (109) imply that for all $n \in \mathbb{N}$ it holds that

$$\begin{aligned} \sup\Bigg(&\left\{ \frac{\left(\mathbb{E}\big[\|X_v - X_w\|_{\mathbb{R}^k}^p\big]\right)^{1/p}}{\|v - w\|_{\mathbb{R}^d}^{\mathfrak{a}}} \colon v, w \in D \cap [-n, n]^d, \, v \neq w \right\} \\ &\cup \left\{ \left(\mathbb{E}\big[\|X_v\|_{\mathbb{R}^k}^p\big]\right)^{1/p} : v \in D \cap [-n, n]^d \right\} \cup \{0\} \Bigg) < \infty. \end{aligned} \tag{113}$$

Therefore, we obtain that for all $n \in \mathbb{N}$ it holds that g_n is a globally bounded and globally $\mathfrak{a}$-Hölder continuous function. Mittmann & Steinwart [55, Theorem 2.2] hence ensures that for every $n \in \mathbb{N}$ there is a globally bounded and globally $\mathfrak{a}$-Hölder continuous function $G_n \colon \mathbb{R}^d \to L^p(\Omega; \mathbb{R}^k)$ which satisfies for all $v \in D \cap [-n, n]^d$ that $G_n(v) = g_n(v)$. This assures that there exist random fields $\xi_n \colon \mathbb{R}^d \times \Omega \to \mathbb{R}^k$, $n \in \mathbb{N}$, which satisfy

(a) that for all $n \in \mathbb{N}$, $v \in \mathbb{R}^d$ it holds $\mathbb{P}$-a.s. that $(\xi_n)_v = G_n(v)$ and

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

$$\begin{aligned} &\sup_{\substack{v,w\in[-n,n+1)^d,\\ v\neq w}} \frac{\big(\mathbb{E}\big[\|(\xi_n)_v - (\xi_n)_w\|_{\mathbb{R}^k}^p\big]\big)^{1/p}}{\|v-w\|_{\mathbb{R}^d}^{\mathfrak{a}}} \\ &\le \sup_{\substack{v,w\in\mathbb{R}^d,\\ v\neq w}} \frac{\big(\mathbb{E}\big[\|(\xi_n)_v - (\xi_n)_w\|_{\mathbb{R}^k}^p\big]\big)^{1/p}}{\|v-w\|_{\mathbb{R}^d}^{\mathfrak{a}}} < \infty \end{aligned} \tag{114}$$

and
$$\sup_{v\in[-n,n+1)^d} \big(\mathbb{E}\big[\|(\xi_n)_v\|_{\mathbb{R}^k}^p\big]\big)^{1/p} \le \sup_{v\in\mathbb{R}^d} \big(\mathbb{E}\big[\|(\xi_n)_v\|_{\mathbb{R}^k}^p\big]\big)^{1/p} < \infty. \tag{115}$$

Combining this and, e.g., Revuz & Yor [65, Theorem 2.1 in Section 2 in Chapter I] (with $X = \xi_n$, $\gamma = p$, $d = d$, $\varepsilon = \mathfrak{a}p - d$ in the notation of [65, Theorem 2.1 in Section 2 in Chapter I]) ensures that there exist random fields $Y_n \colon [-n,n]^d \times \Omega \to \mathbb{R}^k$, $n \in \mathbb{N}$, which satisfy

(A) that for all $n \in \mathbb{N}$, $\omega \in \Omega$ it holds that

$$\big([-n,n]^d \ni v \mapsto (Y_n)_v(\omega) \in \mathbb{R}^k\big) \in C([-n,n]^d, \mathbb{R}^k) \tag{116}$$

and

(B) that for all $n \in \mathbb{N}$, $v \in [-n,n]^d$ it holds that $\mathbb{P}((Y_n)_v = (\xi_n)_v) = 1$.

The fact that for all $n \in \mathbb{N}$, $v \in D \cap [-n,n]^d$ it holds that $\mathbb{P}(X_v = (\xi_n)_v) = 1$ therefore implies that for all $n \in \mathbb{N}$, $v \in D \cap [-n,n]^d$ it holds that

$$\mathbb{P}((Y_n)_v = X_v) = 1. \tag{117}$$

This assures that for all $n \in \mathbb{N}, m \in \mathbb{N} \cap [1,n], v \in D \cap [-m,m]^d$ it holds that

$$\mathbb{P}(\{(Y_n)_v = X_v\} \cap \{(Y_m)_v = X_v\}) = 1. \tag{118}$$

The fact that for all $n \in \mathbb{N}, m \in \mathbb{N} \cap [1,n], v \in D \cap [-m,m]^d$ it holds that

$$\{(Y_n)_v = X_v\} \cap \{(Y_m)_v = X_v\} \subseteq \{(Y_n)_v = (Y_m)_v\} \tag{119}$$

therefore demonstrates that for all $n \in \mathbb{N}, m \in \mathbb{N} \cap [1, n], v \in D \cap [-m, m]^d$ it holds that

$$\mathbb{P}((Y_n)_v = (Y_m)_v) = 1. \tag{120}$$

Combining this with (116) and Lemma 2.16 (with $d = k$, $E = D \cap [-m, m]^d$ for $m \in \mathbb{N} \cap [1, n]$, $n \in \mathbb{N}$ in the notation of Lemma 2.16) establishes that for all $n \in \mathbb{N}$, $m \in \mathbb{N} \cap [1, n]$ it holds that

$$\mathbb{P}\big(\forall\, v \in D \cap [-m, m]^d \colon (Y_n)_v = (Y_m)_v\big) = 1. \tag{121}$$

Next let $\Pi \in \mathcal{F}$ be the event given by

$$\Pi = \big\{\forall\, n \in \mathbb{N},\, m \in \mathbb{N} \cap [1, n],\, v \in D \cap [-m, m]^d \colon (Y_n)_v = (Y_m)_v\big\}. \tag{122}$$

Observe that (122) and (121) show that

$$\mathbb{P}(\Pi) = \mathbb{P}\big(\cap_{n=1}^{\infty} \cap_{m=1}^{n} \big\{\forall\, v \in D \cap [-m, m]^d \colon (Y_n)_v = (Y_m)_v\big\}\big) = 1. \tag{123}$$

Moreover, note that (122) ensures that there exists a unique random field $Z \colon D \times \Omega \to \mathbb{R}^k$ which satisfies

(I) that for all $\omega \in \Pi$, $n \in \mathbb{N}$, $v \in D \cap [-n, n]^d$ it holds that $Z_v(\omega) = (Y_n)_v(\omega)$ and

(II) that for $\omega \in \Omega \backslash \Pi$, $v \in D$ it holds that $Z_v(\omega) = 0$.

Observe that (I), (117), and (123) demonstrate that for all $n \in \mathbb{N}$, $v \in D \cap [-n, n]^d$ it holds that

$$\begin{aligned} \mathbb{P}(Z_v = X_v) &= \mathbb{P}(\{Z_v = X_v\} \cap \Pi) \\ &= \mathbb{P}(\{(Y_n)_v = X_v\} \cap \Pi) = \mathbb{P}((Y_n)_v = X_v) = 1. \end{aligned} \tag{124}$$

This shows that for all $v \in D$ it holds that

$$\mathbb{P}(Z_v = X_v) = 1. \tag{125}$$

Moreover, observe that (116) and (I) imply that for all $\omega \in \Pi$, $n \in \mathbb{N}$ it holds that

$$\begin{aligned} &\big(D \cap [-n, n]^d \ni v \mapsto Z_v(\omega) \in \mathbb{R}^k\big) \\ &= \big(D \cap [-n, n]^d \ni v \mapsto (Y_n)_v(\omega) \in \mathbb{R}^k\big) \in C(D \cap [-n, n]^d, \mathbb{R}^k). \end{aligned} \tag{126}$$

This ensures that for all $\omega \in \Pi$ it holds that

$$\big(D \ni v \mapsto Z_v(\omega) \in \mathbb{R}^k\big) \in C(D, \mathbb{R}^k). \tag{127}$$

In addition, note that (II) assures that for all $\omega \in \Omega \backslash \Pi$ it holds that

$$\big(D \ni v \mapsto Z_v(\omega) \in \mathbb{R}^k\big) = \big(D \ni v \mapsto 0 \in \mathbb{R}^k\big) \in C(D, \mathbb{R}^k). \tag{128}$$

Combining this and (127) demonstrates that for all $\omega \in \Omega$ it holds that

$$\big(D \ni v \mapsto Z_v(\omega) \in \mathbb{R}^k\big) \in C(D, \mathbb{R}^k). \tag{129}$$

This and (125) complete the proof of Lemma 2.19. □

Proposition 2.20. *Let $d \in \mathbb{N}$, $T \in (0, \infty)$, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ be a filtered probability space which fulfils the usual conditions, let $W\colon [0,T] \times \Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion, and let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be functions which satisfy for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that*

$$\mu(\lambda x + y) + \lambda\mu(0) = \lambda\mu(x) + \mu(y) \tag{130}$$

and

$$\sigma(\lambda x + y) + \lambda\sigma(0) = \lambda\sigma(x) + \sigma(y). \tag{131}$$

Then there exist up to indistinguishability unique $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic processes with continuous sample paths $X^x\colon [0,T] \times \Omega \to \mathbb{R}^d$, $x \in \mathbb{R}^d$, which satisfy

(i) that for all $x \in \mathbb{R}^d$, $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$X_t^x = x + \int_0^t \mu(X_s^x)\, ds + \int_0^t \sigma(X_s^x)\, dW_s \tag{132}$$

and

(ii) that for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$, $t \in [0,T]$, $\omega \in \Omega$ it holds that

$$X_t^{\lambda x+y}(\omega) + \lambda X_t^0(\omega) = \lambda X_t^x(\omega) + X_t^y(\omega). \tag{133}$$

Proof of Proposition 2.20. Throughout this proof let $p \in (2(d+1), \infty)$, for every $k \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^k} : \mathbb{R}^k \to [0, \infty)$ be the k-dimensional Euclidean norm, let $|||\cdot|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \colon \mathbb{R}^{d\times d} \to [0, \infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times d}$, let $\mathfrak{m}, \mathfrak{s} \in (0, \infty)$ satisfy for all $x \in \mathbb{R}^d$ that

$$\|\mu(x)\|_{\mathbb{R}^d} \leq \mathfrak{m}(1 + \|x\|_{\mathbb{R}^d}) \qquad \text{and} \qquad \|\mu(x) - \mu(y)\|_{\mathbb{R}^d} \leq \mathfrak{m}\,\|x - y\|_{\mathbb{R}^d} \tag{134}$$

and

$$|||\sigma(x)|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \leq \mathfrak{s}(1 + \|x\|_{\mathbb{R}^d}) \quad \text{and} \quad |||\sigma(x) - \sigma(y)|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \leq \mathfrak{s}\,\|x - y\|_{\mathbb{R}^d} \tag{135}$$

(cf. Corollary 2.9 and Corollary 2.10), and let $C \in (0, \infty)$ be given by

$$C = 4d\sqrt{2}\big(1 + \mathfrak{m}T + \mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T} + \mathfrak{s}p\big]^2 T\Big). \tag{136}$$

Note that (134), (135), e.g., Jentzen & Kloeden [42, Theorem 5.1] (with $T = T$, $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]}) = (\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$, $H = \mathbb{R}^d$, $\|\cdot\|_H = \|\cdot\|_{\mathbb{R}^d}$, $U = \mathbb{R}^d$, $\|\cdot\|_U = \|\cdot\|_{\mathbb{R}^d}$, $Qv = v$, $(W_t)_{t\in[0,T]} = (W_t)_{t\in[0,T]}$, $D(A) = \mathbb{R}^d$, $Av = 0$, $\eta = 1$, $\alpha = 0$, $\delta = 0$, $F(v) = \mu(v)$, $\beta = 0$, $B(v)u = \sigma(v)u$, $\gamma = 0$, $p = 4$, $\xi = (\Omega \ni \omega \mapsto x \in \mathbb{R}^d)$ for $u, v, x \in \mathbb{R}^d$ in the notation of [42, Theorem 5.1]) (cf., e.g., Da Prato & Zabczyk [19, Item (i) in Theorem 7.4] and Klenke [46, Theorem 26.8]), and, e.g., the Kolmogorov-Chentsov type theorem in Lemma 2.19 (with $d = 1$, $k = d$, $p = 4$, $\alpha = 1/2$, $D = [0, T]$ in the notation of Lemma 2.19) assure that there exist $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic processes with continuous sample paths $X^x \colon [0, T] \times \Omega \to \mathbb{R}^d$, $x \in \mathbb{R}^d$, which satisfy that for all $x \in \mathbb{R}^d$, $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$X^x_t = x + \int_0^t \mu(X^x_s)\,ds + \int_0^t \sigma(X^x_s)\,dW_s. \tag{137}$$

Observe that (137) and Lemma 2.15 prove that for all $t \in [0, T]$, $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ it holds $\mathbb{P}$-a.s. that

$$X^{\lambda x + y}_t + \lambda X^0_t = \lambda X^x_t + X^y_t. \tag{138}$$

This implies that for all $t \in [0, T]$, $v \in \mathbb{R}^d \backslash \{0\}$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} X^v_t = X_t^{\|v\|_{\mathbb{R}^d}\frac{v}{\|v\|_{\mathbb{R}^d}}} &= \Big(X_t^{\|v\|_{\mathbb{R}^d}\frac{v}{\|v\|_{\mathbb{R}^d}}} + \|v\|_{\mathbb{R}^d} X^0_t\Big) - \|v\|_{\mathbb{R}^d} X^0_t \\ &= \Big(\|v\|_{\mathbb{R}^d} X_t^{\frac{v}{\|v\|_{\mathbb{R}^d}}} + X^0_t\Big) - \|v\|_{\mathbb{R}^d} X^0_t = \|v\|_{\mathbb{R}^d}\Big(X_t^{\frac{v}{\|v\|_{\mathbb{R}^d}}} - X^0_t\Big) + X^0_t. \end{aligned} \tag{139}$$

Combining this and (138) (with $t = t$, $x = y$, $y = x$, $\lambda = -1$ for $t \in [0,T]$, $x, y \in \mathbb{R}^d$ in the notation of (138)) implies that for all $t \in [0,T]$, $x, y \in \mathbb{R}^d$ with $x \neq y$ it holds that

$$\begin{aligned}
\big(\mathbb{E}\big[\|X_t^x - X_t^y\|_{\mathbb{R}^d}^p\big]\big)^{1/p} &= \big(\mathbb{E}\big[\|X_t^{x-y} - X_t^0\|_{\mathbb{R}^d}^p\big]\big)^{1/p} \\
&= \bigg(\mathbb{E}\Big[\Big\|\|x-y\|_{\mathbb{R}^d}\big(X_t^{\frac{x-y}{\|x-y\|_{\mathbb{R}^d}}} - X_t^0\big) + X_t^0 - X_t^0\Big\|_{\mathbb{R}^d}^p\Big]\bigg)^{1/p} \\
&= \bigg(\mathbb{E}\Big[\Big\|X_t^{\frac{x-y}{\|x-y\|_{\mathbb{R}^d}}} - X_t^0\Big\|_{\mathbb{R}^d}^p\Big]\bigg)^{1/p} \|x-y\|_{\mathbb{R}^d}.
\end{aligned} \tag{140}$$

In addition, observe that (134), (135), (137), Proposition 2.14 (with $d = d$, $m = d$, $p = p$, $T = T$, $\mathfrak{m}_1 = \mathfrak{m}$, $\mathfrak{m}_2 = \mathfrak{m}$, $\mathfrak{s}_1 = \mathfrak{s}$, $\mathfrak{s}_2 = \mathfrak{s}$ in the notation of Proposition 2.14), and the triangle inequality assure that for all $t \in [0,T]$ it holds that

$$\begin{aligned}
&\sup_{v \in \mathbb{R}^d, \|v\|_{\mathbb{R}^d}=1} \big(\mathbb{E}\big[\|X_t^v - X_t^0\|_{\mathbb{R}^d}^p\big]\big)^{1/p} \\
&\leq \big(\mathbb{E}\big[\|X_t^0\|_{\mathbb{R}^d}^p\big]\big)^{1/p} + \sup_{v \in \mathbb{R}^d, \|v\|_{\mathbb{R}^d}=1} \big(\mathbb{E}\big[\|X_t^v\|_{\mathbb{R}^d}^p\big]\big)^{1/p} \\
&\leq \sqrt{2}\big(\mathfrak{m}T + \mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T} + \mathfrak{s}p\big]^2 T\Big) \\
&\quad + \sup_{v \in \mathbb{R}^d, \|v\|_{\mathbb{R}^d}=1} \Big[\sqrt{2}\big(\|v\| + \mathfrak{m}T + \mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T} + \mathfrak{s}p\big]^2 T\Big)\Big] \\
&\leq 2\sqrt{2}\big(1 + \mathfrak{m}T + \mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T} + \mathfrak{s}p\big]^2 T\Big).
\end{aligned} \tag{141}$$

This, (136), (140), and the fact that for all $n \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d), y = (y_1, y_2, \ldots, y_d) \in [-n,n]^d$ it holds that

$$\begin{aligned}
\|x-y\|_{\mathbb{R}^d} &= \|x-y\|_{\mathbb{R}^d}^{1/2}\|x-y\|_{\mathbb{R}^d}^{1/2} \\
&= \big[[|x_1-y_1|^2 + \ldots + |x_d-y_d|^2]^{1/2}\big]^{1/2}\|x-y\|_{\mathbb{R}^d}^{1/2} \\
&\leq \big[[d(2n)^2]^{1/2}\big]^{1/2}\|x-y\|_{\mathbb{R}^d}^{1/2} \\
&= \big[2n\sqrt{d}\big]^{1/2}\|x-y\|_{\mathbb{R}^d}^{1/2} \leq 2dn\|x-y\|_{\mathbb{R}^d}^{1/2}
\end{aligned} \tag{142}$$

assure that for all $t \in [0,T]$, $n \in \mathbb{N}$, $x,y \in [-n,n]^d$ with $x \neq y$ it holds that

$$
\begin{aligned}
&\big(\mathbb{E}\big[\|X_t^x - X_t^y\|_{\mathbb{R}^d}^p\big]\big)^{1/p} = \Big(\mathbb{E}\Big[\Big\|X_t^{\frac{x-y}{\|x-y\|_{\mathbb{R}^d}}} - X_t^0\Big\|_{\mathbb{R}^d}^p\Big]\Big)^{1/p} \|x-y\|_{\mathbb{R}^d} \\
&\leq \left[\sup_{v \in \mathbb{R}^d, \|v\|_{\mathbb{R}^d}=1} \big(\mathbb{E}\big[\|X_t^v - X_t^0\|_{\mathbb{R}^d}^p\big]\big)^{1/p}\right] \|x-y\|_{\mathbb{R}^d} \\
&\leq 2\sqrt{2}\big(1 + \mathfrak{m}T + \mathfrak{s}p\sqrt{T}\big) \exp\Big(\big[\mathfrak{m}\sqrt{T} + \mathfrak{s}p\big]^2 T\Big) \|x-y\|_{\mathbb{R}^d} \\
&\leq 2\sqrt{2}\big(1 + \mathfrak{m}T + \mathfrak{s}p\sqrt{T}\big) \exp\Big(\big[\mathfrak{m}\sqrt{T} + \mathfrak{s}p\big]^2 T\Big) 2dn \|x-y\|_{\mathbb{R}^d}^{1/2} \\
&= nC\|x-y\|_{\mathbb{R}^d}^{1/2}.
\end{aligned}
\tag{143}
$$

Moreover, note that (137) and the triangle inequality imply that for all $x \in \mathbb{R}^d$, $s,t \in [0,T]$ with $s \leq t$ it holds that

$$
\begin{aligned}
&\big(\mathbb{E}\big[\|X_t^x - X_s^x\|_{\mathbb{R}^d}^p\big]\big)^{1/p} \\
&= \Big(\mathbb{E}\Big[\Big\|x + \textstyle\int_0^t \mu(X_u^x)\,du + \int_0^t \sigma(X_u^x)\,dW_u \\
&\qquad - \big(x + \textstyle\int_0^s \mu(X_u^x)\,du + \int_0^s \sigma(X_u^x)\,dW_u\big)\Big\|_{\mathbb{R}^d}^p\Big]\Big)^{1/p} \\
&= \Big(\mathbb{E}\big[\|\textstyle\int_s^t \mu(X_u^x)\,du + \int_s^t \sigma(X_u^x)\,dW_u\|_{\mathbb{R}^d}^p\big]\Big)^{1/p} \\
&\leq \Big(\mathbb{E}\big[\|\textstyle\int_s^t \mu(X_u^x)\,du\|_{\mathbb{R}^d}^p\big]\Big)^{1/p} + \Big(\mathbb{E}\big[\|\textstyle\int_s^t \sigma(X_u^x)\,dW_u\|_{\mathbb{R}^d}^p\big]\Big)^{1/p}.
\end{aligned}
\tag{144}
$$

Furthermore, observe that Lemma 2.12, Proposition 2.14, (134), and the fact that for all $s,t \in [0,T]$ it holds that $|t-s| = |t-s|^{1/2}|t-s|^{1/2} \leq \sqrt{T}|t-s|^{1/2}$

ensure that for all $n \in \mathbb{N}$, $x \in [-n,n]^d$, $s,t \in [0,T]$ with $s \leq t$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left\|\int_s^t \mu(X_u^x)\,du\right\|_{\mathbb{R}^d}^p\right]\right)^{1/p} \leq \left(\mathbb{E}\left[\left|\int_s^t \|\mu(X_u^x)\|_{\mathbb{R}^d}\,du\right|^p\right]\right)^{1/p} \\
&\leq \int_s^t \big(\mathbb{E}\big[\|\mu(X_u^x)\|_{\mathbb{R}^d}^p\big]\big)^{1/p}\,du \leq \int_s^t \mathfrak{m}\big(1+\big(\mathbb{E}\big[\|X_u^x\|_{\mathbb{R}^d}^p\big]\big)^{1/p}\big)\,du \\
&\leq \int_s^t \mathfrak{m}\Big[1+\sqrt{2}\big(\|x\|+\mathfrak{m}T+\mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T}+\mathfrak{s}p\big]^2 T\Big)\Big]\,du \\
&= \mathfrak{m}\Big[1+\sqrt{2}\big(n\sqrt{d}+\mathfrak{m}T+\mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T}+\mathfrak{s}p\big]^2 T\Big)\Big](t-s) \\
&\leq \mathfrak{m}(1+\sqrt{2})nd\big(1+\mathfrak{m}T+\mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T}+\mathfrak{s}p\big]^2 T\Big)|t-s| \\
&= \left(\frac{1+\sqrt{2}}{4\sqrt{2}}\right)\mathfrak{m}nC|t-s| \leq \mathfrak{m}nC\sqrt{T}|t-s|^{1/2}.
\end{aligned} \tag{145}$$

Moreover, note that the fact that for all $x \in \mathbb{R}^d$ it holds that $X^x \colon [0,T]\times\Omega \to \mathbb{R}^d$ is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths and (135) ensure that it holds $\mathbb{P}$-a.s. that

$$\begin{aligned}
&\int_0^T \|\!|\sigma(X_s^x)|\!\|_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)}^2\,ds \leq T\left[\sup_{s\in[0,T]} \|\!|\sigma(X_s^x)|\!\|_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)}^2\right] \\
&\leq T\left[\sup_{s\in[0,T]}\big[\mathfrak{m}^2(1+\|X_s^x\|_{\mathbb{R}^d})^2\big]\right] \leq 2T\mathfrak{m}^2\left(1+\sup_{s\in[0,T]}\|X_s^x\|_{\mathbb{R}^d}^2\right) < \infty.
\end{aligned} \tag{146}$$

Lemma 2.13, Proposition 2.14, the fact that for all $x \in \mathbb{R}^d$ it holds that $X^x \colon [0,T]\times\Omega \to \mathbb{R}^d$ is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths, and (135) hence demonstrate that for all $n \in \mathbb{N}$,

$x \in [-n,n]^d$, $s,t \in [0,T]$ with $s \le t$ it holds that

$$
\begin{aligned}
&\Big(\mathbb{E}\big[\| \textstyle\int_s^t \sigma(X_u^x)\, dW_u\|_{\mathbb{R}^d}^p\big]\Big)^{1/p} \\
&\le \left(\frac{p(p-1)}{2}\right)^{1/2} \left(\int_s^t \big(\mathbb{E}\big[|||\sigma(X_u^x)|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)}^p\big]\big)^{2/p}\, du\right)^{1/2} \\
&\le p\bigg(\int_s^t \Big[\big(\mathbb{E}\big[|||\sigma(X_u^x)|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)}^p\big]\big)^{1/p}\Big]^2\, du\bigg)^{1/2} \\
&\le p\bigg(\int_s^t \Big[\mathfrak{s}\left(1 + \big(\mathbb{E}\big[\|X_u^x\|_{\mathbb{R}^d}^p\big]\big)^{1/p}\right)\Big]^2\, du\bigg)^{1/2} \\
&\le p\mathfrak{s}\bigg[\left(\int_s^t 1^2\, du\right)^{1/2} + \left(\int_s^t \Big[\big(\mathbb{E}\big[\|X_u^x\|_{\mathbb{R}^d}^p\big]\big)^{1/p}\Big]^2\, du\right)^{1/2}\bigg] \\
&\le p\mathfrak{s}\bigg[|t-s|^{1/2} \\
&\quad + \bigg(\int_s^t \Big[\sqrt{2}\big(\|x\| + \mathfrak{m}T + \mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T} + \mathfrak{s}p\big]^2 T\Big)\Big]^2\, du\bigg)^{1/2}\bigg] \\
&\le p\mathfrak{s}|t-s|^{1/2}\bigg[1 + \sqrt{2}\big(n\sqrt{d} + \mathfrak{m}T + \mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T} + \mathfrak{s}p\big]^2 T\Big)\bigg] \\
&\le p\mathfrak{s}|t-s|^{1/2}(1+\sqrt{2})n\sqrt{d}\big(1 + \mathfrak{m}T + \mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T} + \mathfrak{s}p\big]^2 T\Big) \\
&= p\mathfrak{s}nC|t-s|^{1/2}\left(\frac{1+\sqrt{2}}{4\sqrt{2d}}\right) \le p\mathfrak{s}nC|t-s|^{1/2}.
\end{aligned}
\tag{147}
$$

Combining this with (144) and (145) establishes that for all $n \in \mathbb{N}$, $x \in [-n,n]^d$, $s,t \in [0,T]$ it holds that

$$
\begin{aligned}
\big(\mathbb{E}\big[\|X_t^x - X_s^x\|_{\mathbb{R}^d}^p\big]\big)^{1/p} &\le \mathfrak{m}nC\sqrt{T}|t-s|^{1/2} + p\mathfrak{s}nC|t-s|^{1/2} \\
&= nC(\mathfrak{m}\sqrt{T} + p\mathfrak{s})|t-s|^{1/2}.
\end{aligned}
\tag{148}
$$

Moreover, observe that the fact that for all $a,b \in [0,\infty)$ it holds that $a+b \le \sqrt{2}(a^2+b^2)^{1/2}$ ensures that for all $a,b \in [0,\infty)$ it holds that

$$
\sqrt{a} + \sqrt{b} \le \sqrt{2}(a+b)^{1/2} \le \sqrt{2}\big(\sqrt{2}(a^2+b^2)^{1/2}\big)^{1/2} \le 2\big((a^2+b^2)^{1/2}\big)^{1/2}. \tag{149}
$$

This, (143), and (148) demonstrate that for all $n \in \mathbb{N}$, $x,y \in [-n,n]^d$,

$s,t \in [0,T]$ it holds that

$$\begin{aligned}
&\big(\mathbb{E}\big[\|X_t^x - X_s^y\|_{\mathbb{R}^d}^p\big]\big)^{1/p} \\
&\le \big(\mathbb{E}\big[\|X_t^x - X_t^y\|_{\mathbb{R}^d}^p\big]\big)^{1/p} + \big(\mathbb{E}\big[\|X_t^y - X_s^y\|_{\mathbb{R}^d}^p\big]\big)^{1/p} \\
&\le nC\|x-y\|_{\mathbb{R}^d}^{1/2} + nC(\mathfrak{m}\sqrt{T} + p\mathfrak{s})|t-s|^{1/2} \\
&\le nC(1+\mathfrak{m}\sqrt{T} + p\mathfrak{s})(\|x-y\|_{\mathbb{R}^d}^{1/2} + |t-s|^{1/2}) \\
&\le nC(1+\mathfrak{m}\sqrt{T} + p\mathfrak{s})2\left[(\|x-y\|_{\mathbb{R}^d}^2 + |t-s|^2)^{1/2}\right]^{1/2} \\
&= 2nC(1+\mathfrak{m}\sqrt{T} + p\mathfrak{s})\,\|(x,t)-(y,s)\|_{\mathbb{R}^{d+1}}^{1/2}.
\end{aligned} \tag{150}$$

Hence, we obtain for all $n \in \mathbb{N}$ that

$$\begin{aligned}
&\sup_{\substack{(x,t),(y,s)\in(\mathbb{R}^d\times[0,T])\cap[-n,n]^{d+1},\\ (x,t)\neq(y,s)}} \left(\frac{\big(\mathbb{E}\big[\|X_t^x - X_s^y\|_{\mathbb{R}^d}^p\big]\big)^{1/p}}{\|(x,t)-(y,s)\|_{\mathbb{R}^{d+1}}^{1/2}}\right) \\
&\le 2nC(1+\mathfrak{m}\sqrt{T}+p\mathfrak{s}) < \infty.
\end{aligned} \tag{151}$$

In addition, note that (134), (135), and Proposition 2.14 assure that for all $n \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\sup_{(x,t)\in(\mathbb{R}^d\times[0,T])\cap[-n,n]^{d+1}} \left[\big(\mathbb{E}\big[\|X_t^x\|_{\mathbb{R}^d}^p\big]\big)^{1/p}\right] \\
&\le \sup_{(x,t)\in(\mathbb{R}^d\times[0,T])\cap[-n,n]^{d+1}} \left[\sqrt{2}\big(\|x\| + \mathfrak{m}T + \mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T}+\mathfrak{s}p\big]^2 T\Big)\right] \\
&\le \sqrt{2}\big(n\sqrt{d} + \mathfrak{m}T + \mathfrak{s}p\sqrt{T}\big)\exp\Big(\big[\mathfrak{m}\sqrt{T}+\mathfrak{s}p\big]^2 T\Big) < \infty.
\end{aligned} \tag{152}$$

Lemma 2.19 (with $d = d+1$, $k = d$, $p = p$, $\alpha = 1/2$, $(\Omega,\mathcal{F},\mathbb{P}) = (\Omega,\mathcal{F},\mathbb{P})$, $D = \mathbb{R}^d \times [0,T]$, $X = ((\mathbb{R}^d\times[0,T])\times\Omega \ni ((x,t),\omega) \mapsto X_t^x(\omega) \in \mathbb{R}^d)$ in the notation of Lemma 2.19) and (151) hence prove that there exist stochastic processes $Y^x\colon [0,T]\times\Omega \to \mathbb{R}^d$, $x \in \mathbb{R}^d$, which satisfy

(I) that for all $\omega \in \Omega$ it holds that

$$\big(\mathbb{R}^d\times[0,T] \ni (x,t) \mapsto Y_t^x(\omega) \in \mathbb{R}^d\big) \in C(\mathbb{R}^d\times[0,T],\mathbb{R}^d) \tag{153}$$

and

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

$$\mathbb{P}(Y_t^x = X_t^x) = 1. \tag{154}$$

The fact that for all $x \in \mathbb{R}^d$ it holds that X^x is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process and Lemma 2.18 therefore ensure that for all $x \in \mathbb{R}^d$ it holds that Y^x is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process. Next note that (153), (154), and the fact that for all $x \in \mathbb{R}^d$ it holds that X^x has continuous sample paths assure that for all $x \in \mathbb{R}^d$ it holds that

$$\mathbb{P}\big(\forall\, t \in [0,T] \colon Y_t^x = X_t^x\big) = \mathbb{P}\big(\forall\, t \in [0,T] \cap \mathbb{Q} \colon Y_t^x = X_t^x\big) = 1. \tag{155}$$

Moreover, observe that (154) implies that for all $x \in \mathbb{R}^d$, $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$\int_0^t \mu(X_s^x)\, ds = \int_0^t \mu(Y_s^x)\, ds \qquad \text{and} \qquad \int_0^t \sigma(X_s^x)\, dW_s = \int_0^t \sigma(Y_s^x)\, dW_s. \tag{156}$$

Combining this, (154), and (137) ensures that for all $x \in \mathbb{R}^d$, $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} Y_t^x = X_t^x &= x + \int_0^t \mu(X_s^x)\, ds + \int_0^t \sigma(X_s^x)\, dW_s \\ &= x + \int_0^t \mu(Y_s^x)\, ds + \int_0^t \sigma(Y_s^x)\, dW_s. \end{aligned} \tag{157}$$

Next let $\Pi \subseteq \Omega$ be the set given by

$$\Pi = \Big\{\forall\, x, y \in \mathbb{R}^d, \lambda \in \mathbb{R}, t \in [0,T] \colon\ Y_t^{\lambda x + y} + \lambda Y_t^0 = \lambda Y_t^x + Y_t^y\Big\}. \tag{158}$$

Combining (130), (131), (153) (157), and (158) with Proposition 2.17 demonstrates that

$$\Pi \in \mathcal{F} \qquad \text{and} \qquad \mathbb{P}(\Pi) = 1. \tag{159}$$

This proves that there exist unique stochastic processes $Z^x \colon [0,T] \times \Omega \to \mathbb{R}^d$, $x \in \mathbb{R}^d$, which satisfy for all $x \in \mathbb{R}^d$, $t \in [0,T]$, $\omega \in \Omega$ that

$$Z_t^x(\omega) = \begin{cases} Y_t^x(\omega) \colon & \omega \in \Pi \\ 0 \qquad : & \omega \in \Omega \backslash \Pi. \end{cases} \tag{160}$$

Observe that (158) and (160) imply that for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$, $t \in [0,T]$, $\omega \in \Pi$ it holds that

$$\begin{aligned} Z_t^{\lambda x+y}(\omega) + \lambda Z_t^0(\omega) &= Y_t^{\lambda x+y}(\omega) + \lambda Y_t^0(\omega) \\ &= \lambda Y_t^x(\omega) + Y_t^y(\omega) = \lambda Z_t^x(\omega) + Z_t^y(\omega). \end{aligned} \tag{161}$$

Moreover, note that (160) assures that for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$, $t \in [0,T]$, $\omega \in \Omega \backslash \Pi$ it holds that

$$Z_t^{\lambda x+y}(\omega) + \lambda Z_t^0(\omega) = 0 + 0 = \lambda Z_t^x(\omega) + Z_t^y(\omega). \tag{162}$$

Combining this with (161) demonstrates that for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$, $t \in [0,T]$, $\omega \in \Omega$ it holds that

$$Z_t^{\lambda x+y}(\omega) + \lambda Z_t^0(\omega) = \lambda Z_t^x(\omega) + Z_t^y(\omega). \tag{163}$$

Furthermore, observe that (160) ensures for all $x \in \mathbb{R}^d$ that

$$\Pi \subseteq \{\forall\, t \in [0,T] \colon Z_t^x = Y_t^x\}. \tag{164}$$

This and (159) show for all $x \in \mathbb{R}^d$ that

$$\mathbb{P}\big(\forall\, t \in [0,T] \colon Z_t^x = Y_t^x\big) = 1. \tag{165}$$

The fact that for all $x \in \mathbb{R}^d$ it holds that Y^x is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths, Lemma 2.18, and (160) therefore imply that for all $x \in \mathbb{R}^d$ it holds that Z^x is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths. Combining this and (157) with (165) demonstrates that for all $x \in \mathbb{R}^d$, $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$Z_t^x = x + \int_0^t \mu(Z_s^x)\, ds + \int_0^t \sigma(Z_s^x)\, dW_s. \tag{166}$$

This, (134), (135), and, e.g., Jentzen & Kloeden [42, Theorem 5.1] (with $T = T$, $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]}) = (\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$, $H = \mathbb{R}^d$, $\|\cdot\|_H = \|\cdot\|_{\mathbb{R}^d}$, $U = \mathbb{R}^d$, $\|\cdot\|_U = \|\cdot\|_{\mathbb{R}^d}$, $Qv = v$, $(W_t)_{t\in[0,T]} = (W_t)_{t\in[0,T]}$, $D(A) = \mathbb{R}^d$, $Av = 0$, $\eta = 1$, $\alpha = 0$, $\delta = 0$, $F(v) = \mu(v)$, $\beta = 0$, $B(v)u = \sigma(v)u$, $\gamma = 0$, $p = 2$, $\xi = (\Omega \ni \omega \mapsto x \in \mathbb{R}^d)$ for $u, v, x \in \mathbb{R}^d$ in the notation of [42, Theorem 5.1]) (cf., e.g., Da Prato & Zabczyk [19, Item (i) in Theorem 7.4] and Klenke [46, Theorem 26.8]) establish that for all $x \in \mathbb{R}^d$ and all $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic processes with continuous sample paths $V\colon [0,T]\times\Omega \to \mathbb{R}^d$ which satisfy that for all $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$V_t = x + \int_0^t \mu(V_s)\,ds + \int_0^t \sigma(V_s)\,dW_s \tag{167}$$

it holds that $\forall\, t \in [0,T]\colon\ \mathbb{P}(V_t = Z_t^x) = 1$. Lemma 2.16 hence demonstrates that for all $x \in \mathbb{R}^d$ and all $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic processes with continuous sample paths $V\colon [0,T]\times\Omega \to \mathbb{R}^d$ which satisfy that for all $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$V_t = x + \int_0^t \mu(V_s)\,ds + \int_0^t \sigma(V_s)\,dW_s \tag{168}$$

it holds that $\mathbb{P}(\forall\, t \in [0,T]\colon V_t = Z_t^x) = 1$. Combining this with (163) and (166) completes the proof of Proposition 2.20. ☐

2.5 Viscosity solutions for partial differential equations

In this subsection we apply results on viscosity solutions for PDEs from the literature (cf., e.g., Crandall et al. [16], Crandall&Lions [17], and Hairer et al. [31, Subsections 4.3–4.4]) to establish in Proposition 2.22 and Corollary 2.23 the existence, uniqueness, and regularity results for viscosity solutions which we need for our proofs of the ANN approximation results. Our proof of Proposition 2.22 employs the following well-known result, Lemma 2.21, on the existence of a Lyapunov-type function for SDEs under the coercivity-type hypothesis in (169). For the sake of completeness we include in this subsection also a proof of Lemma 2.21.

Lemma 2.21. *Let $d, m \in \mathbb{N}$, $p \in [4,\infty)$, let $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, let $|||\cdot|||\colon \mathbb{R}^{d\times m} \to [0,\infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times m}$, let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be functions which satisfy that*

$$\left[\sup_{x\in\mathbb{R}^d} \frac{\langle x, \mu(x)\rangle}{(1+\|x\|^2)}\right] + \left[\sup_{x\in\mathbb{R}^d} \frac{|||\sigma(x)|||}{(1+\|x\|)}\right] < \infty, \tag{169}$$

and let $V\colon \mathbb{R}^d \to \mathbb{R}$ be the function which satisfies for all $x \in \mathbb{R}^d$ that $V(x) = 1 + \|x\|^p$. Then *(i) it holds that* $V \in C^2(\mathbb{R}^d, (0,\infty))$ *and*

(ii) there exists $\rho \in (0,\infty)$ *such that for all* $x \in \mathbb{R}^d$ *it holds that*

$$\langle \mu(x), (\nabla V)(x) \rangle + \operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess} V)(x)\big) \leq \rho V(x). \tag{170}$$

Proof of Lemma 2.21. Throughout this proof let $\mu_i \colon \mathbb{R}^d \to \mathbb{R}$, $i \in \{1,2,\dots,d\}$, and $\sigma_{i,j} \colon \mathbb{R}^d \to \mathbb{R}$, $i \in \{1,2,\dots,d\}$, $j \in \{1,2,\dots,m\}$, satisfy for all $x \in \mathbb{R}^d$ that $\mu(x) = (\mu_i(x))_{i\in\{1,2,\dots,d\}}$ and $\sigma(x) = (\sigma_{i,j}(x))_{i\in\{1,2,\dots,d\},j\in\{1,2,\dots,m\}}$ and let $c \in [0,\infty)$ satisfy for all $x \in \mathbb{R}^d$ that

$$\langle x, \mu(x) \rangle \leq c(1 + \|x\|^2) \qquad \text{and} \qquad \|\!|\sigma(x)|\!\| \leq c(1 + \|x\|). \tag{171}$$

Note that the fact that for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, $i \in \{1,2,\dots,d\}$ it holds that $V(x) = 1 + \big[\sum_{i=1}^d |x_i|^2\big]^{p/2}$ assures that for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, $i \in \{1,2,\dots,d\}$ it holds that

$$\big(\tfrac{\partial}{\partial x_i} V\big)(x) = \frac{p}{2}\left[\textstyle\sum_{i=1}^d |x_i|^2\right]^{p/2-1} 2x_i = p\,\|x\|^{p-2}\, x_i. \tag{172}$$

This ensures that for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, $i, j \in \{1,2,\dots,d\}$ it holds that

$$\big(\tfrac{\partial^2}{\partial x_j \partial x_i} V\big)(x) = \begin{cases} p(p-2)\,\|x\|^{p-4}\, x_i x_j & : i \neq j \\ p(p-2)\,\|x\|^{p-4}\, |x_i|^2 + p\,\|x\|^{p-2} & : i = j \end{cases}. \tag{173}$$

Combining this and (172) proves item (i). Next observe that (172) and (173) demonstrate that for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\langle \mu(x), (\nabla V)(x) \rangle + \operatorname{Trace}\big(\sigma(x)[\sigma(x)]^* (\operatorname{Hess} V)(x)\big) \\
&= \Big[\textstyle\sum_{i=1}^d \mu_i(x) \big(\tfrac{\partial}{\partial x_i} V\big)(x)\Big] + \Big[\textstyle\sum_{i,j=1}^d \sum_{k=1}^m \sigma_{i,k}(x)\sigma_{j,k}(x) \big(\tfrac{\partial^2}{\partial x_i \partial x_j} V\big)(x)\Big] \\
&= \Big[\textstyle\sum_{i=1}^d \mu_i(x)\, p\, \|x\|^{p-2}\, x_i\Big] + \Big[\textstyle\sum_{i=1}^d \sum_{k=1}^m \sigma_{i,k}(x)\sigma_{i,k}(x)\, p\, \|x\|^{p-2}\Big] \\
&\quad + \Big[\textstyle\sum_{i,j=1}^d \sum_{k=1}^m \sigma_{i,k}(x)\sigma_{j,k}(x)\, p(p-2)\, \|x\|^{p-4}\, x_i x_j\Big] \\
&= p\, \|x\|^{p-2} \langle x, \mu(x) \rangle + p\, \|x\|^{p-2} \Big[\textstyle\sum_{i=1}^d \sum_{k=1}^m |\sigma_{ik}(x)|^2\Big] \\
&\quad + p(p-2)\, \|x\|^{p-4} \Big[\textstyle\sum_{i,j=1}^d \sum_{k=1}^m \sigma_{i,k}(x)\sigma_{j,k}(x)\, x_i x_j\Big] \\
&= p\, \|x\|^{p-2} \langle x, \mu(x) \rangle + p\, \|x\|^{p-2} |||\sigma(x)|||^2 \\
&\quad + p(p-2)\, \|x\|^{p-4} (x^* \sigma(x)[\sigma(x)]^* x) \\
&\leq p\, \|x\|^{p-2} \langle x, \mu(x) \rangle + p\, \|x\|^{p-2} |||\sigma(x)|||^2 + p(p-2)\, \|x\|^{p-2} |||\sigma(x)|||^2 \\
&= p\, \|x\|^{p-2} \langle x, \mu(x) \rangle + p(p-1)\, \|x\|^{p-2} |||\sigma(x)|||^2.
\end{aligned} \tag{174}$$

This and (171) ensure that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\langle \mu(x), (\nabla V)(x) \rangle + \operatorname{Trace}\big(\sigma(x)[\sigma(x)]^* (\operatorname{Hess} V)(x)\big) \\
&\leq p\, \|x\|^{p-2}\, c(1 + \|x\|^2) + p(p-1)\, \|x\|^{p-2}\, c^2(1 + \|x\|)^2 \\
&= (pc + p(p-1)c^2)\, \|x\|^{p-2} + 2p(p-1)c^2\, \|x\|^{p-1} + (pc + p(p-1)c^2)\, \|x\|^p \\
&\leq (pc + p(p-1)c^2)(1 + \|x\|^p) + 2p(p-1)c^2(1 + \|x\|^p) \\
&\quad + (pc + p(p-1)c^2)(1 + \|x\|^p) \\
&= (2pc + 4p(p-1)c^2)(1 + \|x\|^p) = (2pc + 4p(p-1)c^2)V(x).
\end{aligned} \tag{175}$$

This establishes item (ii). The proof of Lemma 2.21 is thus completed. □

Proposition 2.22 (Existence and uniqueness of viscosity solutions)**.** *Let $d, m \in \mathbb{N}$, $c \in [0, \infty)$, let $\langle \cdot, \cdot \rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the d-dimensional Euclidean norm, let $|||\cdot||| \colon \mathbb{R}^{d \times m} \to [0, \infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d \times m}$, let $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ be a continuous and at most polynomially growing function, and let $\mu \colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma \colon \mathbb{R}^d \to \mathbb{R}^{d \times m}$ be functions which satisfy for all $x, y \in \mathbb{R}^d$ that*

$$\langle x, \mu(x) \rangle \leq c(1 + \|x\|^2), \qquad |||\sigma(x)||| \leq c(1 + \|x\|), \tag{176}$$

$$\text{and} \qquad \|\mu(x) - \mu(y)\| + |||\sigma(x) - \sigma(y)||| \le c\,\|x - y\|\,. \tag{177}$$

Then

(i) *there exists a continuous function* $u\colon [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ *which satisfies for all* $x \in \mathbb{R}^d$ *that* $u(0,x) = \varphi(x)$*, which satisfies for all* $T \in (0,\infty)$ *that* $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty$*, and which satisfies that* $u|_{(0,\infty)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle(\nabla_x u)(t,x), \mu(x)\rangle \tag{178}$$

for $(t,x) \in (0,\infty) \times \mathbb{R}^d$*,*

(ii) *for all* $T \in (0,\infty)$ *it holds that* $u|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle(\nabla_x u)(t,x), \mu(x)\rangle \tag{179}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$*,*

(iii) *for all* $T \in (0,\infty)$ *and all continuous functions* $v\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ *which satisfy for all* $x \in \mathbb{R}^d$ *that* $v(0,x) = \varphi(x)$*, which satisfy that* $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|v(t,x)|}{1+\|x\|^q} < \infty$*, and which satisfy that* $v|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}v)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x v)(t,x)\big) + \langle(\nabla_x v)(t,x), \mu(x)\rangle \tag{180}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ *it holds that* $v = u|_{[0,T]\times\mathbb{R}^d}$*, and*

(iv) *for every* $T \in (0,\infty)$*,* $x \in \mathbb{R}^d$*, every filtered probability space* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ *which fulfils the usual conditions, every standard* $(\mathbb{F}_t)_{t\in[0,T]}$*-Brownian motion* $W\colon [0,T] \times \Omega \to \mathbb{R}^m$*, and every* $(\mathbb{F}_t)_{t\in[0,T]}$*-adapted stochastic process with continuous sample paths* $X\colon [0,T] \times \Omega \to \mathbb{R}^d$ *which satisfies that for all* $t \in [0,T]$ *it holds* $\mathbb{P}$*-a.s. that*

$$X_t = x + \int_0^t \mu(X_s)\,ds + \int_0^t \sigma(X_s)\,dW_s \tag{181}$$

it holds that

$$u(T,x) = \mathbb{E}\big[\varphi(X_T)\big]. \tag{182}$$

Proof of Proposition 2.22. Throughout this proof let $C_n \subseteq \mathbb{R}^d$, $n \in \mathbb{N}$, be the sets which satisfy for all $n \in \mathbb{N}$ that $C_n = \{x \in \mathbb{R}^d \colon \|x\| > n\}$ and for every $p \in (0,\infty)$ let $V_p \colon \mathbb{R}^d \to (0,\infty)$ be the function which satisfies for all $x \in \mathbb{R}^d$ that $V_p(x) = 1 + \|x\|^p$. Note that the hypothesis that φ is a continuous and at most polynomially growing function, (176), (177), and Hairer et al. [31, Corollary 4.17] demonstrate that there exists a continuous function $u \colon [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ which satisfies for all $x \in \mathbb{R}^d$ that $u(0,x) = \varphi(x)$, which satisfies for all $T \in (0,\infty)$ that

$$\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty, \tag{183}$$

and which satisfies that $u|_{(0,\infty)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle(\nabla_x u)(t,x), \mu(x)\rangle \tag{184}$$

for $(t,x) \in (0,\infty) \times \mathbb{R}^d$. This establishes items (i)–(ii). For the next step let $T \in (0,\infty)$ and let $v = (v(t,x))_{(t,x)\in[0,T]\times\mathbb{R}^d} \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ be a continuous function which satisfies for all $x \in \mathbb{R}^d$ that $v(0,x) = \varphi(x)$, which satisfies that

$$\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|v(t,x)|}{1+\|x\|^q} < \infty, \tag{185}$$

and which satisfies that $v|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}v)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x v)(t,x)\big) + \langle(\nabla_x v)(t,x), \mu(x)\rangle \tag{186}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$. Observe that (183) and (185) show that

$$\inf_{q\in[3,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty \quad \text{and} \quad \inf_{q\in[3,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|v(t,x)|}{1+\|x\|^q} < \infty. \tag{187}$$

Therefore, we obtain that there exists $p \in [3,\infty)$ such that

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^p} < \infty \qquad \text{and} \qquad \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|v(t,x)|}{1+\|x\|^p} < \infty. \tag{188}$$

Next note that the fact that there exists $r_0 > 0$ such that the function $\big([r_0,\infty) \ni r \mapsto \frac{1+r^p}{1+r^{p+1}}\big)$ is monotonically decreasing ensures that there exists $n_0 \in \mathbb{N}$ such that for all $n \in \mathbb{N} \cap [n_0,\infty)$, $x \in C_n$ it holds that

$$\tfrac{1+\|x\|^p}{1+\|x\|^{p+1}} \leq \tfrac{1+n^p}{1+n^{p+1}}.$$

Therefore, we obtain that for all $w \in \{u, v\}$ it holds that

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{(t,x)\in[0,T]\times C_n} \tfrac{|w(t,x)|}{V_{p+1}(x)} \\
&= \lim_{n\to\infty} \sup_{(t,x)\in[0,T]\times C_n} \left(\tfrac{|w(t,x)|}{1+\|x\|^p} \tfrac{1+\|x\|^p}{1+\|x\|^{p+1}} \right) \\
&\leq \left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \tfrac{|w(t,x)|}{1+\|x\|^p} \right] \lim_{n\to\infty} \sup_{(t,x)\in[0,T]\times C_n} \tfrac{1+\|x\|^p}{1+\|x\|^{p+1}} \\
&\leq \left[\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \tfrac{|w(t,x)|}{1+\|x\|^p} \right] \lim_{n\to\infty} \tfrac{1+n^p}{1+n^{p+1}} = 0.
\end{aligned} \tag{189}$$

In addition, note that (176), the fact that $p + 1 \in [4, \infty)$, and Lemma 2.21 (with $d = d$, $m = m$, $p = p+1$, $\mu = \mu$, $\sigma = 2^{-1/2}\sigma$, $V = V_{p+1}$ in the notation of Lemma 2.21) prove that there exists $\rho \in (0, \infty)$ such that for all $x \in \mathbb{R}^d$ it holds that $V_{p+1} \in C^2(\mathbb{R}^d, (0, \infty))$ and

$$\langle \mu(x), (\nabla V_{p+1})(x) \rangle + \tfrac{1}{2} \operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess} V_{p+1})(x)\big) \leq \rho V_{p+1}(x). \tag{190}$$

Combining this, (177), item (ii), the fact that for all $x \in \mathbb{R}^d$ it holds that $u(0, x) = \varphi(x) = v(0, x)$, (186), and (189) with Hairer et al. [31, Corollary 4.14] (with $T = T$, $d = d$, $m = m$, $\rho = \rho$, $O = \mathbb{R}^d$, $\varphi = \varphi$, $v = 0$, $\mu = \mu$, $\sigma = 2^{-1/2}\sigma$, $V = V_{p+1}$ in the notation of Hairer et al. [31, Corollary 4.14]) establishes that $v = u|_{[0,T]\times\mathbb{R}^d}$. This proves item (iii). It thus remains to prove item (iv). For this let $T \in (0, \infty)$, $x \in \mathbb{R}^d$, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ be a filtered probability space which fulfils the usual conditions, let $W\colon [0, \infty) \times \Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t\in[0,\infty)}$-Brownian motion, and let $X\colon [0, T] \times \Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths which satisfies that for all $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that

$$X_t = x + \int_0^t \mu(X_s)\, ds + \int_0^t \sigma(X_s)\, dW_s. \tag{191}$$

Observe that (177) and Klenke [46, Theorem 26.8] assure that there exists an $(\mathbb{F}_t)_{t\in[0,\infty)}$-adapted stochastic process with continuous sample paths $Y\colon [0, \infty) \times \Omega \to \mathbb{R}^d$ which satisfies that for all $t \in [0, \infty)$ it holds $\mathbb{P}$-a.s. that

$$Y_t = x + \int_0^t \mu(Y_s)\, ds + \int_0^t \sigma(Y_s)\, dW_s \tag{192}$$

and $\mathbb{P}(\forall s \in [0,T]\colon Y_s = X_s) = 1$. This, item (i), and the Feynman-Kac formula in Hairer et al. [31, Corollary 4.17] demonstrate that

$$u(T,x) = \mathbb{E}\big[\varphi(Y_T)\big] = \mathbb{E}\big[\varphi(X_T)\big]. \tag{193}$$

This implies item (iv). The proof of Proposition 2.22 is thus completed. □

Corollary 2.23. *Let $d, m \in \mathbb{N}$, $T \in (0,\infty)$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ be a continuous and at most polynomially growing function, and let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be functions which satisfy for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that*

$$\begin{aligned} &\mu(\lambda x + y) + \lambda\mu(0) = \lambda\mu(x) + \mu(y) \\ \textit{and} \qquad &\sigma(\lambda x + y) + \lambda\sigma(0) = \lambda\sigma(x) + \sigma(y). \end{aligned} \tag{194}$$

Then

(i) *there exists a unique continuous function $u\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty$, which satisfies for all $x \in \mathbb{R}^d$ that $u(0,x) = \varphi(x)$, and which satisfies that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle(\nabla_x u)(t,x), \mu(x)\rangle \tag{195}$$

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

(ii) *for every $x \in \mathbb{R}^d$, every filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ which fulfils the usual conditions, every standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion $W\colon [0,T] \times \Omega \to \mathbb{R}^m$, and every $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths $X\colon [0,T] \times \Omega \to \mathbb{R}^d$ which satisfies that for all $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that $X_t = x + \int_0^t \mu(X_s)\,ds + \int_0^t \sigma(X_s)\,dW_s$ it holds that*

$$u(T,x) = \mathbb{E}\big[\varphi(X_T)\big]. \tag{196}$$

Proof of Corollary 2.23. Throughout this proof let $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product and let $|||\cdot|||\colon \mathbb{R}^{d\times m} \to [0,\infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times m}$. Note that (194), Corollary 2.9, and

Corollary 2.10 assure that there exists $\kappa \in (0, \infty)$ such that for all $x, y \in \mathbb{R}^d$ it holds that

$$\|\mu(x)\| + |||\sigma(x)||| \leq \kappa(1 + \|x\|) \tag{197}$$

and

$$\|\mu(x) - \mu(y)\| + |||\sigma(x) - \sigma(y)||| \leq \kappa \|x - y\| . \tag{198}$$

This ensures that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \langle x, \mu(x) \rangle &\leq \|x\| \|\mu(x)\| \leq \|x\| \kappa(1 + \|x\|) = \kappa(\|x\| + \|x\|^2) \\ &\leq \kappa(1 + \|x\|^2 + \|x\|^2) \leq 2\kappa(1 + \|x\|^2). \end{aligned} \tag{199}$$

Combining this, the hypothesis that φ is a continuous function, and (197)–(198) with items (i)–(iii) in Proposition 2.22 (with $d = d$, $m = m$, $c = 2\kappa$, $\varphi = \varphi$, $\mu = \mu$, $\sigma = \sigma$ in the notation of Proposition 2.22) proves item (i). Moreover, note that item (iv) in Proposition 2.22 and item (i) establish item (ii). The proof of Corollary 2.23 is thus completed. □

3 Artificial neural network approximations

3.1 Construction of a realization on the artificial probability space

In Theorem 3.14 in Subsection 3.6 below we establish that the number of required parameters of an ANN to approximate the solution of the Black-Scholes PDE grows at most polynomially in both the reciprocal of the prescribed approximation accuracy $\varepsilon > 0$ and the PDE dimension $d \in \mathbb{N}$. An important ingredient in our proof of Theorem 3.14 is an artificial probability space on which we establish the existence of a suitable realization with the desired approximation properties. In this subsection we provide, roughly speaking, in the elementary result in Proposition 3.3 below on a very abstract level the argument for the existence of such a realization on the artificial probability space. Proposition 3.3 is an immediate consequence from the elementary result in Corollary 3.2 below. Corollary 3.2, in turn, follows from the following elementary lemma, Lemma 3.1.

Lemma 3.1. *Let $\varepsilon \in \mathbb{R}$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable which satisfies that $\mathbb{P}(X > \varepsilon) = 1$. Then*

(i) it holds that $\mathbb{E}[\max\{-X, 0\}] < \infty$ and

(ii) *it holds that* $\mathbb{E}[X] > \varepsilon$.

Proof of Lemma 3.1. Observe that the hypothesis that $\mathbb{P}(X > \varepsilon) = 1$ establishes item (i). Next note that the fact that

$$\{X > \varepsilon\} = \cup_{n\in\mathbb{N}}\{X \geq \varepsilon + \tfrac{1}{n}\}, \tag{200}$$

the hypothesis that $\mathbb{P}(X > \varepsilon) = 1$, and the fact that $\mathbb{P}$ is continuous from below imply that

$$0 < 1 = \mathbb{P}(X > \varepsilon) = \mathbb{P}\big(\cup_{n\in\mathbb{N}}\{X \geq \varepsilon + \tfrac{1}{n}\}\big) = \lim_{n\to\infty} \mathbb{P}(X \geq \varepsilon + \tfrac{1}{n}). \tag{201}$$

Hence, we obtain that there exists $\delta \in (0,\infty)$ such that

$$\mathbb{P}(X \geq \varepsilon + \delta) > 0. \tag{202}$$

The hypothesis that $\mathbb{P}(X > \varepsilon) = 1$ therefore ensures that

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\big[X\mathbb{1}_{\{X\geq\varepsilon+\delta\}} + X\mathbb{1}_{\{X\in(\varepsilon,\varepsilon+\delta)\}}\big] \\
&\geq \mathbb{E}\big[(\varepsilon+\delta)\mathbb{1}_{\{X\geq\varepsilon+\delta\}} + \varepsilon\mathbb{1}_{\{X\in(\varepsilon,\varepsilon+\delta)\}}\big] \\
&= (\varepsilon+\delta)\,\mathbb{P}(X \geq \varepsilon+\delta) + \varepsilon\,\mathbb{P}(X \in (\varepsilon,\varepsilon+\delta)) \\
&= \delta\,\mathbb{P}(X \geq \varepsilon+\delta) + \varepsilon\,\big[\mathbb{P}(X \geq \varepsilon+\delta) + \mathbb{P}(X \in (\varepsilon,\varepsilon+\delta))\big] \\
&= \delta\,\mathbb{P}(X \geq \varepsilon+\delta) + \varepsilon\,\mathbb{P}(X > \varepsilon) \\
&= \delta\,\mathbb{P}(X \geq \varepsilon+\delta) + \varepsilon > \varepsilon.
\end{aligned} \tag{203}$$

This establishes item (ii). The proof of Lemma 3.1 is thus completed. □

Corollary 3.2. *Let* $\varepsilon \in [0,\infty)$, *let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, and let* $X\colon \Omega \to \mathbb{R}$ *be a random variable which satisfies that* $\mathbb{P}(|X| > \varepsilon) = 1$. *Then it holds that*

$$\mathbb{E}[|X|] > \varepsilon. \tag{204}$$

Proof of Corollary 3.2. Note that item (ii) in Lemma 3.1 (with $\varepsilon = \varepsilon$, $X = |X|$ in the notation of Lemma 2.11) ensures that $\mathbb{E}[|X|] > \varepsilon$. The proof of Corollary 3.2 is thus completed. □

Proposition 3.3. *Let* $\varepsilon \in [0,\infty)$, *let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, and let* $X\colon \Omega \to \mathbb{R}$ *be a random variable which satisfies that*

$$\mathbb{E}[|X|] \leq \varepsilon. \tag{205}$$

Then it holds that $\mathbb{P}(|X| \leq \varepsilon) > 0$.

Proof of Proposition 3.3. Note that Corollary 3.2 and (205) ensure that $\mathbb{P}(|X| > \varepsilon) < 1$. Therefore, we obtain that

$$\mathbb{P}(|X| \le \varepsilon) = 1 - \mathbb{P}(|X| > \varepsilon) > 0. \tag{206}$$

The proof of Proposition 3.3 is thus completed. □

3.2 Approximation error estimates

Proposition 3.4. *Let $d, n \in \mathbb{N}$, $p \in [2, \infty)$, $T \in (0, \infty)$, $c, \varepsilon, L \in [0, \infty)$, $\mathbf{v}, \mathbf{w} \in [1/p, \infty)$, let $\langle \cdot, \cdot \rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the d-dimensional Euclidean norm, let $|||\cdot||| \colon \mathbb{R}^{d \times d} \to [0, \infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d \times d}$, let $\nu \colon \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ be a probability measure, let $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ be a continuous function, let $\phi \colon \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})$-measurable function which satisfies for all $x \in \mathbb{R}^d$ that*

$$|\phi(x)| \le c\,(1 + \|x\|^{\mathbf{v}}) \qquad \text{and} \qquad |\varphi(x) - \phi(x)| \le \varepsilon(1 + \|x\|^{\mathbf{w}}), \tag{207}$$

and let $\mu \colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma \colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$ be functions which satisfy for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that

$$\mu(\lambda x + y) + \lambda \mu(0) = \lambda \mu(x) + \mu(y), \tag{208}$$

$$\sigma(\lambda x + y) + \lambda \sigma(0) = \lambda \sigma(x) + \sigma(y), \tag{209}$$

and $\|\mu(x)\| + |||\sigma(x)||| \le L(1 + \|x\|)$. Then

(i) there exists a unique continuous function $u \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ which satisfies that $\inf_{q \in (0,\infty)} \sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \frac{|u(t,x)|}{1 + \|x\|^q} < \infty$, which satisfies for all $x \in \mathbb{R}^d$ that $u(0, x) = \varphi(x)$, and which satisfies that $u|_{(0,T) \times \mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t} u)(t, x) = \tfrac{1}{2} \operatorname{Trace}\big(\sigma(x)[\sigma(x)]^* (\operatorname{Hess}_x u)(t, x)\big) + \langle (\nabla_x u)(t, x), \mu(x) \rangle \tag{210}$$

for $(t, x) \in (0, T) \times \mathbb{R}^d$ and *(ii) there exist* $A_1, A_2, \dots, A_n \in \mathbb{R}^{d\times d}$*,* $b_1, b_2, \dots, b_n \in \mathbb{R}^d$ *such that*

$$\begin{aligned}
&\left[\int_{\mathbb{R}^d} \left|u(T,x) - \tfrac{1}{n}\left[\textstyle\sum_{i=1}^n \phi(A_i x + b_i)\right]\right|^p \nu(dx)\right]^{1/p} \\
&\leq \varepsilon\Bigg(1 + 2^{\mathbf{w}/2}\exp\Big(\big[\sqrt{T} + \max\{2,\mathbf{w}\}\big]^2 L^2\, T\mathbf{w}\Big) \\
&\qquad \cdot \left[L\big(T + \max\{2,\mathbf{w}\}\sqrt{T}\big) + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{w}p}\,\nu(dx)\right]^{1/(\mathbf{w}p)}\right]^{\mathbf{w}}\Bigg) \\
&\quad + n^{-1/2} 4\, c\,(p-1)^{1/2}\Bigg(1 + 2^{\mathbf{v}/2}\exp\Big(\big[\sqrt{T} + \max\{2,\mathbf{v}p\}\big]^2 L^2\, T\mathbf{v}\Big) \\
&\qquad \cdot \left[L\big(T + \max\{2,\mathbf{v}p\}\sqrt{T}\big) + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p}\,\nu(dx)\right]^{1/(\mathbf{v}p)}\right]^{\mathbf{v}}\Bigg).
\end{aligned} \tag{211}$$

Proof of Proposition 3.4. Throughout this proof let $e_j \in \mathbb{R}^d$, $j \in \{1, 2, \dots, d\}$, be given by $e_1 = (1, 0, \dots, 0), e_2 = (0, 1, 0, \dots, 0), \dots, e_d = (0, \dots, 0, 1)$, let $m\colon (0,\infty) \to [2,\infty)$ be the function which satisfies for all $z \in (0,\infty)$ that $m(z) = \max\{2, z\}$, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ be a filtered probability space which fulfils the usual conditions, let $W^i\colon [0,T]\times\Omega \to \mathbb{R}^d$, $i \in \mathbb{N}$, be independent standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motions, let $X^{i,x}\colon [0,T]\times\Omega \to \mathbb{R}^d$, $i \in \mathbb{N}$, $x \in \mathbb{R}^d$, be $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic processes with continuous sample paths which satisfy that for all $i \in \mathbb{N}$, $x \in \mathbb{R}^d$, $t \in [0,T]$ it holds $\mathbb{P}$-a.s. that

$$X_t^{i,x} = x + \int_0^t \mu(X_s^{i,x})\, ds + \int_0^t \sigma(X_s^{i,x})\, dW_s^i \tag{212}$$

and that for all $i \in \mathbb{N}$, $t \in [0,T]$, $\omega \in \Omega$, $\lambda \in \mathbb{R}$, $x, y \in \mathbb{R}^d$ it holds that

$$X_t^{i,\lambda x + y}(\omega) + \lambda X_t^{i,0}(\omega) = \lambda X_t^{i,x}(\omega) + X_t^{i,y}(\omega) \tag{213}$$

(cf. Proposition 2.20), and let $\mathscr{A}_i\colon \Omega \to \mathbb{R}^{d\times d}$, $i \in \mathbb{N}$, and $\mathscr{B}_i\colon \Omega \to \mathbb{R}^d$, $i \in \mathbb{N}$, be the random variables which satisfy that for all $i \in \mathbb{N}$, $\omega \in \Omega$ it holds that $\mathscr{B}_i(\omega) = X_T^{i,0}(\omega)$ and

$\mathscr{A}_i(\omega)$

$$= \left(X_T^{i,e_1}(\omega) - X_T^{i,0}(\omega) \middle| X_T^{i,e_2}(\omega) - X_T^{i,0}(\omega) \middle| \cdots \middle| X_T^{i,e_d}(\omega) - X_T^{i,0}(\omega) \right). \tag{214}$$

Observe that (207) assures for all $x \in \mathbb{R}^d$ that

$$\begin{aligned} |\varphi(x)| &\leq |\phi(x)| + |\varphi(x) - \phi(x)| \leq c(1 + \|x\|^{\mathbf{v}}) + \varepsilon(1 + \|x\|^{\mathbf{w}}) \\ &\leq c(2 + \|x\|^{\max\{\mathbf{v},\mathbf{w}\}}) + \varepsilon(2 + \|x\|^{\max\{\mathbf{v},\mathbf{w}\}}) \\ &\leq 2(c + \varepsilon)(1 + \|x\|^{\max\{\mathbf{v},\mathbf{w}\}}). \end{aligned} \tag{215}$$

Therefore, we obtain that φ is an at most polynomially growing function. This, (208), (209), and item (i) in Corollary 2.23 (with $d = d$, $m = d$, $T = T$, $\varphi = \varphi$, $\mu = \mu$, $\sigma = \sigma$ in the notation of Corollary 2.23) demonstrate that there exists a unique continuous function $u\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ which satisfies for all $x \in \mathbb{R}^d$ that $u(0,x) = \varphi(x)$, which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty$, and which satisfies that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2} \operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle(\nabla_x u)(t,x), \mu(x)\rangle \tag{216}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$. This proves item (i). Moreover, note that item (ii) in Corollary 2.23, item (i), and (212) establish that for all $x \in \mathbb{R}^d$ it holds that

$$u(T,x) = \mathbb{E}\big[\varphi(X_T^{1,x})\big]. \tag{217}$$

Moreover, note that (213), (214), and Lemma 2.7 demonstrate that for all $i \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\omega \in \Omega$ it holds that

$$X_T^{i,x}(\omega) = \mathscr{A}_i(\omega)x + \mathscr{B}_i(\omega). \tag{218}$$

This ensures that for all $i \in \mathbb{N}$, $\omega \in \Omega$ it holds that the function

$$\mathbb{R}^d \ni x \mapsto X_T^{i,x}(\omega) \in \mathbb{R}^d \tag{219}$$

is continuous. Combining this and the fact that for all $i \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $X_T^{i,x}\colon \Omega \to \mathbb{R}^d$ is $\mathcal{F}/\mathcal{B}(\mathbb{R}^d)$-measurable with Beck et al. [4, Lemma 2.4] establishes that for all $i \in \mathbb{N}$ it holds that the function

$$\mathbb{R}^d \times \Omega \ni (x,\omega) \mapsto X_T^{i,x}(\omega) \in \mathbb{R}^d \tag{220}$$

is $(\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{F})/\mathcal{B}(\mathbb{R}^d)$-measurable. This, (207), and the triangle inequality assure that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left[\int_{\mathbb{R}^d} \left|\mathbb{E}\big[\varphi(X_T^{1,x})\big] - \mathbb{E}\big[\phi(X_T^{1,x})\big]\right|^p \nu(dx)\right]^{1/p} \\
&= \left[\int_{\mathbb{R}^d} \left|\mathbb{E}\big[\varphi(X_T^{1,x}) - \phi(X_T^{1,x})\big]\right|^p \nu(dx)\right]^{1/p} \\
&\leq \left[\int_{\mathbb{R}^d} \big(\mathbb{E}\big[\big|\varphi(X_T^{1,x}) - \phi(X_T^{1,x})\big|\big]\big)^p \nu(dx)\right]^{1/p} \\
&\leq \left[\int_{\mathbb{R}^d} \big(\mathbb{E}\big[\varepsilon\big(1 + \|X_T^{1,x}\|^{\mathbf{w}}\big)\big]\big)^p \nu(dx)\right]^{1/p} \\
&= \varepsilon \left[\int_{\mathbb{R}^d} \big(1 + \mathbb{E}\big[\|X_T^{1,x}\|^{\mathbf{w}}\big]\big)^p \nu(dx)\right]^{1/p} \\
&\leq \varepsilon \left(1 + \left[\int_{\mathbb{R}^d} \big(\mathbb{E}\big[\|X_T^{1,x}\|^{\mathbf{w}}\big]\big)^p \nu(dx)\right]^{1/p}\right).
\end{aligned} \tag{221}$$

Furthermore, observe that Jensen's inequality, the hypothesis that for all $x \in \mathbb{R}^d$ it holds that $\|\mu(x)\| + |||\sigma(x)||| \leq L(1 + \|x\|)$, (212), and Proposition 2.14 (with $d = d$, $p = m(z)$, $\mathfrak{m}_1 = \mathfrak{m}_2 = \mathfrak{s}_1 = \mathfrak{s}_2 = L$, $T = T$, $\xi = x$, $\mu = \mu$, $\sigma = \sigma$ for $z \in (0, \infty)$, $x \in \mathbb{R}^d$ in the notation of Proposition 2.14) prove that for all $x \in \mathbb{R}^d$, $z \in (0, \infty)$ it holds that

$$\begin{aligned}
\mathbb{E}\big[\|X_T^{1,x}\|^z\big] &= \mathbb{E}\left[\left[\|X_T^{1,x}\|^{m(z)}\right]^{z/m(z)}\right] \leq \left(\mathbb{E}\left[\|X_T^{1,x}\|^{m(z)}\right]\right)^{z/m(z)} \\
&\leq \left[\sqrt{2}\big(\|x\| + LT + Lm(z)\sqrt{T}\big) \exp\Big(\big[L\sqrt{T} + Lm(z)\big]^2 T\Big)\right]^z \\
&= 2^{z/2}\left[\|x\| + L\big(T + m(z)\sqrt{T}\big)\right]^z \exp\Big(\big[\sqrt{T} + m(z)\big]^2 L^2 Tz\Big).
\end{aligned} \tag{222}$$

The fact that $\mathbf{w}p \geq 1$ and the triangle inequality hence prove that for all

$x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
&\left[\int_{\mathbb{R}^d} \left(\mathbb{E}\big[\|X_T^{1,x}\|^{\mathbf{w}}\big]\right)^p \, \nu(dx)\right]^{1/p} \\
&\leq \Bigg[\int_{\mathbb{R}^d} \bigg[2^{\mathbf{w}/2}\Big[\|x\| + L\big(T + m(\mathbf{w})\sqrt{T}\big)\Big]^{\mathbf{w}} \\
&\qquad \cdot \exp\Big(\big[\sqrt{T} + m(\mathbf{w})\big]^2 L^2 \, T\mathbf{w}\Big)\bigg]^p \, \nu(dx)\Bigg]^{1/p} \\
&= 2^{\mathbf{w}/2} \exp\Big(\big[\sqrt{T} + m(\mathbf{w})\big]^2 L^2 \, T\mathbf{w}\Big) \\
&\qquad \cdot \left[\left[\int_{\mathbb{R}^d} \Big[\|x\| + L\big(T + m(\mathbf{w})\sqrt{T}\big)\Big]^{\mathbf{w}p} \, \nu(dx)\right]^{1/(\mathbf{w}p)}\right]^{\mathbf{w}} \\
&\leq 2^{\mathbf{w}/2} \exp\Big(\big[\sqrt{T} + m(\mathbf{w})\big]^2 L^2 \, T\mathbf{w}\Big) \\
&\qquad \cdot \left[\left[\int_{\mathbb{R}^d} \|x\|^{\mathbf{w}p} \, \nu(dx)\right]^{1/(\mathbf{w}p)} + L\big(T + m(\mathbf{w})\sqrt{T}\big)\right]^{\mathbf{w}}.
\end{aligned} \tag{223}
$$

Combining this and (221) demonstrates that for all $x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
&\left[\int_{\mathbb{R}^d} \big|\mathbb{E}\big[\varphi(X_T^{1,x})\big] - \mathbb{E}\big[\phi(X_T^{1,x})\big]\big|^p \, \nu(dx)\right]^{1/p} \\
&\leq \varepsilon \Bigg(1 + 2^{\mathbf{w}/2} \exp\Big(\big[\sqrt{T} + m(\mathbf{w})\big]^2 L^2 \, T\mathbf{w}\Big) \\
&\qquad \cdot \left[\left[\int_{\mathbb{R}^d} \|x\|^{\mathbf{w}p} \, \nu(dx)\right]^{1/(\mathbf{w}p)} + L\big(T + m(\mathbf{w})\sqrt{T}\big)\right]^{\mathbf{w}}\Bigg).
\end{aligned} \tag{224}
$$

Moreover, observe that the fact that $W^i\colon [0,T] \times \Omega \to \mathbb{R}^d$, $i \in \mathbb{N}$, are independent Brownian motions ensures for every $x \in \mathbb{R}^d$ that $X_T^{i,x}\colon \Omega \to \mathbb{R}^d$, $i \in \mathbb{N}$, are i.i.d. random variables (cf., e.g., Beck et al. [4, Theorem 2.8] and Klenke [46, Theorem 15.8]). Combining this, (207), (222), and the hypothesis that $\phi\colon \mathbb{R}^d \to \mathbb{R}$ is a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})$-measurable function proves for every $x \in \mathbb{R}^d$ that $\phi(X_T^{i,x})\colon \Omega \to \mathbb{R}$, $i \in \{1, 2, \ldots, n\}$, are i.i.d. random variables which satisfy for every $x \in \mathbb{R}^d$ that

$$
\mathbb{E}\big[\|\phi(X_T^{1,x})\|\big] \leq c \left(1 + \mathbb{E}\big[\|X_T^{1,x}\|^{\mathbf{v}}\big]\right) < \infty. \tag{225}
$$

This, Hölder's inequality, Fubini's theorem, and Corollary 2.5 demonstrate that

$$
\begin{aligned}
&\mathbb{E}\bigg[\bigg[\int_{\mathbb{R}^d} \big|\mathbb{E}\big[\phi(X_T^{1,x})\big] - \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n \phi(X_T^{i,x})\big]\big|^p \,\nu(dx)\bigg]^{1/p}\bigg] \\
&\leq \bigg(\mathbb{E}\bigg[\int_{\mathbb{R}^d} \big|\mathbb{E}\big[\phi(X_T^{1,x})\big] - \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n \phi(X_T^{i,x})\big]\big|^p \,\nu(dx)\bigg]\bigg)^{1/p} \\
&= \bigg(\int_{\mathbb{R}^d} \mathbb{E}\Big[\big|\mathbb{E}\big[\phi(X_T^{1,x})\big] - \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n \phi(X_T^{i,x})\big]\big|^p\Big] \,\nu(dx)\bigg)^{1/p} \\
&\leq \bigg(\int_{\mathbb{R}^d} \bigg[\frac{2(p-1)^{1/2}}{n^{1/2}}\bigg]^p \mathbb{E}\big[\big|\phi(X_T^{1,x}) - \mathbb{E}\big[\phi(X_T^{1,x})\big]\big|^p\big] \,\nu(dx)\bigg)^{1/p} \\
&= \frac{2(p-1)^{1/2}}{n^{1/2}} \bigg(\int_{\mathbb{R}^d} \mathbb{E}\big[\big|\phi(X_T^{1,x}) - \mathbb{E}\big[\phi(X_T^{1,x})\big]\big|^p\big] \,\nu(dx)\bigg)^{1/p}.
\end{aligned}
\tag{226}
$$

Moreover, observe that Hölder's inequality demonstrates that for all $x \in \mathbb{R}^d$ it holds that

$$
\mathbb{E}\Big[\big|\mathbb{E}\big[\phi(X_T^{1,x})\big]\big|^p\Big] = \big|\mathbb{E}\big[\phi(X_T^{1,x})\big]\big|^p \leq \mathbb{E}\big[\big|\phi(X_T^{1,x})\big|^p\big]. \tag{227}
$$

The triangle inequality, Hölder's inequality, (207), and (226) hence imply

that

$$
\begin{aligned}
&\mathbb{E}\Bigg[\bigg[\int_{\mathbb{R}^d}\big|\mathbb{E}\big[\phi(X_T^{1,x})\big]-\tfrac{1}{n}\big[\textstyle\sum_{i=1}^n\phi(X_T^{i,x})\big]\big|^p\,\nu(dx)\bigg]^{1/p}\Bigg]\\
&\leq\frac{2(p-1)^{1/2}}{n^{1/2}}\Bigg[\bigg(\int_{\mathbb{R}^d}\mathbb{E}\big[\big|\phi(X_T^{1,x})\big|^p\big]\,\nu(dx)\bigg)^{1/p}\\
&\qquad\qquad+\bigg(\int_{\mathbb{R}^d}\mathbb{E}\Big[\big|\mathbb{E}\big[\phi(X_T^{1,x})\big]\big|^p\Big]\nu(dx)\bigg)^{1/p}\Bigg]\\
&\leq\frac{4(p-1)^{1/2}}{n^{1/2}}\bigg(\int_{\mathbb{R}^d}\mathbb{E}\big[\big|\phi(X_T^{1,x})\big|^p\big]\,\nu(dx)\bigg)^{1/p}\\
&\leq\frac{4(p-1)^{1/2}}{n^{1/2}}\bigg(\int_{\mathbb{R}^d}\mathbb{E}\big[\big|c(1+\|X_T^{1,x}\|^{\mathbf{v}})\big|^p\big]\,\nu(dx)\bigg)^{1/p}\\
&=\frac{4c(p-1)^{1/2}}{n^{1/2}}\bigg(\int_{\mathbb{R}^d}\mathbb{E}\big[\big(1+\|X_T^{1,x}\|^{\mathbf{v}}\big)^p\big]\,\nu(dx)\bigg)^{1/p}\\
&\leq\frac{4c(p-1)^{1/2}}{n^{1/2}}\Bigg(1+\bigg[\int_{\mathbb{R}^d}\mathbb{E}\big[\|X_T^{1,x}\|^{\mathbf{v}p}\big]\,\nu(dx)\bigg]^{1/p}\Bigg).
\end{aligned}
\tag{228}
$$

This, (222), the triangle inequality, and the fact that $\mathbf{v}p \geq 1$ assure that

$$\begin{aligned}
&\mathbb{E}\Bigg[\bigg[\int_{\mathbb{R}^d} \big|\mathbb{E}\big[\phi(X_T^{1,x})\big] - \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n \phi(X_T^{i,x})\big]\big|^p \,\nu(dx)\bigg]^{1/p}\Bigg] \\
&\leq \frac{4c(p-1)^{1/2}}{n^{1/2}}\Bigg(1 + \bigg[\int_{\mathbb{R}^d} 2^{(\mathbf{v}p)/2}\Big[\|x\| + L\big(T + m(\mathbf{v}p)\sqrt{T}\big)\Big]^{\mathbf{v}p} \\
&\qquad \cdot \exp\Big(\big[\sqrt{T} + m(\mathbf{v}p)\big]^2 L^2\, T\mathbf{v}p\Big)\nu(dx)\bigg]^{1/p}\Bigg) \\
&= \frac{4c(p-1)^{1/2}}{n^{1/2}}\Bigg(1 + 2^{\mathbf{v}/2}\exp\Big(\big[\sqrt{T} + m(\mathbf{v}p)\big]^2 L^2\, T\mathbf{v}\Big) \qquad (229) \\
&\qquad \cdot \bigg[\bigg[\int_{\mathbb{R}^d} \Big[\|x\| + L\big(T + m(\mathbf{v}p)\sqrt{T}\big)\Big]^{\mathbf{v}p} \nu(dx)\bigg]^{1/(\mathbf{v}p)}\bigg]^{\mathbf{v}}\Bigg) \\
&\leq \frac{4c(p-1)^{1/2}}{n^{1/2}}\Bigg(1 + 2^{\mathbf{v}/2}\exp\Big(\big[\sqrt{T} + m(\mathbf{v}p)\big]^2 L^2\, T\mathbf{v}\Big) \\
&\qquad \cdot \bigg[\bigg[\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p}\,\nu(dx)\bigg]^{1/(\mathbf{v}p)} + L\big(T + m(\mathbf{v}p)\sqrt{T}\big)\bigg]^{\mathbf{v}}\Bigg).
\end{aligned}$$

Proposition 3.3 hence demonstrates that there exists $\boldsymbol{\omega} \in \Omega$ such that

$$\begin{aligned}
&\bigg[\int_{\mathbb{R}^d} \big|\mathbb{E}\big[\phi(X_T^{1,x})\big] - \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n \phi(X_T^{i,x}(\boldsymbol{\omega}))\big]\big|^p \,\nu(dx)\bigg]^{1/p} \\
&\leq \frac{4c(p-1)^{1/2}}{n^{1/2}}\Bigg(1 + 2^{\mathbf{v}/2}\exp\Big(\big[\sqrt{T} + m(\mathbf{v}p)\big]^2 L^2\, T\mathbf{v}\Big) \qquad (230) \\
&\qquad \cdot \bigg[\bigg[\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p}\,\nu(dx)\bigg]^{1/(\mathbf{v}p)} + L\big(T + m(\mathbf{v}p)\sqrt{T}\big)\bigg]^{\mathbf{v}}\Bigg).
\end{aligned}$$

Combining this, the triangle inequality, (217), (218), and (224) establishes

that

$$
\begin{aligned}
&\left[\int_{\mathbb{R}^d} \left|u(T,x) - \tfrac{1}{n}\Big[\textstyle\sum_{i=1}^n \phi(\mathscr{A}_i(\boldsymbol{\omega})x + \mathscr{B}_i(\boldsymbol{\omega}))\Big]\right|^p \nu(dx)\right]^{1/p} \\
&= \left[\int_{\mathbb{R}^d} \left|\mathbb{E}\big[\varphi(X_T^{1,x})\big] - \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n \phi(X_T^{1,x}(\boldsymbol{\omega}))\big]\right|^p \nu(dx)\right]^{1/p} \\
&\leq \left[\int_{\mathbb{R}^d} \left|\mathbb{E}\big[\varphi(X_T^{1,x})\big] - \mathbb{E}\big[\phi(X_T^{1,x})\big]\right|^p \nu(dx)\right]^{1/p} \\
&\quad + \left[\int_{\mathbb{R}^d} \left|\mathbb{E}\big[\phi(X_T^{1,x})\big] - \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n \phi(X_T^{i,x}(\boldsymbol{\omega}))\big]\right|^p \nu(dx)\right]^{1/p} \\
&\leq \varepsilon\Bigg(1 + 2^{\mathbf{w}/2}\exp\Big(\big[\sqrt{T} + m(\mathbf{w})\big]^2 L^2\, T\mathbf{w}\Big) \\
&\qquad \cdot \left[\left[\int_{\mathbb{R}^d} \|x\|^{\mathbf{w}p}\,\nu(dx)\right]^{1/(\mathbf{w}p)} + L\big(T + m(\mathbf{w})\sqrt{T}\big)\right]^{\mathbf{w}}\Bigg) \\
&\quad + \frac{4c(p-1)^{1/2}}{n^{1/2}}\Bigg(1 + 2^{\mathbf{v}/2}\exp\Big(\big[\sqrt{T} + m(\mathbf{v}p)\big]^2 L^2\, T\mathbf{v}\Big) \\
&\qquad \cdot \left[\left[\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p}\,\nu(dx)\right]^{1/(\mathbf{v}p)} + L\big(T + m(\mathbf{v}p)\sqrt{T}\big)\right]^{\mathbf{v}}\Bigg).
\end{aligned}
\tag{231}
$$

The proof of Proposition 3.4 is thus completed. ☐

Corollary 3.5. *Let $d, n \in \mathbb{N}$, $T \in (0,\infty)$, $\varepsilon, c, L, C \in [0,\infty)$, $\mathbf{v}, p \in [2,\infty)$, let $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, let $|||\cdot|||\colon \mathbb{R}^{d\times d} \to [0,\infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times d}$, let $\nu\colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a probability measure, assume that*

$$
C = (p-1)^{1/2}\exp\big(3\mathbf{v}(1+L^2T(\sqrt{T}+\mathbf{v}p)^2)\big)\big(1+\big[\textstyle\int_{\mathbb{R}^d}\|x\|^{p\mathbf{v}}\,\nu(dx)\big]^{1/p}\big), \tag{232}
$$

let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ be a continuous function, let $\phi\colon \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})$-measurable function which satisfies for all $x \in \mathbb{R}^d$ that

$$
|\phi(x)| \leq c\,(1+\|x\|^{\mathbf{v}}) \qquad \text{and} \qquad |\varphi(x) - \phi(x)| \leq \varepsilon(1+\|x\|^{\mathbf{v}}), \tag{233}
$$

and let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be functions which satisfy for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that

$$
\mu(\lambda x + y) + \lambda\mu(0) = \lambda\mu(x) + \mu(y), \tag{234}
$$

$$\sigma(\lambda x + y) + \lambda\sigma(0) = \lambda\sigma(x) + \sigma(y), \tag{235}$$

and $\|\mu(x)\| + |\!|\!|\sigma(x)|\!|\!| \leq L(1 + \|x\|)$*. Then*

(i) there exists a unique continuous function $u\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ *which satisfies that* $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty$*, which satisfies for all* $x \in \mathbb{R}^d$ *that* $u(0,x) = \varphi(x)$*, and which satisfies that* $u|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle(\nabla_x u)(t,x), \mu(x)\rangle \tag{236}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ *and*

(ii) there exist $A_1, A_2, \ldots, A_n \in \mathbb{R}^{d\times d}$*,* $b_1, b_2, \ldots, b_n \in \mathbb{R}^d$ *such that*

$$\left[\int_{\mathbb{R}^d} \big|u(T,x) - \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n \phi(A_i x + b_i)\big]\big|^p \,\nu(dx)\right]^{1/p} \leq (\varepsilon + n^{-1/2}\, c)C. \tag{237}$$

Proof of Corollary 3.5. Throughout this proof let $r \in (0,\infty)$ be given by $r = L\sqrt{T}(\sqrt{T} + \mathbf{v}p)$. Note that (233) implies that φ is an at most polynomially growing function. Combining this, (234), and (235) with item (i) in Corollary 2.23 (with $d = d$, $m = d$, $T = T$, $\varphi = \varphi$, $\mu = \mu$, $\sigma = \sigma$ in the notation of Corollary 2.23) demonstrates that there exists a unique continuous function $u\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ which satisfies for all $x \in \mathbb{R}^d$ that $u(0,x) = \varphi(x)$, which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty$, and which satisfies that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle(\nabla_x u)(t,x), \mu(x)\rangle \tag{238}$$

for $(t,x) \in (0,T)\times\mathbb{R}^d$. This proves item (i). Furthermore, note that item (ii) in Proposition 3.4 (with $d = d$, $n = n$, $p = p$, $T = T$, $c = c$, $\varepsilon = \varepsilon$, $L = L$, $\mathbf{v} = \mathbf{v}$, $\mathbf{w} = \mathbf{v}$, $\nu = \nu$, $\varphi = \varphi$, $\phi = \phi$, $\mu = \mu$, $\sigma = \sigma$ in the notation of Proposition 3.4) assures that there exist $A_1, A_2, \ldots, A_n \in \mathbb{R}^{d\times d}$,

$b_1, b_2, \ldots, b_n \in \mathbb{R}^d$ such that

$$\begin{aligned}
&\left[\int_{\mathbb{R}^d} \left|u(T,x) - \tfrac{1}{n}\left[\textstyle\sum_{i=1}^n \phi(A_i x + b_i)\right]\right|^p \nu(dx)\right]^{1/p} \\
&\leq \varepsilon \Bigg(1 + 2^{\mathbf{v}/2} \exp\Big(\big[\sqrt{T} + \max\{2,\mathbf{v}\}\big]^2 L^2\, T\mathbf{v}\Big) \\
&\qquad \cdot \left[L\big(T + \max\{2,\mathbf{v}\}\sqrt{T}\big) + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p}\, \nu(dx)\right]^{1/(\mathbf{v}p)}\right]^{\mathbf{v}}\Bigg) \qquad (239) \\
&\quad + n^{-1/2}\, 4\, c\, (p-1)^{1/2} \Bigg(1 + 2^{\mathbf{v}/2} \exp\Big(\big[\sqrt{T} + \max\{2,\mathbf{v}p\}\big]^2 L^2\, T\mathbf{v}\Big) \\
&\qquad \cdot \left[L\big(T + \max\{2,\mathbf{v}p\}\sqrt{T}\big) + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p}\, \nu(dx)\right]^{1/(\mathbf{v}p)}\right]^{\mathbf{v}}\Bigg).
\end{aligned}$$

Next note that the fact that $\mathbf{v}p \geq 2$ and the fact that $p \geq 1$ imply that

$$\max\{2,\mathbf{v}\} \leq \max\{2,\mathbf{v}p\} = \mathbf{v}p. \qquad (240)$$

This demonstrates that

$$\begin{aligned}
&1 + 2^{\mathbf{v}/2} \exp\Big(\big[\sqrt{T} + \max\{2,\mathbf{v}\}\big]^2 L^2\, T\mathbf{v}\Big) \\
&\qquad \cdot \left[L\big(T + \max\{2,\mathbf{v}\}\sqrt{T}\big) + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p}\, \nu(dx)\right]^{1/(\mathbf{v}p)}\right]^{\mathbf{v}} \\
&\leq 1 + 2^{\mathbf{v}/2} \exp\Big(\big[\sqrt{T} + \max\{2,\mathbf{v}p\}\big]^2 L^2\, T\mathbf{v}\Big) \\
&\qquad \cdot \left[L\big(T + \max\{2,\mathbf{v}p\}\sqrt{T}\big) + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p}\, \nu(dx)\right]^{1/(\mathbf{v}p)}\right]^{\mathbf{v}} \qquad (241) \\
&= 1 + 2^{\mathbf{v}/2} \exp\Big(L^2\, T\big[\sqrt{T} + \mathbf{v}p\big]^2 \mathbf{v}\Big) \\
&\qquad \cdot \left[L\sqrt{T}\big(\sqrt{T} + \mathbf{v}p\big) + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p}\, \nu(dx)\right]^{1/(\mathbf{v}p)}\right]^{\mathbf{v}} \\
&= 1 + 2^{\mathbf{v}/2} \exp\big(r^2 \mathbf{v}\big) \left[r + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p}\, \nu(dx)\right]^{1/(\mathbf{v}p)}\right]^{\mathbf{v}}.
\end{aligned}$$

In addition, note that the fact that for all $x \in \mathbb{R}$ it holds that $1 + x \leq \exp(x)$ and the fact that for all $y \in (0,\infty)$ it holds that $1 + y + y^2 \leq \frac{3}{2}(1 + y^2)$ ensure

that

$$\begin{aligned}
&1 + 2^{\mathbf{v}/2} \exp\bigl(r^2\mathbf{v}\bigr) \left[r + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p} \, \nu(dx) \right]^{1/(\mathbf{v}p)} \right]^{\mathbf{v}} \\
&\le 1 + \exp(\mathbf{v}/2) \exp\bigl(r^2\mathbf{v}\bigr) \left[(1+r) \max\left\{ 1, \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p} \, \nu(dx) \right]^{1/(\mathbf{v}p)} \right\} \right]^{\mathbf{v}} \\
&\le 1 + \exp(\mathbf{v}/2) \exp\bigl(r^2\mathbf{v}\bigr) \left[\exp(r) \max\left\{ 1, \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p} \, \nu(dx) \right]^{1/(\mathbf{v}p)} \right\} \right]^{\mathbf{v}} \\
&\le 2 \exp\bigl(\mathbf{v} + r^2\mathbf{v}\bigr) \exp(r)^{\mathbf{v}} \max\left\{ 1, \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p} \, \nu(dx) \right]^{1/p} \right\} \\
&\le 2 \exp\bigl((1 + r + r^2)\mathbf{v}\bigr) \left(1 + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p} \, \nu(dx) \right]^{1/p} \right) \\
&\le 2 \exp\bigl(\tfrac{3}{2}(1 + r^2)\mathbf{v}\bigr) \left(1 + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p} \, \nu(dx) \right]^{1/p} \right).
\end{aligned} \tag{242}$$

Moreover, note that the fact that $\mathbf{v} \ge 2$ implies that $8 = 2^3 \le \exp(3) \le \exp\bigl(\tfrac{3}{2}\mathbf{v}\bigr) \le \exp\bigl(\tfrac{3}{2}(1 + r^2)\mathbf{v}\bigr)$. Hence, we obtain that

$$8 \exp\bigl(\tfrac{3}{2}(1 + r^2)\mathbf{v}\bigr) \le \exp\bigl(\tfrac{3}{2}(1 + r^2)\mathbf{v}\bigr) \exp\bigl(\tfrac{3}{2}(1 + r^2)\mathbf{v}\bigr) = \exp\bigl(3(1 + r^2)\mathbf{v}\bigr). \tag{243}$$

Combining this, (232), (239), (241), and (242) establishes that

$$\begin{aligned}
&\left[\int_{\mathbb{R}^d} \left| u(T, x) - \tfrac{1}{n} \bigl[\textstyle\sum_{i=1}^{n} \phi(A_i x + b_i) \bigr] \right|^p \, \nu(dx) \right]^{1/p} \\
&\le \varepsilon \left(1 + 2^{\mathbf{v}/2} \exp\bigl(r^2\mathbf{v}\bigr) \left[r + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p} \, \nu(dx) \right]^{1/(\mathbf{v}p)} \right]^{\mathbf{v}} \right) \\
&\quad + n^{-1/2} \, 4 \, c \, (p-1)^{1/2} \left(1 + 2^{\mathbf{v}/2} \exp\bigl(r^2\mathbf{v}\bigr) \left[r + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p} \, \nu(dx) \right]^{1/(\mathbf{v}p)} \right]^{\mathbf{v}} \right) \\
&\le \bigl[\varepsilon + n^{-1/2} \, 4 \, c \, (p-1)^{1/2} \bigr] \; 2 \exp\bigl(\tfrac{3}{2}(1 + r^2)\mathbf{v}\bigr) \left(1 + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p} \, \nu(dx) \right]^{1/p} \right) \\
&\le (\varepsilon + n^{-1/2} \, c)(p-1)^{1/2} \, 8 \exp\bigl(\tfrac{3}{2}(1 + r^2)\mathbf{v}\bigr) \left(1 + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p} \, \nu(dx) \right]^{1/p} \right) \\
&\le (\varepsilon + n^{-1/2} \, c) \, (p-1)^{1/2} \exp\bigl(3(1 + r^2)\mathbf{v}\bigr) \left(1 + \left[\textstyle\int_{\mathbb{R}^d} \|x\|^{\mathbf{v}p} \, \nu(dx) \right]^{1/p} \right) \\
&= (\varepsilon + n^{-1/2} \, c) C.
\end{aligned} \tag{244}$$

The proof of Corollary 3.5 is thus completed. ◻

3.3 Cost estimates

Proposition 3.6. *Let $d \in \mathbb{N}$, $T, \varepsilon \in (0,\infty)$, $c, L, C \in [0,\infty)$, $\mathbf{v}, p \in [2,\infty)$, $n \in \mathbb{N} \cap [c^2 C^2 \varepsilon^{-2}, \infty)$, let $\langle \cdot, \cdot \rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, let $\|\cdot\| \colon \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, let $|||\cdot||| \colon \mathbb{R}^{d\times d} \to [0,\infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times d}$, let $\nu \colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a probability measure, assume that*

$$C = 2(p-1)^{1/2} \exp\bigl(3\mathbf{v}(1+L^2 T(\sqrt{T}+\mathbf{v}p)^2)\bigr)\bigl(1+\bigl[\textstyle\int_{\mathbb{R}^d} \|x\|^{p\mathbf{v}}\,\nu(dx)\bigr]^{1/p}\bigr), \tag{245}$$

let $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ be a continuous function, let $\phi \colon \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})$-measurable function which satisfies for all $x \in \mathbb{R}^d$ that

$$|\phi(x)| \leq c(1+\|x\|^{\mathbf{v}}) \qquad \text{and} \qquad |\varphi(x) - \phi(x)| \leq C^{-1}\varepsilon(1+\|x\|^{\mathbf{v}}) \tag{246}$$

and let $\mu \colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma \colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be functions which satisfy for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that

$$\mu(\lambda x + y) + \lambda\mu(0) = \lambda\mu(x) + \mu(y), \tag{247}$$

$$\sigma(\lambda x + y) + \lambda\sigma(0) = \lambda\sigma(x) + \sigma(y), \tag{248}$$

and $\|\mu(x)\| + |||\sigma(x)||| \leq L(1+\|x\|)$. Then

(i) there exists a unique continuous function $u \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty$, which satisfies for all $x \in \mathbb{R}^d$ that $u(0,x) = \varphi(x)$, and which satisfies that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\bigl(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\bigr) + \langle(\nabla_x u)(t,x), \mu(x)\rangle \tag{249}$$

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

(ii) there exist $A_1, A_2, \ldots, A_n \in \mathbb{R}^{d\times d}$, $b_1, b_2, \ldots, b_n \in \mathbb{R}^d$ such that

$$\left[\int_{\mathbb{R}^d} \bigl|u(T,x) - \tfrac{1}{n}\bigl[\textstyle\sum_{i=1}^n \phi(A_i x + b_i)\bigr]\bigr|^p \,\nu(dx)\right]^{1/p} \leq \varepsilon. \tag{250}$$

Proof of Proposition 3.6. Note that (246) implies that φ is an at most polynomially growing function. Combining this, (247), and (248) with item (i) in Corollary 2.23 (with $d = d$, $m = d$, $T = T$, $\varphi = \varphi$, $\mu = \mu$, $\sigma = \sigma$ in the notation of Corollary 2.23) establishes that there exists a unique continuous function $u\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ which satisfies for all $x \in \mathbb{R}^d$ that $u(0,x) = \varphi(x)$, which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty$, and which satisfies that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle(\nabla_x u)(t,x), \mu(x)\rangle \tag{251}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$. This proves item (i). Next note that Corollary 3.5 (with $d = d$, $n = n$, $T = T$, $\varepsilon = C^{-1}\varepsilon$, $c = c$, $L = L$, $\mathbf{v} = \mathbf{v}$, $p = p$, $\nu = \nu$, $\phi = \phi$, $\mu = \mu$, $\sigma = \sigma$ in the notation of Corollary 3.5) assures that there exist $A_1, A_2, \dots, A_n \in \mathbb{R}^{d\times d}$, $b_1, b_2, \dots, b_n \in \mathbb{R}^d$ such that

$$\left[\int_{\mathbb{R}^d} \big|u(T,x) - \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n \phi(A_i x + b_i)\big]\big|^p \, \nu(dx)\right]^{1/p} \le \big(C^{-1}\varepsilon + n^{-1/2}\,c\big)\,\frac{C}{2}. \tag{252}$$

The hypothesis that $n \ge c^2 C^2 \varepsilon^{-2}$ hence assures that

$$\begin{aligned}
&\left[\int_{\mathbb{R}^d} \big|u(T,x) - \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n \phi(A_i x + b_i)\big]\big|^p \, \nu(dx)\right]^{1/p} \\
&\le \big(C^{-1}\varepsilon + (c^2C^2\varepsilon^{-2})^{-1/2}\,c\big)\,\frac{C}{2} \\
&= \big(C^{-1}\varepsilon + C^{-1}\varepsilon\big)\,\frac{C}{2} = \varepsilon.
\end{aligned} \tag{253}$$

This establishes item (ii). The proof of Proposition 3.6 is thus completed. ☐

3.4 Representation properties for artificial neural networks

Setting 3.7. *For every $l \in \mathbb{N}$ let $\mathcal{M}_l$ be the set of all Borel measurable functions from $\mathbb{R}^l$ to $\mathbb{R}$, let*

$$\mathcal{N} = \cup_{\mathcal{L}\in\{2,3,\dots\}} \cup_{(l_0,l_1,\dots,l_{\mathcal{L}})\in((\mathbb{N}^{\mathcal{L}})\times\{1\})} \left(\times_{k=1}^{\mathcal{L}} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\right), \tag{254}$$

let $\mathbf{A}_l\colon \mathbb{R}^l \to \mathbb{R}^l$, $l \in \mathbb{N}$, *and* $\mathbf{a} \in \mathcal{M}_1$ *be functions which satisfy for all* $l \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_l) \in \mathbb{R}^l$ *that*

$$\mathbf{A}_l(x) = (\mathbf{a}(x_1), \mathbf{a}(x_2), \ldots, \mathbf{a}(x_l)), \tag{255}$$

and let $\mathcal{P}, \mathscr{P}\colon \mathcal{N} \to \mathbb{N}$ *and* $\mathcal{R}\colon \mathcal{N} \to \cup_{l=1}^{\infty} \mathcal{M}_l$ *be the functions which satisfy for all* $\mathcal{L} \in \{2, 3, \ldots\}$, $(l_0, l_1, \ldots, l_{\mathcal{L}}) \in ((\mathbb{N}^{\mathcal{L}}) \times \{1\})$, $\Phi = ((W_1, B_1), \ldots, (W_{\mathcal{L}}, B_{\mathcal{L}})) = ((W_k^{(i,j)})_{i \in \{1,2,\ldots,l_k\}, j \in \{1,2,\ldots,l_{k-1}\}}, (B_k^{(i)})_{i \in \{1,2,\ldots,l_k\}})_{k \in \{1,2,\ldots,\mathcal{L}\}} \in (\times_{k=1}^{\mathcal{L}} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$, $x_0 \in \mathbb{R}^{l_0}, x_1 \in \mathbb{R}^{l_1}, \ldots, x_{\mathcal{L}-1} \in \mathbb{R}^{l_{\mathcal{L}-1}}$ *with* $\forall\, k \in \mathbb{N} \cap (0, \mathcal{L})\colon x_k = \mathbf{A}_{l_k}(W_k x_{k-1} + B_k)$ *that*

$$\mathcal{R}(\Phi) \in \mathcal{M}_{l_0}, \qquad (\mathcal{R}(\Phi))(x_0) = W_{\mathcal{L}} x_{\mathcal{L}-1} + B_{\mathcal{L}}, \tag{256}$$

$$\mathscr{P}(\Phi) = \sum_{k=1}^{\mathcal{L}} \sum_{i=1}^{l_k} \left(\mathbb{1}_{\mathbb{R}\setminus\{0\}}(B_k^{(i)}) + \sum_{j=1}^{l_{k-1}} \mathbb{1}_{\mathbb{R}\setminus\{0\}}(W_k^{(i,j)}) \right), \tag{257}$$

and $\mathcal{P}(\Phi) = \sum_{k=1}^{\mathcal{L}} l_k(l_{k-1} + 1)$.

Lemma 3.8. *Assume Setting 3.7 and let* $d, n \in \mathbb{N}$, $A_1, A_2, \ldots, A_n \in \mathbb{R}^{d \times d}$, $b_1, b_2, \ldots, b_n \in \mathbb{R}^d$, $\phi \in \mathcal{N}$ *satisfy that* $\mathcal{R}(\phi) \in \mathcal{M}_d$. *Then there exists* $\psi \in \mathcal{N}$ *such that for all* $x \in \mathbb{R}^d$ *it holds that* $\mathcal{P}(\psi) \leq n^2\, \mathcal{P}(\phi)$, $\mathscr{P}(\psi) \leq n\, \mathcal{P}(\phi)$, $\mathcal{R}(\psi) \in \mathcal{M}_d$, *and*

$$(\mathcal{R}(\psi))(x) = \tfrac{1}{n} \big[\textstyle\sum_{i=1}^{n} (\mathcal{R}(\phi))(A_i x + b_i) \big]. \tag{258}$$

Proof of Lemma 3.8. Throughout this proof let $x \in \mathbb{R}^d$, for all $\mathcal{L} \in \{2, 3, \ldots\}$, $(l_0, l_1, \ldots, l_{\mathcal{L}}) \in ((\mathbb{N}^{\mathcal{L}}) \times \{1\})$, $\Phi \in (\times_{k=1}^{\mathcal{L}} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$ let $W_1^{(\Phi)} \in \mathbb{R}^{l_1 \times l_0}$, $W_2^{(\Phi)} \in \mathbb{R}^{l_2 \times l_1}, \ldots, W_{\mathcal{L}}^{(\Phi)} \in \mathbb{R}^{l_{\mathcal{L}} \times l_{\mathcal{L}-1}}$, $B_1^{(\Phi)} \in \mathbb{R}^{l_1}$, $B_2^{(\Phi)} \in \mathbb{R}^{l_2}, \ldots, B_{\mathcal{L}}^{(\Phi)} \in \mathbb{R}^{l_{\mathcal{L}}}$ satisfy that

$$\Phi = \big((W_1^{(\Phi)}, B_1^{(\Phi)}), (W_2^{(\Phi)}, B_2^{(\Phi)}), \ldots, (W_{\mathcal{L}}^{(\Phi)}, B_{\mathcal{L}}^{(\Phi)}) \big), \tag{259}$$

let $N \in \{2, 3, \ldots\}$, $(u_0, u_1, \ldots, u_N) \in (\{d\} \times (\mathbb{N}^{N-1}) \times \{1\})$ satisfy that $\phi \in \times_{k=1}^{N} (\mathbb{R}^{u_k \times u_{k-1}} \times \mathbb{R}^{u_k})$ (i.e., ϕ corresponds to a fully connected feedforward artificial neural network with $N+1$ layers with dimensions $(u_0, u_1, \ldots, u_N)$), let $\psi \in (\mathbb{R}^{(nu_1) \times u_0} \times \mathbb{R}^{nu_1}) \times (\times_{k=2}^{N-1} (\mathbb{R}^{(nu_k) \times (nu_{k-1})} \times \mathbb{R}^{nu_k})) \times (\mathbb{R}^{u_N \times (nu_{N-1})} \times \mathbb{R}^{u_N}) \subseteq \mathcal{N}$ (i.e., ψ corresponds to a fully connected feedforward artificial neural network with $N+1$ layers with dimensions $(u_0, nu_1, nu_2, \ldots, nu_{N-1}, u_N)$) satisfy for all $k \in \{2, 3, \dots, N-1\}$ that

$$W_1^{(\psi)} = \begin{pmatrix} W_1^{(\phi)} A_1 \\ W_1^{(\phi)} A_2 \\ \vdots \\ W_1^{(\phi)} A_n \end{pmatrix} \in \mathbb{R}^{(nu_1) \times u_0} = \mathbb{R}^{(nu_1) \times d}, \tag{260}$$

$$B_1^{(\psi)} = \begin{pmatrix} W_1^{(\phi)} b_1 + B_1^{(\phi)} \\ W_1^{(\phi)} b_2 + B_1^{(\phi)} \\ \vdots \\ W_1^{(\phi)} b_n + B_1^{(\phi)} \end{pmatrix} \in \mathbb{R}^{nu_1}, \qquad B_k^{(\psi)} = \begin{pmatrix} B_k^{(\phi)} \\ B_k^{(\phi)} \\ \vdots \\ B_k^{(\phi)} \end{pmatrix} \in \mathbb{R}^{nu_k}, \tag{261}$$

$$W_k^{(\psi)} = \begin{pmatrix} W_k^{(\phi)} & 0 & \cdots & 0 \\ 0 & W_k^{(\phi)} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & W_k^{(\phi)} \end{pmatrix} \in \mathbb{R}^{(nu_k) \times (nu_{k-1})}, \tag{262}$$

$$W_N^{(\psi)} = \begin{pmatrix} \tfrac{1}{n} W_N^{(\phi)} & \tfrac{1}{n} W_N^{(\phi)} & \cdots & \tfrac{1}{n} W_N^{(\phi)} \end{pmatrix} \in \mathbb{R}^{u_N \times (nu_{N-1})} = \mathbb{R}^{1 \times (nu_{N-1})}, \tag{263}$$

$$\text{and} \qquad B_N^{(\psi)} = B_N^{(\phi)} \in \mathbb{R}^{u_N} = \mathbb{R}, \tag{264}$$

let $y_{i,k} \in \mathbb{R}^{u_k}$, $i \in \{1, 2, \dots, n\}$, $k \in \{0, 1, \dots, N\}$, satisfy for all $i \in \{1, 2, \dots, n\}$, $k \in \{1, 2, \dots, N-1\}$ that

$$y_{i,0} = A_i x + b_i, \qquad y_{i,k} = \mathbf{A}_{u_k}(W_k^{(\phi)} y_{i,k-1} + B_k^{(\phi)}), \tag{265}$$

$$\text{and} \qquad y_{i,N} = W_N^{(\phi)} y_{i,N-1} + B_N^{(\phi)}, \tag{266}$$

and let $z_0 \in \mathbb{R}^{u_0}, z_1 \in \mathbb{R}^{nu_1}, z_2 \in \mathbb{R}^{nu_2}, \dots, z_{N-1} \in \mathbb{R}^{nu_{N-1}}, z_N \in \mathbb{R}^{u_N}$ satisfy that

$$z_0 = x, \qquad z_k = \mathbf{A}_{nu_k}(W_k^{(\psi)} z_{k-1} + B_k^{(\psi)}), \tag{267}$$

$$\text{and} \qquad z_N = W_N^{(\psi)} z_{N-1} + B_N^{(\psi)}. \tag{268}$$

Observe that (255) proves that for all $l, L \in \mathbb{N}$, $v = (v_1, v_2, \dots, v_{Ll}) \in \mathbb{R}^{(Ll)}$ it holds that

$$\begin{aligned} \mathbf{A}_{Ll}(v) &= \big(\mathbf{a}(v_1), \mathbf{a}(v_2), \dots, \mathbf{a}(v_{Ll})\big) \\ &= \big(\mathbf{A}_l(v_1, v_2, \dots, v_l), \mathbf{A}_l(v_{l+1}, v_{l+2}, \dots, v_{2l}), \dots, \\ &\qquad \mathbf{A}_l(v_{(L-1)l+1}, v_{(L-1)l+2}, \dots, v_{Ll})\big). \end{aligned} \tag{269}$$

Furthermore, note that the fact that $u_0 = d$ ensures that $\mathcal{R}(\psi) \in \mathcal{M}_{u_0} = \mathcal{M}_d$. Next observe that (256), (259), (265), and (266) imply that for all $i \in \{1, 2, \ldots, n\}$ it holds that

$$y_{i,N} = (\mathcal{R}(\phi))(y_{i,0}) = (\mathcal{R}(\phi))(A_i x + b_i). \tag{270}$$

Moreover, observe that (256), (259), (267), and (268) ensure that

$$z_N = (\mathcal{R}(\psi))(z_0) = (\mathcal{R}(\psi))(x). \tag{271}$$

Next we claim that for all $k \in \{1, 2, \ldots, N-1\}$ it holds that

$$z_k = \big(y_{1,k}, y_{2,k}, \ldots, y_{n,k}\big). \tag{272}$$

We now prove (272) by induction on $k \in \{1, 2, \ldots, N-1\}$. For the base case $k = 1$ note that (265) assures that for all $i \in \{1, 2, \ldots, n\}$ it holds that

$$\begin{aligned} y_{i,1} &= \mathbf{A}_{u_1}(W_1^{(\phi)} y_0^{(i)} + B_1^{(\phi)}) \\ &= \mathbf{A}_{u_1}(W_1^{(\phi)}(A_i x + b_i) + B_1^{(\phi)}) \\ &= \mathbf{A}_{u_1}(W_1^{(\phi)} A_i x + W_1^{(\phi)} b_i + B_1^{(\phi)}). \end{aligned} \tag{273}$$

This, (260), (261), (267), and (269) demonstrate that

$$\begin{aligned} z_1 &= \mathbf{A}_{nu_1}(W_1^{(\psi)} x + B_1^{(\psi)}) \\ &= \mathbf{A}_{nu_1}\left(\begin{pmatrix} W_1^{(\phi)} A_1 x \\ W_1^{(\phi)} A_2 x \\ \vdots \\ W_1^{(\phi)} A_n x \end{pmatrix} + \begin{pmatrix} W_1^{(\phi)} b_1 + B_1^{(\phi)} \\ W_1^{(\phi)} b_2 + B_1^{(\phi)} \\ \vdots \\ W_1^{(\phi)} b_n + B_1^{(\phi)} \end{pmatrix}\right) \\ &= \begin{pmatrix} \mathbf{A}_{u_1}(W_1^{(\phi)} A_1 x + W_1^{(\phi)} b_1 + B_1^{(\phi)}) \\ \mathbf{A}_{u_1}(W_1^{(\phi)} A_2 x + W_1^{(\phi)} b_2 + B_1^{(\phi)}) \\ \vdots \\ \mathbf{A}_{u_1}(W_1^{(\phi)} A_n x + W_1^{(\phi)} b_n + B_1^{(\phi)}) \end{pmatrix} = \begin{pmatrix} y_{1,1} \\ y_{2,1} \\ \vdots \\ y_{n,1} \end{pmatrix}. \end{aligned} \tag{274}$$

This establishes (272) in the base case $k = 1$. For the induction step $\{1, 2, \ldots, N-2\} \ni k-1 \to k \in \{2, 3 \ldots, N-1\}$ observe that (261), (262), (265), (267), and (269) imply that for all $k \in \{2, 3 \ldots, N\}$ with $z_{k-1} = (y_{1,k-1}, y_{2,k-1}, \ldots, y_{n,k-1})$ it holds that

$$\begin{aligned} z_k &= \mathbf{A}_{nu_k}\left(W_k^{(\psi)}\begin{pmatrix} y_{1,k-1} \\ y_{2,k-1} \\ \vdots \\ y_{n,k-1}\end{pmatrix} + B_k^{(\psi)}\right) \\ &= \mathbf{A}_{nu_k}\left(\begin{pmatrix} W_k^{(\phi)} y_{1,k-1} \\ W_k^{(\phi)} y_{2,k-1} \\ \vdots \\ W_k^{(\phi)} y_{n,k-1}\end{pmatrix} + \begin{pmatrix} B_k^{(\phi)} \\ B_k^{(\phi)} \\ \vdots \\ B_k^{(\phi)}\end{pmatrix}\right) \\ &= \begin{pmatrix} \mathbf{A}_{u_k}(W_k^{(\phi)} y_{1,k-1} + B_k^{(\phi)}) \\ \mathbf{A}_{u_k}(W_k^{(\phi)} y_{2,k-1} + B_k^{(\phi)}) \\ \vdots \\ \mathbf{A}_{u_k}(W_k^{(\phi)} y_{n,k-1} + B_k^{(\phi)})\end{pmatrix} = \begin{pmatrix} y_{1,k} \\ y_{2,k} \\ \vdots \\ y_{n,k}\end{pmatrix}. \end{aligned} \tag{275}$$

Induction thus proves (272). Next note that (263), (264), (266), (268), and (272) demonstrate that

$$\begin{aligned} z_N &= W_N^{(\psi)} z_{N-1} + B_N^{(\psi)} = \begin{pmatrix} \frac{1}{n} W_N^{(\phi)} & \frac{1}{n} W_N^{(\phi)} & \cdots & \frac{1}{n} W_N^{(\phi)}\end{pmatrix}\begin{pmatrix} y_{1,N-1} \\ y_{2,N-1} \\ \vdots \\ y_{n,N-1}\end{pmatrix} + B_N^{(\phi)} \\ &= \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n W_N^{(\phi)} y_{i,N-1}\big] + B_N^{(\phi)} = \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n W_N^{(\phi)} y_{i,N-1} + B_N^{(\phi)}\big] \\ &= \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n y_{i,N}\big]. \end{aligned} \tag{276}$$

Combining this with (270) and (271) establishes that

$$(\mathcal{R}(\psi))(x) = z_N = \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n y_{i,N}\big] = \tfrac{1}{n}\big[\textstyle\sum_{i=1}^n (\mathcal{R}(\phi))(A_i x + b_i)\big]. \tag{277}$$

In addition, observe that the fact that $\mathcal{P}(\phi) = \sum_{k=1}^N u_k(u_{k-1} + 1)$ assures

that

$$\begin{aligned}\mathcal{P}(\psi) &= nu_1(u_0+1) + \left[\sum_{k=2}^{N-1} nu_k(nu_{k-1}+1)\right] + u_N(nu_{N-1}+1)\\ &\leq n^2u_1(u_0+1) + \left[\sum_{k=2}^{N-1} n^2u_k(u_{k-1}+1)\right] + n^2u_N(u_{N-1}+1)\\ &\leq n^2\left[\sum_{k=1}^{N} u_k(u_{k-1}+1)\right] = n^2\,\mathcal{P}(\phi).\end{aligned} \tag{278}$$

Furthermore, note that (260) – (264) and (278) demonstrate that

$$\begin{aligned}\mathscr{P}(\psi) &\leq nu_1u_0 + nu_1 + \left[\sum_{k=2}^{N-1} nu_ku_{k-1} + nu_k\right] + u_Nnu_{N-1} + u_N\\ &\leq n\left[\sum_{k=1}^{N} u_k(u_{k-1}+1)\right] = n\,\mathcal{P}(\phi).\end{aligned} \tag{279}$$

Combining this, (277), and (278) completes the proof of Lemma 3.8. ☐

3.5 Cost estimates for artificial neural networks

Lemma 3.9. *Assume Setting 3.7, let* $d \in \mathbb{N}$, $T, \varepsilon \in (0,\infty)$, $L, C, \mathbf{C} \in [0,\infty)$, $c \in [1,\infty)$, $\mathbf{v}, p \in [2,\infty)$, *let* $\langle\cdot,\cdot\rangle\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *be the d-dimensional Euclidean scalar product, let* $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ *be the d-dimensional Euclidean norm, let* $|||\cdot|||\colon \mathbb{R}^{d\times d} \to [0,\infty)$ *be the Hilbert-Schmidt norm on* $\mathbb{R}^{d\times d}$, *let* $\nu\colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ *be a probability measure, assume that*

$$C = 2(p-1)^{1/2}\exp\bigl(3\mathbf{v}(1+L^2T(\sqrt{T}+\mathbf{v}p)^2)\bigr)\Bigl(1+\bigl[\textstyle\int_{\mathbb{R}^d}\|x\|^{p\mathbf{v}}\,\nu(dx)\bigr]^{1/p}\Bigr), \tag{280}$$

assume that $\mathbf{C} = 4(\max\{C,\varepsilon\})^4$, *let* $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ *be a continuous function, let* $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ *and* $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ *be functions which satisfy for all* $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ *that*

$$\mu(\lambda x + y) + \lambda\mu(0) = \lambda\mu(x) + \mu(y), \tag{281}$$

$$\sigma(\lambda x + y) + \lambda\sigma(0) = \lambda\sigma(x) + \sigma(y), \tag{282}$$

and $\|\mu(x)\| + |||\sigma(x)||| \leq L(1+\|x\|)$, *and let* $\phi \in \mathcal{N}$ *satisfy for all* $x \in \mathbb{R}^d$ *that* $\mathcal{R}(\phi) \in \mathcal{M}_d$, $|(\mathcal{R}(\phi))(x)| \leq c(1+\|x\|^{\mathbf{v}})$, *and*

$$|\varphi(x) - (\mathcal{R}(\phi))(x)| \leq C^{-1}\varepsilon(1+\|x\|^{\mathbf{v}}). \tag{283}$$

Then

(i) there exists a unique continuous function $u\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ *which satisfies that* $\inf_{q\in(0,\infty)}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\frac{|u(t,x)|}{1+\|x\|^q} < \infty$*, which satisfies for all* $x\in\mathbb{R}^d$ *that* $u(0,x)=\varphi(x)$*, and which satisfies that* $u|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle(\nabla_x u)(t,x),\mu(x)\rangle \tag{284}$$

for $(t,x)\in(0,T)\times\mathbb{R}^d$ *and*

(ii) there exists $\psi\in\mathcal{N}$ *such that* $\mathcal{P}(\psi)\le c^4\,\mathbf{C}\,\mathcal{P}(\phi)\,\varepsilon^{-4}$*,* $\mathscr{P}(\psi)\le c^2\,\mathbf{C}\,\mathcal{P}(\phi)\,\varepsilon^{-2}$*,* $\mathcal{R}(\psi)\in\mathcal{M}_d$*, and*

$$\left[\int_{\mathbb{R}^d}|u(T,x)-(\mathcal{R}(\psi))(x)|^p\,\nu(dx)\right]^{1/p}\le\varepsilon. \tag{285}$$

Proof of Lemma 3.9. Throughout this proof let $n=\min(\mathbb{N}\cap[c^2C^2\varepsilon^{-2},\infty))$. Note that (283) implies that φ is an at most polynomially growing function. Combining this, (281), and (282) with item (i) in Corollary 2.23 (with $d=d$, $m=d$, $T=T$, $\varphi=\varphi$, $\mu=\mu$, $\sigma=\sigma$ in the notation of Corollary 2.23) establishes that there exists a unique continuous function $u\colon[0,T]\times\mathbb{R}^d\to\mathbb{R}$ which satisfies for all $x\in\mathbb{R}^d$ that $u(0,x)=\varphi(x)$, which satisfies that $\inf_{q\in(0,\infty)}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\frac{|u(t,x)|}{1+\|x\|^q}<\infty$, and which satisfies that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle(\nabla_x u)(t,x),\mu(x)\rangle \tag{286}$$

for $(t,x)\in(0,T)\times\mathbb{R}^d$. This proves item (i). Next note that Proposition 3.6 (with $d=d$, $T=T$, $\varepsilon=\varepsilon$, $L=L$, $c=c$, $\mathbf{v}=\mathbf{v}$, $p=p$, $\nu=\nu$, $n=n$, $\varphi=\varphi$, $\phi=\mathcal{R}(\phi)$, $\mu=\mu$, $\sigma=\sigma$ in the notation of Proposition 3.6) ensures that there exist $A_1,A_2,\dots,A_n\in\mathbb{R}^{d\times d}$, $b_1,b_2,\dots,b_n\in\mathbb{R}^d$ such that

$$\left[\int_{\mathbb{R}^d}\left|u(T,x)-\tfrac{1}{n}\big[\textstyle\sum_{i=1}^n(\mathcal{R}(\phi))(A_ix+b_i)\big]\right|^p\,\nu(dx)\right]^{1/p}\le\varepsilon. \tag{287}$$

Moreover, observe that Lemma 3.8 demonstrates that there exists $\psi\in\mathcal{N}$ such that for all $x\in\mathbb{R}^d$ it holds that $\mathcal{P}(\psi)\le n^2\,\mathcal{P}(\phi)$, $\mathscr{P}(\psi)\le n\,\mathcal{P}(\phi)$, $\mathcal{R}(\psi)\in\mathcal{M}_d$, and

$$(\mathcal{R}(\psi))(x)=\tfrac{1}{n}\big[\textstyle\sum_{i=1}^n(\mathcal{R}(\phi))(A_ix+b_i)\big]. \tag{288}$$

This and (287) assure that

$$\begin{aligned}&\left[\int_{\mathbb{R}^d} |u(T,x) - (\mathcal{R}(\psi))(x)|^p \, \nu(dx)\right]^{1/p} \\ &= \left[\int_{\mathbb{R}^d} \left|u(T,x) - \tfrac{1}{n}\left[\textstyle\sum_{i=1}^n (\mathcal{R}(\phi))(A_i x + b_i)\right]\right|^p \, \nu(dx)\right]^{1/p} \leq \varepsilon.\end{aligned} \tag{289}$$

Moreover, note that the hypothesis that $\mathbf{C} = 4(\max\{C, \varepsilon\})^4$ implies that $c^2 \sqrt{\mathbf{C}} \, \varepsilon^{-2} \geq \sqrt{4\varepsilon^4} \, \varepsilon^{-2} = 2$. This ensures that

$$\begin{aligned} n &\leq c^2 \, C^2 \, \varepsilon^{-2} + 1 \leq 2 \max\{c^2 \, C^2 \, \varepsilon^{-2}, 1\} \\ &= \max\{c^2 \, 2 \, C^2 \, \varepsilon^{-2}, 2\} \\ &\leq \max\{c^2 \sqrt{\mathbf{C}} \, \varepsilon^{-2}, 2\} \\ &= c^2 \sqrt{\mathbf{C}} \, \varepsilon^{-2}. \end{aligned} \tag{290}$$

This and the fact that $\mathcal{P}(\psi) \leq n^2 \, \mathcal{P}(\phi)$ imply that

$$\mathcal{P}(\psi) \leq (c^2 \sqrt{\mathbf{C}} \, \varepsilon^{-2})^2 \, \mathcal{P}(\phi) = c^4 \, \mathbf{C} \, \mathcal{P}(\phi) \, \varepsilon^{-4}. \tag{291}$$

Furthermore, note that (290), the fact that $\mathscr{P}(\psi) \leq n \, \mathcal{P}(\phi)$, and the fact that $\mathbf{C} \geq 1$ ensure that

$$\mathscr{P}(\psi) \leq c^2 \sqrt{\mathbf{C}} \, \varepsilon^{-2} \, \mathcal{P}(\phi) \leq c^2 \, \mathbf{C} \, \mathcal{P}(\phi) \, \varepsilon^{-2}. \tag{292}$$

This, (289), (291), and the fact that $\mathcal{R}(\psi) \in \mathcal{M}_d$ establish item (ii). The proof of Lemma 3.9 is thus completed. ◻

Proposition 3.10. *Assume Setting 3.7, let $d \in \mathbb{N}$, $T, a, r, R \in (0, \infty)$, $L, \mathbf{C}, \mathbf{z} \in [0, \infty)$, $b, c \in [1, \infty)$, $\mathbf{v}, p \in [2, \infty)$, let $\langle \cdot, \cdot \rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, let $\|\cdot\| \colon \mathbb{R}^d \to [0, \infty)$ be the d-dimensional Euclidean norm, let $|\!|\!|\cdot|\!|\!| \colon \mathbb{R}^{d \times d} \to [0, \infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d \times d}$, let $\nu \colon \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ be a probability measure, assume that*

$$\begin{aligned}\mathbf{C} = 4\big[\max\{1, \tfrac{R}{r}\} \max\big\{&2(p-1)^{1/2} \exp\big(3\mathbf{v}(1 + L^2 T(\sqrt{T} + \mathbf{v}p)^2)\big) \\ &\big(1 + \big[\textstyle\int_{\mathbb{R}^d} \|x\|^{p\mathbf{v}} \, \nu(dx)\big]^{1/p}\big), R\big\}\big]^{4+\mathbf{z}},\end{aligned} \tag{293}$$

let $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ be a continuous function, let $\mu \colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma \colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$ be functions which satisfy for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that

$$\mu(\lambda x + y) + \lambda \mu(0) = \lambda \mu(x) + \mu(y), \tag{294}$$

$$\sigma(\lambda x + y) + \lambda\sigma(0) = \lambda\sigma(x) + \sigma(y), \tag{295}$$

and $\|\mu(x)\| + \|\!|\sigma(x)|\!\| \leq L(1 + \|x\|)$, *and let* $(\phi_\delta)_{\delta\in(0,r]} \subseteq \mathcal{N}$ *satisfy for all* $\delta \in (0, r]$, $x \in \mathbb{R}^d$ *that* $\mathcal{P}(\phi_\delta) \leq a\,\delta^{-\mathbf{z}}$, $\mathcal{R}(\phi_\delta) \in \mathcal{M}_d$, $|(\mathcal{R}(\phi_\delta))(x)| \leq c\,(1 + \|x\|^{\mathbf{v}})$, *and*

$$|\varphi(x) - (\mathcal{R}(\phi_\delta))(x)| \leq b\,\delta\,(1 + \|x\|^{\mathbf{v}}). \tag{296}$$

Then

(i) there exists a unique continuous function $u\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ *which satisfies that* $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty$, *which satisfies for all* $x \in \mathbb{R}^d$ *that* $u(0, x) = \varphi(x)$, *and which satisfies that* $u|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u)(t, x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t, x)\big) + \langle(\nabla_x u)(t, x), \mu(x)\rangle \tag{297}$$

for $(t, x) \in (0, T) \times \mathbb{R}^d$ *and*

(ii) there exist $(\psi_\varepsilon)_{\varepsilon\in(0,R]} \subseteq \mathcal{N}$ *such that for all* $\varepsilon \in (0, R]$ *it holds that* $\mathcal{P}(\psi_\varepsilon) \leq \mathbf{C}\,a\,b^{\mathbf{z}}\,c^4\,\varepsilon^{-4-\mathbf{z}}$, $\mathscr{P}(\psi_\varepsilon) \leq \mathbf{C}\,a\,b^{\mathbf{z}}\,c^2\,\varepsilon^{-2-\mathbf{z}}$, $\mathcal{R}(\psi_\varepsilon) \in \mathcal{M}_d$, *and*

$$\left[\int_{\mathbb{R}^d} |u(T, x) - (\mathcal{R}(\psi_\varepsilon))(x)|^p\ \nu(dx)\right]^{1/p} \leq \varepsilon. \tag{298}$$

Proof of Proposition 3.10. Throughout this proof let $\varepsilon \in (0, R]$ and let $C \in [0, \infty)$ be given by

$$C = 2(p-1)^{1/2}\exp\big(3\mathbf{v}(1+L^2T(\sqrt{T}+\mathbf{v}p)^2)\big)\big(1+\big[\textstyle\int_{\mathbb{R}^d}\|x\|^{p\mathbf{v}}\,\nu(dx)\big]^{1/p}\big). \tag{299}$$

Note that (296) implies that φ is an at most polynomially growing function. Combining this, (294), and (295) with item (i) in Corollary 2.23 (with $d = d$, $m = d$, $T = T$, $\varphi = \varphi$, $\mu = \mu$, $\sigma = \sigma$ in the notation of Corollary 2.23) establishes that there exists a unique continuous function $u\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ which satisfies for all $x \in \mathbb{R}^d$ that $u(0, x) = \varphi(x)$, which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty$, and which satisfies that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u)(t, x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t, x)\big) + \langle(\nabla_x u)(t, x), \mu(x)\rangle \tag{300}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$. This proves item (i). Next observe that (296) ensures that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \left|\varphi(x) - \big(\mathcal{R}(\phi_{\min\{b^{-1}C^{-1}\varepsilon,r\}})\big)(x)\right| &\leq \min\{b^{-1}C^{-1}\varepsilon, r\}\, b(1+\|x\|^{\mathbf{v}}) \\ &\leq b^{-1}C^{-1}\varepsilon\, b(1+\|x\|^{\mathbf{v}}) \\ &= C^{-1}\varepsilon(1+\|x\|^{\mathbf{v}}). \end{aligned} \tag{301}$$

Lemma 3.9 (with $d=d$, $T=T$, $\varepsilon=\varepsilon$, $L=L$, $c=c$, $\mathbf{v}=\mathbf{v}$, $p=p$, $\nu=\nu$, $\varphi=\varphi$, $\mu=\mu$, $\sigma=\sigma$, $\phi=\phi_{\min\{b^{-1}C^{-1}\varepsilon,r\}}$ in the notation of Lemma 3.9) hence assures that there exists $\psi \in \mathcal{N}$ such that

$$\mathcal{P}(\psi) \leq c^4\, 4(\max\{C,\varepsilon\})^4\, \mathcal{P}(\phi_{\min\{b^{-1}C^{-1}\varepsilon,r\}})\, \varepsilon^{-4}, \tag{302}$$

$$\mathscr{P}(\psi) \leq c^2\, 4(\max\{C,\varepsilon\})^4\, \mathcal{P}(\phi_{\min\{b^{-1}C^{-1}\varepsilon,r\}})\, \varepsilon^{-2}, \tag{303}$$

$$\mathcal{R}(\psi) \in \mathcal{M}_d, \qquad \text{and} \qquad \left[\int_{\mathbb{R}^d} |u(T,x) - (\mathcal{R}(\psi))(x)|^p\, \nu(dx)\right]^{1/p} \leq \varepsilon. \tag{304}$$

Moreover, note that the fact that $b, C \geq 1$ and the fact that $\varepsilon R^{-1} \leq 1$ assures that

$$\begin{aligned} \min\{b^{-1}C^{-1}\varepsilon, r\} &\geq \min\left\{b^{-1}C^{-1}\varepsilon, r\, b^{-1}C^{-1}\varepsilon R^{-1}\right\} \\ &= \min\{1, \tfrac{r}{R}\}\, b^{-1}C^{-1}\varepsilon. \end{aligned} \tag{305}$$

This, the fact that $C \geq 1$, the fact that $\varepsilon \in (0,R]$, (293), and (299) ensure that

$$\begin{aligned} &4(\max\{C,\varepsilon\})^4(\min\{b^{-1}C^{-1}\varepsilon, r\})^{-\mathbf{z}} \\ &\leq 4(\max\{C,R\})^4\left(\min\{1,\tfrac{r}{R}\}\, b^{-1}C^{-1}\varepsilon\right)^{-\mathbf{z}} \\ &= 4(\max\{C,R\})^4(\max\{1,\tfrac{R}{r}\})^{\mathbf{z}}\, b^{\mathbf{z}}\, C^{\mathbf{z}}\, \varepsilon^{-\mathbf{z}} \\ &\leq 4(\max\{C,R\})^{4+\mathbf{z}}(\max\{1,\tfrac{R}{r}\})^{4+\mathbf{z}}\, b^{\mathbf{z}}\, \varepsilon^{-\mathbf{z}} \\ &= \mathbf{C}\, b^{\mathbf{z}}\, \varepsilon^{-\mathbf{z}}. \end{aligned} \tag{306}$$

Combining this with the hypothesis that for all $\delta \in (0,r]$ it holds that $\mathcal{P}(\phi_\delta) \leq a\delta^{-\mathbf{z}}$ and (302) demonstrates that

$$\begin{aligned} \mathcal{P}(\psi) &\leq c^4\, 4(\max\{C,\varepsilon\})^4\, a\,(\min\{b^{-1}C^{-1}\varepsilon, r\})^{-\mathbf{z}}\, \varepsilon^{-4} \\ &\leq c^4\, a\, \mathbf{C}\, b^{\mathbf{z}}\, \varepsilon^{-\mathbf{z}}\varepsilon^{-4} = \mathbf{C}\, a\, b^{\mathbf{z}}\, c^4\, \varepsilon^{-4-\mathbf{z}}. \end{aligned} \tag{307}$$

Furthermore, observe that the hypothesis that for all $\delta \in (0,r]$ it holds that $\mathcal{P}(\phi_\delta) \leq a\delta^{-\mathbf{z}}$, (303), and (306) demonstrate that

$$\begin{aligned} \mathscr{P}(\psi) &\leq c^2\, 4(\max\{C,\varepsilon\})^4\, a\, (\min\{b^{-1}C^{-1}\varepsilon, r\})^{-\mathbf{z}}\, \varepsilon^{-2} \\ &\leq c^2\, a\, \mathbf{C}\, b^{\mathbf{z}}\, \varepsilon^{-\mathbf{z}}\, \varepsilon^{-2} = \mathbf{C}\, a\, b^{\mathbf{z}}\, c^2\, \varepsilon^{-2-\mathbf{z}}. \end{aligned} \tag{308}$$

Combining this, (304), and (307) establishes item (ii). The proof of Proposition 3.10 is thus completed. □

Corollary 3.11. *Assume Setting 3.7, let $d \in \mathbb{N}$, $T, r, R \in (0,\infty)$, $L, \mathfrak{C}, v, w, z, \mathbf{z} \in [0,\infty)$, $\mathfrak{c} \in [1,\infty)$, $\mathbf{v}, p \in [2,\infty)$, let $\langle \cdot,\cdot \rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, let $\|\cdot\| \colon \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, let $|||\cdot||| \colon \mathbb{R}^{d\times d} \to [0,\infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times d}$, let $\nu \colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a probability measure, assume that*

$$\mathfrak{C} = \Big[\mathfrak{c} \max\{1, \tfrac{R}{r}\} \max\Big\{2(p-1)^{1/2} \exp\big(3\mathbf{v}(1 + L^2 T(\sqrt{T} + \mathbf{v}p)^2)\big) \Big(1 + \Big[\textstyle\int_{\mathbb{R}^d} \|x\|^{p\mathbf{v}}\, \nu(dx)\Big]^{1/p}\Big), R\Big\}\Big]^{5+\mathbf{z}}, \tag{309}$$

let $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ be a continuous function, let $\mu \colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma \colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be functions which satisfy for all $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that

$$\mu(\lambda x + y) + \lambda\mu(0) = \lambda\mu(x) + \mu(y), \tag{310}$$

$$\sigma(\lambda x + y) + \lambda\sigma(0) = \lambda\sigma(x) + \sigma(y), \tag{311}$$

and $\|\mu(x)\| + |||\sigma(x)||| \leq L(1 + \|x\|)$, and let $(\phi_\delta)_{\delta\in(0,r]} \subseteq \mathcal{N}$ satisfy for all $\delta \in (0,r]$, $x \in \mathbb{R}^d$ that $\mathcal{P}(\phi_\delta) \leq \mathfrak{c}\, d^z \delta^{-\mathbf{z}}$, $\mathcal{R}(\phi_\delta) \in \mathcal{M}_d$, $|(\mathcal{R}(\phi_\delta))(x)| \leq \mathfrak{c}\, d^v(1 + \|x\|^{\mathbf{v}})$, and

$$|\varphi(x) - (\mathcal{R}(\phi_\delta))(x)| \leq \mathfrak{c}\, d^w\, \delta\, (1 + \|x\|^{\mathbf{v}}). \tag{312}$$

Then

(i) *there exists a unique continuous function $u \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty$, which satisfies for all $x \in \mathbb{R}^d$ that $u(0,x) = \varphi(x)$, and which satisfies that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle (\nabla_x u)(t,x), \mu(x)\rangle \tag{313}$$

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

(ii) *there exist* $(\psi_\varepsilon)_{\varepsilon\in(0,R]} \subseteq \mathcal{N}$ *such that for all* $\varepsilon \in (0,R]$ *it holds that* $\mathcal{P}(\psi_\varepsilon) \leq \mathfrak{C}\, d^{z+w\mathbf{z}+4v}\, \varepsilon^{-4-\mathbf{z}}$, $\mathscr{P}(\psi_\varepsilon) \leq \mathfrak{C}\, d^{z+w\mathbf{z}+2v}\, \varepsilon^{-2-\mathbf{z}}$, $\mathcal{R}(\psi_\varepsilon) \in \mathcal{M}_d$, *and*

$$\left[\int_{\mathbb{R}^d} |u(T,x) - (\mathcal{R}(\psi_\varepsilon))(x)|^p \, \nu(dx)\right]^{1/p} \leq \varepsilon. \tag{314}$$

Proof of Corollary 3.11. Throughout this proof let $C, \mathbf{C} \in [0,\infty)$ be given by

$$C = 2(p-1)^{1/2} \exp\big(3\mathbf{v}(1+L^2T(\sqrt{T}+\mathbf{v}p)^2)\big)\big(1+\big[\textstyle\int_{\mathbb{R}^d} \|x\|^{p\mathbf{v}}\,\nu(dx)\big]^{1/p}\big) \tag{315}$$

and

$$\mathbf{C} = 4\big[\max\{1, \tfrac{R}{r}\} \max\{C, R\}\big]^{4+\mathbf{z}}. \tag{316}$$

Note that (312) implies that φ is an at most polynomially growing function. Combining this, (310), and (311) with item (i) in Corollary 2.23 (with $d = d$, $m = d$, $T = T$, $\varphi = \varphi$, $\mu = \mu$, $\sigma = \sigma$ in the notation of Corollary 2.23) establishes that there exists a unique continuous function $u\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|^q} < \infty$, which satisfies for all $x \in \mathbb{R}^d$ that $u(0,x) = \varphi(x)$, and which satisfies that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t,x)\big) + \langle(\nabla_x u)(t,x), \mu(x)\rangle \tag{317}$$

for $(t,x) \in (0,T)\times\mathbb{R}^d$. This proves item (i). Next observe that Proposition 3.10 (with $d = d$, $T = T$, $a = \mathfrak{c}\, d^z$, $r = r$, $R = R$, $L = L$, $\mathbf{z} = \mathbf{z}$, $b = \mathfrak{c}\, d^w$, $c = \mathfrak{c}\, d^v$, $\mathbf{v} = \mathbf{v}$, $p = p$, $\nu = \nu$, $\varphi = \varphi$, $\mu = \mu$, $\sigma = \sigma$, $(\phi_\delta)_{\delta\in(0,r]} = (\phi_\delta)_{\delta\in(0,r]}$ in the notation of Proposition 3.10) proves that there exist $(\psi_\varepsilon)_{\varepsilon\in(0,R]} \subseteq \mathcal{N}$ such that for all $\varepsilon \in (0,R]$ it holds that

$$\mathcal{P}(\psi_\varepsilon) \leq \mathbf{C}\mathfrak{c}\, d^z(\mathfrak{c}\, d^w)^{\mathbf{z}}(\mathfrak{c}\, d^v)^4\varepsilon^{-4-\mathbf{z}}, \tag{318}$$

$$\mathscr{P}(\psi_\varepsilon) \leq \mathbf{C}\mathfrak{c}\, d^z(\mathfrak{c}\, d^w)^{\mathbf{z}}(\mathfrak{c}\, d^v)^2\varepsilon^{-2-\mathbf{z}}, \tag{319}$$

$$\mathcal{R}(\psi_\varepsilon) \in \mathcal{M}_d, \qquad \text{and} \qquad \left[\int_{\mathbb{R}^d} |u(T,x) - (\mathcal{R}(\psi_\varepsilon))(x)|^p \, \nu(dx)\right]^{1/p} \leq \varepsilon. \tag{320}$$

Furthermore, note that the fact that $p, \mathbf{v} \geq 2$ assures that

$$C \geq 2(p-1)^{1/2} \exp\big(3\mathbf{v}(1+L^2T(\sqrt{T}+\mathbf{v}p)^2)\big) \geq 2\exp(6) \geq 4. \tag{321}$$

This implies that

$$\begin{aligned}
\mathfrak{c}^{5+\mathbf{z}}\,\mathbf{C} &= \mathfrak{c}^{5+\mathbf{z}}\,4\big[\max\{1,\tfrac{R}{r}\}\max\{C,R\}\big]^{4+\mathbf{z}} \\
&\le \mathfrak{c}^{5+\mathbf{z}}\,C\big[\max\{1,\tfrac{R}{r}\}\max\{C,R\}\big]^{4+\mathbf{z}} \\
&\le \mathfrak{c}^{5+\mathbf{z}}\big[\max\{1,\tfrac{R}{r}\}\max\{C,R\}\big]^{5+\mathbf{z}} \\
&= \mathfrak{C}.
\end{aligned} \tag{322}$$

This and (318) demonstrate that for all $\varepsilon \in (0,R]$ it holds that

$$\mathcal{P}(\psi_\varepsilon) \le \mathbf{C}\,\mathfrak{c}^{5+\mathbf{z}} d^{z+w\mathbf{z}+4v}\varepsilon^{-4-\mathbf{z}} \le \mathfrak{C}\,d^{z+w\mathbf{z}+4v}\varepsilon^{-4-\mathbf{z}}. \tag{323}$$

Moreover, observe that (319), (322), and the fact that $\mathfrak{c} \ge 1$ assure that

$$\begin{aligned}
\mathscr{P}(\psi_\varepsilon) &\le \mathbf{C}\,\mathfrak{c}^{3+\mathbf{z}} d^{z+w\mathbf{z}+2v}\varepsilon^{-2-\mathbf{z}} \\
&\le \mathbf{C}\,\mathfrak{c}^{5+\mathbf{z}} d^{z+w\mathbf{z}+2v}\varepsilon^{-2-\mathbf{z}} \le \mathfrak{C}\,d^{z+w\mathbf{z}+2v}\,\varepsilon^{-2-\mathbf{z}}.
\end{aligned} \tag{324}$$

Combining this, (320), and (323) establishes item (ii). The proof of Corollary 3.11 is thus completed. □

Proposition 3.12. *Assume Setting 3.7, let $\mathcal{I}$ be a set, let $\mathfrak{d} = (\mathfrak{d}_i)_{i\in\mathcal{I}} \colon \mathcal{I} \to \mathbb{N}$ be a function, for every $d \in \mathbb{N}$ let $\langle\cdot,\cdot\rangle_{\mathbb{R}^d} \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} \colon \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm and let $|||\cdot|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \colon \mathbb{R}^{d\times d} \to [0,\infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times d}$, let $T, r, R \in (0,\infty)$, $\mathfrak{C}, L, v, w, z, \mathbf{z}, \theta \in [0,\infty)$, $\mathfrak{c} \in [1,\infty)$, $\mathbf{v}, p \in [2,\infty)$, let $\nu_d \colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$, $d \in \operatorname{Im}(\mathfrak{d})$, be probability measures, assume that*

$$\mathfrak{C} = \Big[\mathfrak{c}\max\{1,\tfrac{R}{r}\}\max\Big\{2(p-1)^{1/2}\exp\!\big(3\mathbf{v}(1+L^2T(\sqrt{T}+\mathbf{v}p)^2)\big) \\ \Big(1+\sup_{i\in\mathcal{I}}\Big[(\mathfrak{d}_i)^{-\theta}\Big[\textstyle\int_{\mathbb{R}^{\mathfrak{d}_i}}\|x\|_{\mathbb{R}^{\mathfrak{d}_i}}^{p\mathbf{v}}\,\nu_{\mathfrak{d}_i}(dx)\Big]^{1/p}\Big]\Big), R\Big\}\Big]^{5+\mathbf{z}}, \tag{325}$$

let $\varphi_i \colon \mathbb{R}^{\mathfrak{d}_i} \to \mathbb{R}$, $i \in \mathcal{I}$, be continuous functions, let $\mu_i \colon \mathbb{R}^{\mathfrak{d}_i} \to \mathbb{R}^{\mathfrak{d}_i}$, $i \in \mathcal{I}$, and $\sigma_i \colon \mathbb{R}^{\mathfrak{d}_i} \to \mathbb{R}^{\mathfrak{d}_i\times\mathfrak{d}_i}$, $i \in \mathcal{I}$, be functions which satisfy for all $i \in \mathcal{I}$, $x, y \in \mathbb{R}^{\mathfrak{d}_i}$, $\lambda \in \mathbb{R}$ that

$$\mu_i(\lambda x + y) + \lambda\mu_i(0) = \lambda\mu_i(x) + \mu_i(y), \tag{326}$$

$$\sigma_i(\lambda x + y) + \lambda\sigma_i(0) = \lambda\sigma_i(x) + \sigma_i(y), \tag{327}$$

and $\|\mu_i(x)\|_{\mathbb{R}^{\mathfrak{d}_i}} + |||\sigma_i(x)|||_{\mathrm{HS}(\mathbb{R}^{\mathfrak{d}_i},\mathbb{R}^{\mathfrak{d}_i})} \le L(1+\|x\|_{\mathbb{R}^{\mathfrak{d}_i}})$, and let $(\phi_{i,\delta})_{i\in\mathcal{I},\,\delta\in(0,r]} \subseteq \mathcal{N}$ satisfy for all $i \in \mathcal{I}$, $\delta \in (0,r]$, $x \in \mathbb{R}^{\mathfrak{d}_i}$ that $\mathcal{P}(\phi_{i,\delta}) \le \mathfrak{c}\,(\mathfrak{d}_i)^z\delta^{-\mathbf{z}}$, $\mathcal{R}(\phi_{i,\delta}) \in \mathcal{M}_{\mathfrak{d}_i}$, $|(\mathcal{R}(\phi_{i,\delta}))(x)| \le \mathfrak{c}\,(\mathfrak{d}_i)^v(1+\|x\|_{\mathbb{R}^{\mathfrak{d}_i}}^{\mathbf{v}})$, and

$$|\varphi_i(x) - (\mathcal{R}(\phi_{i,\delta}))(x)| \le \mathfrak{c}\,(\mathfrak{d}_i)^w\,\delta\,(1+\|x\|_{\mathbb{R}^{\mathfrak{d}_i}}^{\mathbf{v}}). \tag{328}$$

Then

(i) there exist unique continuous functions $u_i\colon [0,T]\times\mathbb{R}^{\mathfrak{d}_i}\to\mathbb{R}$, $i\in\mathcal{I}$, which satisfy for all $i\in\mathcal{I}$, $x\in\mathbb{R}^{\mathfrak{d}_i}$ that $u_i(0,x)=\varphi_i(x)$, which satisfy for all $i\in\mathcal{I}$ that $\inf_{q\in(0,\infty)}\sup_{(t,x)\in[0,T]\times\mathbb{R}^{\mathfrak{d}_i}}\frac{|u_i(t,x)|}{1+\|x\|^q_{\mathbb{R}^{\mathfrak{d}_i}}}<\infty$, and which satisfy for all $i\in\mathcal{I}$ that $u_i|_{(0,T)\times\mathbb{R}^{\mathfrak{d}_i}}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u_i)(t,x)=\tfrac{1}{2}\operatorname{Trace}\big(\sigma_i(x)[\sigma_i(x)]^*(\operatorname{Hess}_x u_i)(t,x)\big)\\ +\langle(\nabla_x u_i)(t,x),\mu_i(x)\rangle_{\mathbb{R}^d} \tag{329}$$

for $(t,x)\in(0,T)\times\mathbb{R}^{\mathfrak{d}_i}$ and

(ii) there exist $(\psi_{i,\varepsilon})_{i\in\mathcal{I},\,\varepsilon\in(0,R]}\subseteq\mathcal{N}$ such that for all $i\in\mathcal{I}$, $\varepsilon\in(0,R]$ it holds that

$$\mathcal{P}(\psi_{i,\varepsilon})\leq\mathfrak{C}\,(\mathfrak{d}_i)^{(5+\mathbf{z})\theta+z+w\mathbf{z}+4v}\,\varepsilon^{-4-\mathbf{z}}, \tag{330}$$

$$\mathscr{P}(\psi_{i,\varepsilon})\leq\mathfrak{C}\,(\mathfrak{d}_i)^{(5+\mathbf{z})\theta+z+w\mathbf{z}+2v}\,\varepsilon^{-2-\mathbf{z}},\qquad \mathcal{R}(\psi_{i,\varepsilon})\in\mathcal{M}_{\mathfrak{d}_i}, \tag{331}$$

$$\text{and}\qquad \left[\int_{\mathbb{R}^d}|u_i(T,x)-(\mathcal{R}(\psi_{i,\varepsilon}))(x)|^p\;\nu_{\mathfrak{d}_i}(dx)\right]^{1/p}\leq\varepsilon. \tag{332}$$

Proof of Proposition 3.12. Throughout this proof let $i\in\mathcal{I}$, let $c_0\in(0,\infty)$ be given by

$$c_0=2(p-1)^{1/2}\exp\big(3\mathbf{v}(1+L^2T(\sqrt{T}+\mathbf{v}p)^2)\big), \tag{333}$$

and let $\mathcal{C}\in[0,\infty)$ be given by

$$\mathcal{C}=\big[\mathfrak{c}\max\{1,\tfrac{R}{r}\}\max\big\{R,c_0\big(1+\big[\textstyle\int_{\mathbb{R}^{\mathfrak{d}_i}}\|x\|^{p\mathbf{v}}_{\mathbb{R}^{\mathfrak{d}_i}}\,\nu_{\mathfrak{d}_i}(dx)\big]^{1/p}\big)\big\}\big]^{5+\mathbf{z}}. \tag{334}$$

Note that the fact that for all $\theta\in[0,\infty)$, $i\in\mathcal{I}$ it holds that $(\mathfrak{d}_i)^{-\theta}\leq 1$ implies that

$$\begin{aligned}
&\max\big\{c_0\big(1+\big[\textstyle\int_{\mathbb{R}^{\mathfrak{d}_i}}\|x\|^{p\mathbf{v}}_{\mathbb{R}^{\mathfrak{d}_i}}\,\nu_{\mathfrak{d}_i}(dx)\big]^{1/p}\big),R\big\}\\
&\leq\max\left\{c_0\left((\mathfrak{d}_i)^{\theta}(\mathfrak{d}_i)^{-\theta}+(\mathfrak{d}_i)^{\theta}\sup_{i\in\mathcal{I}}\left[(\mathfrak{d}_i)^{-\theta}\left(\textstyle\int_{\mathbb{R}^{\mathfrak{d}_i}}\|x\|^{p\mathbf{v}}_{\mathbb{R}^{\mathfrak{d}_i}}\,\nu_{\mathfrak{d}_i}(dx)\right)^{1/p}\right]\right),R\right\}\\
&\leq\max\left\{c_0(\mathfrak{d}_i)^{\theta}\left(1+\sup_{i\in\mathcal{I}}\left[(\mathfrak{d}_i)^{-\theta}\left(\textstyle\int_{\mathbb{R}^{\mathfrak{d}_i}}\|x\|^{p\mathbf{v}}_{\mathbb{R}^{\mathfrak{d}_i}}\,\nu_{\mathfrak{d}_i}(dx)\right)^{1/p}\right]\right),(\mathfrak{d}_i)^{\theta}(\mathfrak{d}_i)^{-\theta}R\right\}\\
&\leq(\mathfrak{d}_i)^{\theta}\max\left\{c_0\left(1+\sup_{i\in\mathcal{I}}\left[(\mathfrak{d}_i)^{-\theta}\left(\textstyle\int_{\mathbb{R}^{\mathfrak{d}_i}}\|x\|^{p\mathbf{v}}_{\mathbb{R}^{\mathfrak{d}_i}}\,\nu_{\mathfrak{d}_i}(dx)\right)^{1/p}\right]\right),R\right\}.
\end{aligned} \tag{335}$$

Therefore, (334), (333), and (325) ensure that

$$
\begin{aligned}
\mathcal{C} &= \Big[\mathfrak{c}\max\{1,\tfrac{R}{r}\}\max\Big\{c_0\big(1+\big[\textstyle\int_{\mathbb{R}^{\mathfrak{d}_i}}\|x\|_{\mathbb{R}^{\mathfrak{d}_i}}^{p\mathbf{v}}\,\nu_{\mathfrak{d}_i}(dx)\big]^{1/p}\big),R\Big\}\Big]^{5+\mathbf{z}} \\
&\leq (\mathfrak{d}_i)^{\theta(5+\mathbf{z})}\Bigg[\mathfrak{c}\max\{1,\tfrac{R}{r}\} \\
&\qquad \max\Bigg\{c_0\left(1+\sup_{i\in\mathcal{I}}\left[(\mathfrak{d}_i)^{-\theta}\left(\textstyle\int_{\mathbb{R}^{\mathfrak{d}_i}}\|x\|_{\mathbb{R}^{\mathfrak{d}_i}}^{p\mathbf{v}}\,\nu_{\mathfrak{d}_i}(dx)\right)^{1/p}\right]\right),R\Bigg\}\Bigg]^{5+\mathbf{z}} \\
&\leq (\mathfrak{d}_i)^{\theta(5+\mathbf{z})}\Bigg[\mathfrak{c}\max\{1,\tfrac{R}{r}\}\max\Bigg\{2(p-1)^{1/2}\exp\big(3\mathbf{v}(1+L^2T(\sqrt{T}+\mathbf{v}p)^2)\big) \\
&\qquad \left(1+\sup_{i\in\mathcal{I}}\left[(\mathfrak{d}_i)^{-\theta}\left(\textstyle\int_{\mathbb{R}^{\mathfrak{d}_i}}\|x\|_{\mathbb{R}^{\mathfrak{d}_i}}^{p\mathbf{v}}\,\nu_{\mathfrak{d}_i}(dx)\right)^{1/p}\right]\right),R\Bigg\}\Bigg]^{5+\mathbf{z}} \\
&= (\mathfrak{d}_i)^{\theta(5+\mathbf{z})}\mathfrak{C}.
\end{aligned}
\tag{336}
$$

Moreover, observe that (328) implies that φ_i is an at most polynomially growing function. Combining this, (326), and (327) with item (i) in Corollary 2.23 (with $d=\mathfrak{d}_i$, $m=\mathfrak{d}_i$, $T=T$, $\varphi=\varphi_i$, $\mu=\mu_i$, $\sigma=\sigma_i$ in the notation of Corollary 2.23) establishes that there exists a unique continuous function $u_i\colon[0,T]\times\mathbb{R}^{\mathfrak{d}_i}\to\mathbb{R}$ which satisfies for all $x\in\mathbb{R}^{\mathfrak{d}_i}$ that $u_i(0,x)=\varphi_i(x)$, which satisfies that $\inf_{q\in(0,\infty)}\sup_{(t,x)\in[0,T]\times\mathbb{R}^{\mathfrak{d}_i}}\frac{|u_i(t,x)|}{1+\|x\|_{\mathbb{R}^{\mathfrak{d}_i}}^q}<\infty$, and which satisfies that $u_i|_{(0,T)\times\mathbb{R}^{\mathfrak{d}_i}}$ is a viscosity solution of

$$
(\tfrac{\partial}{\partial t}u_i)(t,x)=\tfrac{1}{2}\operatorname{Trace}\big(\sigma_i(x)[\sigma_i(x)]^*(\operatorname{Hess}_x u_i)(t,x)\big)+\langle(\nabla_x u_i)(t,x),\mu_i(x)\rangle_{\mathbb{R}^d}
\tag{337}
$$

for $(t,x)\in(0,T)\times\mathbb{R}^{\mathfrak{d}_i}$. This proves item (i). Next observe that Corollary 3.11 (with $d=\mathfrak{d}_i$, $T=T$, $r=r$, $R=R$, $v=v$, $w=w$, $z=z$, $\mathbf{z}=\mathbf{z}$, $\mathfrak{c}=\mathfrak{c}$, $\mathbf{v}=\mathbf{v}$, $p=p$, $\nu=\nu_{\mathfrak{d}_i}$, $\varphi=\varphi_i$, $\mu=\mu_i$, $\sigma=\sigma_i$, $(\phi_\delta)_{\delta\in(0,r]}=(\phi_{i,\delta})_{\delta\in(0,r]}$ in the notation of Corollary 3.11) demonstrates that there exists $(\psi_{i,\varepsilon})_{\varepsilon\in(0,R]}\subseteq\mathcal{N}$ such that for all $\varepsilon\in(0,R]$ it holds that

$$
\mathcal{P}(\psi_{i,\varepsilon})\leq\mathcal{C}\,(\mathfrak{d}_i)^{z+w\mathbf{z}+4v}\,\varepsilon^{-4-\mathbf{z}},
\tag{338}
$$

$$
\mathscr{P}(\psi_{i,\varepsilon})\leq\mathcal{C}\,(\mathfrak{d}_i)^{z+w\mathbf{z}+2v}\,\varepsilon^{-2-\mathbf{z}},\qquad \mathcal{R}(\psi_{i,\varepsilon})\in\mathcal{M}_{\mathfrak{d}_i},
\tag{339}
$$

$$
\text{and}\qquad \left[\int_{\mathbb{R}^d}|u_i(T,x)-(\mathcal{R}(\psi_{i,\varepsilon}))(x)|^p\,\nu_{\mathfrak{d}_i}(dx)\right]^{1/p}\leq\varepsilon.
\tag{340}
$$

Combining this with (336) establishes item (ii). The proof of Proposition 3.12 is thus completed. □

Corollary 3.13. *Assume Setting 3.7, let $T, r, R \in (0,\infty)$, $\mathfrak{C}, L, v, w, z, \mathbf{z}, \theta \in [0,\infty)$, $\mathfrak{c} \in [1,\infty)$, $\mathbf{v}, p \in [2,\infty)$, for every $d \in \mathbb{N}$ let $\langle \cdot,\cdot \rangle_{\mathbb{R}^d} \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} \colon \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm and let $|||\cdot|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \colon \mathbb{R}^{d\times d} \to [0,\infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times d}$, let $\nu_d \colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$, $d \in \mathbb{N}$, be probability measures, assume that*

$$\begin{aligned}\mathfrak{C} = \Big[\mathfrak{c}\max\{1, \tfrac{R}{r}\}\max\Big\{2(p-1)^{1/2}\exp\big(3\mathbf{v}(1 + L^2T(\sqrt{T} + \mathbf{v}p)^2)\big)\\ \big(1 + \sup\nolimits_{d\in\mathbb{N}}\big[d^{-\theta}\big[\textstyle\int_{\mathbb{R}^d}\|x\|_{\mathbb{R}^d}^{p\mathbf{v}}\,\nu_d(dx)\big]^{1/p}\big]\big), R\Big\}\Big]^{5+\mathbf{z}},\end{aligned} \tag{341}$$

let $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, be continuous functions, let $\mu_d \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, and $\sigma_d \colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, be functions which satisfy for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that

$$\mu_d(\lambda x + y) + \lambda\mu_d(0) = \lambda\mu_d(x) + \mu_d(y), \tag{342}$$

$$\sigma_d(\lambda x + y) + \lambda\sigma_d(0) = \lambda\sigma_d(x) + \sigma_d(y), \tag{343}$$

and $\|\mu_d(x)\|_{\mathbb{R}^d} + |||\sigma_d(x)|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \le L(1 + \|x\|_{\mathbb{R}^d})$, and let $(\phi_{d,\delta})_{d\in\mathbb{N},\,\delta\in(0,r]} \subseteq \mathcal{N}$ satisfy for all $d \in \mathbb{N}$, $\delta \in (0,r]$, $x \in \mathbb{R}^d$ that $\mathcal{P}(\phi_{d,\delta}) \le \mathfrak{c}\, d^z \delta^{-\mathbf{z}}$, $\mathcal{R}(\phi_{d,\delta}) \in \mathcal{M}_d$, $|(\mathcal{R}(\phi_{d,\delta}))(x)| \le \mathfrak{c}\, d^v(1 + \|x\|_{\mathbb{R}^d}^{\mathbf{v}})$, and

$$|\varphi_d(x) - (\mathcal{R}(\phi_{d,\delta}))(x)| \le \mathfrak{c}\, d^w\, \delta\,(1 + \|x\|_{\mathbb{R}^d}^{\mathbf{v}}). \tag{344}$$

Then

(i) *there exist unique continuous functions $u_d \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $u_d(0,x) = \varphi_d(x)$, which satisfy for all $d \in \mathbb{N}$ that $\inf_{q\in(0,\infty)}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $u_d|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of*

$$\begin{aligned}(\tfrac{\partial}{\partial t}u_d)(t,x) &= \tfrac{1}{2}\operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t,x)\big)\\ &\quad + \langle(\nabla_x u_d)(t,x), \mu_d(x)\rangle_{\mathbb{R}^d}\end{aligned} \tag{345}$$

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

(ii) *there exist $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\,\varepsilon\in(0,R]} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,R]$ it holds that*

$$\mathcal{P}(\psi_{d,\varepsilon}) \le \mathfrak{C}\, d^{(5+\mathbf{z})\theta + z + w\mathbf{z} + 4v}\,\varepsilon^{-4-\mathbf{z}}, \tag{346}$$

$$\mathscr{P}(\psi_{d,\varepsilon}) \le \mathfrak{C}\, d^{(5+\mathbf{z})\theta + z + w\mathbf{z} + 2v}\,\varepsilon^{-2-\mathbf{z}}, \qquad \mathcal{R}(\psi_{d,\varepsilon}) \in \mathcal{M}_d, \tag{347}$$

$$\text{and} \qquad \left[\int_{\mathbb{R}^d}|u_d(T,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p\;\nu_d(dx)\right]^{1/p} \le \varepsilon. \tag{348}$$

Proof of Corollary 3.13. Observe that Proposition 3.12 (with $\mathcal{I} = \mathbb{N}$, $\mathfrak{d} = \mathrm{id}_{\mathbb{N}}$, $T = T$, $r = r$, $R = R$, $L = L$, $v = v$, $w = w$, $z = z$, $\mathbf{z} = \mathbf{z}$, $\theta = \theta$, $\mathfrak{c} = \mathfrak{c}$, $\mathbf{v} = \mathbf{v}$, $p = p$, $(\nu_d)_{d \in \mathrm{Im}(\mathfrak{d})} = (\nu_d)_{d \in \mathbb{N}}$, $(\varphi_i)_{i \in \mathcal{I}} = (\varphi_d)_{d \in \mathbb{N}}$, $(\mu_i)_{i \in \mathcal{I}} = (\mu_d)_{d \in \mathbb{N}}$, $(\sigma_i)_{i \in \mathcal{I}} = (\sigma_d)_{d \in \mathbb{N}}$, $(\phi_{i,\delta})_{i \in \mathcal{I},\, \delta \in (0,r]} = (\phi_{d,\delta})_{d \in \mathbb{N},\, \delta \in (0,r]}$ in the notation of Proposition 3.12) establishes items (i)–(ii). The proof of Corollary 3.13 is thus completed. ☐

3.6 Artificial neural networks with continuous activation functions

In this subsection we establish in Theorem 3.14 below the main result of this article. Theorem 3.14 proves, roughly speaking, that fully-connected artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes PDEs (see (357) in item (ii) in Theorem 3.14 for details). In Theorem 3.14 the approximation error between the solution of the PDE and the artificial neural network is measured by means of L^p-norms with respect to the general probability measures ν_d, $d \in \mathbb{N}$, in Theorem 3.14. To make Theorem 3.14 easier accessible, we derive a simplified and specialized version of Theorem 3.14 in Corollary 3.16 below. In particular, in Corollary 3.16 below we specialize Theorem 3.14 to the case where the general probability measures ν_d, $d \in \mathbb{N}$, are nothing else but the continuous uniform distribution on $[0,1]^d$. Our proof of Corollary 3.16 uses the elementary estimate in Lemma 3.15 below. For the sake of completeness we also present in this subsection a detailed proof of Lemma 3.15.

Theorem 3.14. *Let $T, r, R \in (0,\infty)$, $v, w, z, \mathbf{z}, \theta \in [0,\infty)$, $\mathfrak{c} \in [1,\infty)$, $\mathbf{v}, p \in [2,\infty)$, for every $d \in \mathbb{N}$ let $\langle \cdot, \cdot \rangle_{\mathbb{R}^d} \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} \colon \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm and let $|||\cdot|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \colon \mathbb{R}^{d\times d} \to [0,\infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times d}$, let $\nu_d \colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$, $d \in \mathbb{N}$, be probability measures, let $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, be continuous functions, let $\mu_d \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, and $\sigma_d \colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, be functions which satisfy for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that*

$$\mu_d(\lambda x + y) + \lambda \mu_d(0) = \lambda \mu_d(x) + \mu_d(y), \tag{349}$$

$$\sigma_d(\lambda x + y) + \lambda \sigma_d(0) = \lambda \sigma_d(x) + \sigma_d(y), \tag{350}$$

and $\|\mu_d(x)\|_{\mathbb{R}^d} + \|\!|\sigma_d(x)|\!\|_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \leq \mathfrak{c}(1 + \|x\|_{\mathbb{R}^d})$, *let*

$$\mathcal{N} = \cup_{\mathcal{L}\in\{2,3,\ldots\}} \cup_{(l_0,l_1,\ldots,l_\mathcal{L})\in((\mathbb{N}^\mathcal{L})\times\{1\})} \left(\times_{k=1}^{\mathcal{L}} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\right), \tag{351}$$

assume that $\sup_{d\in\mathbb{N}} \left[d^{-\theta p} \int_{\mathbb{R}^d} \|x\|_{\mathbb{R}^d}^{p\mathbf{v}} \,\nu_d(dx)\right] < \infty$, *let* $\mathbf{A}_d \in C(\mathbb{R}^d, \mathbb{R}^d)$, $d \in \mathbb{N}$, *and* $\mathbf{a} \in C(\mathbb{R}, \mathbb{R})$ *be functions which satisfy for all* $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ *that*

$$\mathbf{A}_d(x) = (\mathbf{a}(x_1), \mathbf{a}(x_2), \ldots, \mathbf{a}(x_d)), \tag{352}$$

let $\mathcal{P}, \mathscr{P} \colon \mathcal{N} \to \mathbb{N}$ *and* $\mathcal{R} \colon \mathcal{N} \to \cup_{d=1}^{\infty} C(\mathbb{R}^d, \mathbb{R})$ *be the functions which satisfy for all* $\mathcal{L} \in \{2, 3, \ldots\}$, $(l_0, l_1, \ldots, l_\mathcal{L}) \in ((\mathbb{N}^\mathcal{L}) \times \{1\})$, $\Phi = ((W_1, B_1), \ldots, (W_\mathcal{L}, B_\mathcal{L})) = ((W_k^{(i,j)})_{i\in\{1,2,\ldots,l_k\},j\in\{1,2,\ldots,l_{k-1}\}}, (B_k^{(i)})_{i\in\{1,2,\ldots,l_k\}})_{k\in\{1,2,\ldots,\mathcal{L}\}} \in (\times_{k=1}^{\mathcal{L}} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $x_0 \in \mathbb{R}^{l_0}, x_1 \in \mathbb{R}^{l_1}, \ldots, x_{\mathcal{L}-1} \in \mathbb{R}^{l_{\mathcal{L}-1}}$ *with* $\forall\, k \in \mathbb{N} \cap (0, \mathcal{L}) \colon x_k = \mathbf{A}_{l_k}(W_k x_{k-1} + B_k)$ *that*

$$\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}), \qquad (\mathcal{R}(\Phi))(x_0) = W_\mathcal{L} x_{\mathcal{L}-1} + B_\mathcal{L}, \tag{353}$$

$$\mathscr{P}(\Phi) = \sum_{k=1}^{\mathcal{L}} \sum_{i=1}^{l_k} \left(\mathbb{1}_{\mathbb{R}\setminus\{0\}}(B_k^{(i)}) + \sum_{j=1}^{l_{k-1}} \mathbb{1}_{\mathbb{R}\setminus\{0\}}(W_k^{(i,j)}) \right), \tag{354}$$

and $\mathcal{P}(\Phi) = \sum_{k=1}^{\mathcal{L}} l_k(l_{k-1} + 1)$, *and let* $(\phi_{d,\delta})_{d\in\mathbb{N},\, \delta\in(0,r]} \subseteq \mathcal{N}$ *satisfy for all* $d \in \mathbb{N}$, $\delta \in (0, r]$, $x \in \mathbb{R}^d$ *that* $\mathcal{P}(\phi_{d,\delta}) \leq \mathfrak{c}\, d^z \delta^{-\mathbf{z}}$, $\mathcal{R}(\phi_{d,\delta}) \in C(\mathbb{R}^d, \mathbb{R})$, $|(\mathcal{R}(\phi_{d,\delta}))(x)| \leq \mathfrak{c}\, d^v (1 + \|x\|_{\mathbb{R}^d}^{\mathbf{v}})$, *and*

$$|\varphi_d(x) - (\mathcal{R}(\phi_{d,\delta}))(x)| \leq \mathfrak{c}\, d^w\, \delta\, (1 + \|x\|_{\mathbb{R}^d}^{\mathbf{v}}). \tag{355}$$

Then

(i) there exist unique continuous functions $u_d \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, *which satisfy for all* $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ *that* $u_d(0, x) = \varphi_d(x)$, *which satisfy for all* $d \in \mathbb{N}$ *that* $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, *and which satisfy for all* $d \in \mathbb{N}$ *that* $u_d|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$\begin{aligned}(\tfrac{\partial}{\partial t} u_d)(t, x) = \tfrac{1}{2} \operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^* (\operatorname{Hess}_x u_d)(t, x)\big) \\ + \langle (\nabla_x u_d)(t, x), \mu_d(x) \rangle_{\mathbb{R}^d}\end{aligned} \tag{356}$$

for $(t, x) \in (0, T) \times \mathbb{R}^d$ *and* *(ii) there exist* $\mathfrak{C} \in (0,\infty)$, $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\,\varepsilon\in(0,R]} \subseteq \mathcal{N}$ *such that for all* $d \in \mathbb{N}$, $\varepsilon \in (0,R]$ *it holds that* $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{(5+\mathbf{z})\theta+z+w\mathbf{z}+4v}\,\varepsilon^{-4-\mathbf{z}}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{(5+\mathbf{z})\theta+z+w\mathbf{z}+2v}\,\varepsilon^{-2-\mathbf{z}}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, *and*

$$\left[\int_{\mathbb{R}^d} |u_d(T,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \;\nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{357}$$

Proof of Theorem 3.14. Throughout this proof let $\mathfrak{C} \in (0,\infty)$ be given by

$$\begin{split}\mathfrak{C} = \Big[\mathfrak{c}\max\{1,\tfrac{R}{r}\}\max\Big\{R, 2(p-1)^{1/2}\exp\big(3\mathbf{v}(1+\mathfrak{c}^2T(\sqrt{T}+\mathbf{v}p)^2)\big)\\ \cdot\big(1+\sup\nolimits_{d\in\mathbb{N}}\big[d^{-\theta}\,\big[\textstyle\int_{\mathbb{R}^d}\|x\|_{\mathbb{R}^d}^{p\mathbf{v}}\,\nu_d(dx)\big]^{1/p}\,\big]\big)\Big\}\Big]^{5+\mathbf{z}}\end{split} \tag{358}$$

and for every $d \in \mathbb{N}$ let $\mathcal{M}_d$ be the set of all Borel measurable functions from $\mathbb{R}^d$ to $\mathbb{R}$. Note that Corollary 3.13 (with $T = T$, $r = r$, $R = R$, $L = \mathfrak{c}$, $v = v$, $w = w$, $z = z$, $\mathbf{z} = \mathbf{z}$, $\theta = \theta$, $\mathfrak{c} = \mathfrak{c}$, $\mathbf{v} = \mathbf{v}$, $p = p$, $(\nu_d)_{d\in\mathbb{N}} = (\nu_d)_{d\in\mathbb{N}}$, $(\varphi_d)_{d\in\mathbb{N}} = (\varphi_d)_{d\in\mathbb{N}}$, $(\mu_d)_{d\in\mathbb{N}} = (\mu_d)_{d\in\mathbb{N}}$, $(\sigma_d)_{d\in\mathbb{N}} = (\sigma_d)_{d\in\mathbb{N}}$, $(\phi_{d,\delta})_{d\in\mathbb{N},\,\delta\in(0,r]} = (\phi_{d,\delta})_{d\in\mathbb{N},\,\delta\in(0,r]}$ in the notation of Corollary 3.13) demonstrates that there exist unique continuous functions $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $u_d(0,x) = \varphi_d(x)$, which satisfy for all $d \in \mathbb{N}$ that $\inf_{q\in(0,\infty)}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $u_d|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u_d)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t,x)\big) + \langle(\nabla_x u_d)(t,x),\mu_d(x)\rangle_{\mathbb{R}^d} \tag{359}$$

for $(t,x) \in (0,T)\times\mathbb{R}^d$ and that there exist $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\,\varepsilon\in(0,R]} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,R]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{(5+\mathbf{z})\theta+z+w\mathbf{z}+4v}\,\varepsilon^{-4-\mathbf{z}}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{(5+\mathbf{z})\theta+z+w\mathbf{z}+2v}\,\varepsilon^{-2-\mathbf{z}}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in \mathcal{M}_d$, and

$$\left[\int_{\mathbb{R}^d} |u_d(T,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \;\nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{360}$$

The fact that $\operatorname{Im}(\mathcal{R}) \subseteq \cup_{d=1}^{\infty} C(\mathbb{R}^d,\mathbb{R})$ hence demonstrates that for all $d \in \mathbb{N}$, $\varepsilon \in (0,R]$ it holds that $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R}) \subseteq \mathcal{M}_d$. Combining this with (359) and (360) establishes items (i)–(ii). The proof of Theorem 3.14 is thus completed. ∎

Lemma 3.15. *Let $d \in \mathbb{N}$, $p \in [2,\infty)$, $u \in \mathbb{R}$, $v \in (u,\infty)$, and let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm. Then it holds that*

$$\frac{1}{(v-u)^d} \int_{[u,v]^d} \|x\|^p \, dx \leq d^{p/2} \max\big\{|u|^p, |v|^p\big\}. \tag{361}$$

Proof of Lemma 3.15. Observe that the Hölder inequality implies that for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
\|x\|^2 = \sum_{i=1}^d |x_i|^2 &\leq \left(\sum_{i=1}^d |x_i|^p\right)^{2/p} \left(\sum_{i=1}^d 1\right)^{1-2/p} \\
&= \left(\sum_{i=1}^d |x_i|^p\right)^{2/p} d^{1-2/p}.
\end{aligned} \tag{362}$$

Next note that the Fubini theorem ensures that

$$\begin{aligned}
&\int_{[u,v]^d} \left(\sum_{i=1}^d |x_i|^p\right) d(x_1, x_2, \ldots, x_d) = \sum_{i=1}^d \int_{[u,v]^d} |x_i|^p \, d(x_1, x_2, \ldots, x_d) \\
&= \sum_{i=1}^d \left(\int_u^v |x_i|^p \, dx_i\right) \left(\int_u^v 1 \, dt\right)^{d-1} = d \left(\int_u^v |t|^p \, dt\right) (v-u)^{d-1} \\
&\leq d(v-u)^d \sup_{t \in [u,v]} \big[|t|^p\big] = d(v-u)^d \max\big\{|u|^p, |v|^p\big\}.
\end{aligned} \tag{363}$$

Combining this with (362) demonstrates that

$$\begin{aligned}
&\frac{1}{(v-u)^d} \int_{[u,v]^d} \|x\|^p \, dx \\
&\leq \frac{1}{(v-u)^d} \, d^{p/2-1} \int_{[u,v]^d} \left(\sum_{i=1}^d |x_i|^p\right) d(x_1, x_2, \ldots, x_d) \\
&\leq d^{p/2} \max\big\{|u|^p, |v|^p\big\}.
\end{aligned} \tag{364}$$

The proof of Lemma 3.15 is thus completed. ◻

Corollary 3.16. *Let $T, r, \mathfrak{c}, p \in (0,\infty)$, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm and let $|||\cdot|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} : \mathbb{R}^{d\times d} \to [0,\infty)$ be the Hilbert-Schmidt norm on $\mathbb{R}^{d\times d}$, let $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, be* *continuous functions, let* $\mu_d\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, *and* $\sigma_d\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, *be functions which satisfy for all* $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ *that*

$$\mu_d(\lambda x + y) + \lambda \mu_d(0) = \lambda \mu_d(x) + \mu_d(y), \tag{365}$$

$$\sigma_d(\lambda x + y) + \lambda \sigma_d(0) = \lambda \sigma_d(x) + \sigma_d(y), \tag{366}$$

and $\|\mu_d(x)\|_{\mathbb{R}^d} + |||\sigma_d(x)|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \leq \mathfrak{c}(1 + \|x\|_{\mathbb{R}^d})$, *let*

$$\mathcal{N} = \cup_{\mathcal{L}\in\{2,3,\dots\}} \cup_{(l_0,l_1,\dots,l_{\mathcal{L}})\in((\mathbb{N}^{\mathcal{L}})\times\{1\})} \left(\times_{k=1}^{\mathcal{L}} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})\right), \tag{367}$$

let $\mathbf{A}_d \in C(\mathbb{R}^d, \mathbb{R}^d)$, $d \in \mathbb{N}$, *and* $\mathbf{a} \in C(\mathbb{R}, \mathbb{R})$ *be functions which satisfy for all* $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ *that*

$$\mathbf{A}_d(x) = (\mathbf{a}(x_1), \mathbf{a}(x_2), \dots, \mathbf{a}(x_d)), \tag{368}$$

let $\mathcal{P}\colon \mathcal{N} \to \mathbb{N}$ *and* $\mathcal{R}\colon \mathcal{N} \to \cup_{d=1}^{\infty} C(\mathbb{R}^d, \mathbb{R})$ *be the functions which satisfy for all* $\mathcal{L} \in \{2, 3, \dots\}$, $(l_0, l_1, \dots, l_{\mathcal{L}}) \in ((\mathbb{N}^{\mathcal{L}}) \times \{1\})$, $\Phi = ((W_1, B_1), \dots, (W_{\mathcal{L}}, B_{\mathcal{L}})) = ((W_k^{(i,j)})_{i\in\{1,2,\dots,l_k\}, j\in\{1,2,\dots,l_{k-1}\}}, (B_k^{(i)})_{i\in\{1,2,\dots,l_k\}})_{k\in\{1,2,\dots,\mathcal{L}\}} \in (\times_{k=1}^{\mathcal{L}} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $x_0 \in \mathbb{R}^{l_0}, x_1 \in \mathbb{R}^{l_1}, \dots, x_{\mathcal{L}-1} \in \mathbb{R}^{l_{\mathcal{L}-1}}$ *with* $\forall\, k \in \mathbb{N} \cap (0, \mathcal{L})\colon x_k = \mathbf{A}_{l_k}(W_k x_{k-1} + B_k)$ *that*

$$\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}), \qquad (\mathcal{R}(\Phi))(x_0) = W_{\mathcal{L}} x_{\mathcal{L}-1} + B_{\mathcal{L}}, \tag{369}$$

and $\mathcal{P}(\Phi) = \sum_{k=1}^{\mathcal{L}} l_k(l_{k-1} + 1)$, *and let* $(\phi_{d,\delta})_{d\in\mathbb{N},\, \delta\in(0,r]} \subseteq \mathcal{N}$ *satisfy for all* $d \in \mathbb{N}$, $\delta \in (0, r]$, $x \in \mathbb{R}^d$ *that* $\mathcal{P}(\phi_{d,\delta}) \leq \mathfrak{c}\, d^{\mathfrak{c}} \delta^{-\mathfrak{c}}$, $\mathcal{R}(\phi_{d,\delta}) \in C(\mathbb{R}^d, \mathbb{R})$, $|(\mathcal{R}(\phi_{d,\delta}))(x)| \leq \mathfrak{c}\, d^{\mathfrak{c}}(1 + \|x\|_{\mathbb{R}^d}^{\mathfrak{c}})$, *and*

$$|\varphi_d(x) - (\mathcal{R}(\phi_{d,\delta}))(x)| \leq \mathfrak{c}\, d^{\mathfrak{c}}\, \delta\, (1 + \|x\|_{\mathbb{R}^d}^{\mathfrak{c}}). \tag{370}$$

Then

(i) there exist unique continuous functions $u_d\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, *which satisfy for all* $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ *that* $u_d(0, x) = \varphi_d(x)$, *which satisfy for all* $d \in \mathbb{N}$ *that* $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, *and which satisfy for all* $d \in \mathbb{N}$ *that* $u_d|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$\begin{aligned}(\tfrac{\partial}{\partial t} u_d)(t, x) = \operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t, x)\big) \\ + (\tfrac{\partial}{\partial x} u_d)(t, x)\, \mu_d(x)\end{aligned} \tag{371}$$

for $(t, x) \in (0, T) \times \mathbb{R}^d$ *and* *(ii) there exist* $\mathfrak{C} \in (0,\infty)$*,* $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\,\varepsilon\in(0,1]} \subseteq \mathcal{N}$ *such that for all* $d \in \mathbb{N}$*,* $\varepsilon \in (0,1]$ *it holds that* $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{\mathfrak{C}}\, \varepsilon^{-\mathfrak{C}}$*,* $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$*, and*

$$\left[\int_{[0,1]^d} |u_d(T,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \, dx\right]^{1/p} \leq \varepsilon. \tag{372}$$

Proof of Corollary 3.16. Throughout this proof let $m\colon (0,\infty) \to [2,\infty)$ be the function which satisfies for all $z \in (0,\infty)$ that $m(z) = \max\{2, z\}$ and for every $d \in \mathbb{N}$ let $\nu_d\colon \mathcal{B}(\mathbb{R}^d) \to [0,\infty]$ be the d-dimensional Lebesgue measure. Note that Lemma 3.15 (with $d = d$, $p = m(\mathfrak{c})m(p)$, $u = 0$, $v = 1$, $\|\cdot\| = \|\cdot\|_{\mathbb{R}^d}$ in the notation of Lemma 3.15) implies for all $d \in \mathbb{N}$ that

$$\int_{[0,1]^d} \|x\|_{\mathbb{R}^d}^{(m(\mathfrak{c})\, m(p))} \, dx \leq d^{1/2(m(p)\, m(\mathfrak{c}))}. \tag{373}$$

This ensures that

$$\begin{aligned} &\sup_{d\in\mathbb{N}} \left[d^{-1/2(m(p)\, m(\mathfrak{c}))} \int_{\mathbb{R}^d} \|x\|_{\mathbb{R}^d}^{(m(\mathfrak{c})\, m(p))} \, \nu_d|_{[0,1]^d}(dx)\right] \\ &= \sup_{d\in\mathbb{N}} \left[d^{-1/2(m(p)\, m(\mathfrak{c}))} \int_{[0,1]^d} \|x\|_{\mathbb{R}^d}^{(m(\mathfrak{c})\, m(p))} \, dx\right] \leq 1 < \infty. \end{aligned} \tag{374}$$

Theorem 3.14 (with $T = T$, $r = r$, $R = 1$, $v = \mathfrak{c}$, $w = \mathfrak{c}$, $z = \mathfrak{c}$, $\mathbf{z} = \mathfrak{c}$, $\theta = m(\mathfrak{c})/2$, $\mathfrak{c} = \max\{\sqrt{2}\,\mathfrak{c}, 1\}$, $\mathbf{v} = m(\mathfrak{c})$, $p = m(p)$, $(\nu_d)_{d\in\mathbb{N}} = (\nu_d|_{[0,1]^d})_{d\in\mathbb{N}}$, $(\varphi_d)_{d\in\mathbb{N}} = (\varphi_d)_{d\in\mathbb{N}}$, $(\mu_d)_{d\in\mathbb{N}} = (\mu_d)_{d\in\mathbb{N}}$, $(\sigma_d)_{d\in\mathbb{N}} = (\sqrt{2}\,\sigma_d)_{d\in\mathbb{N}}$, $(\phi_{d,\delta})_{d\in\mathbb{N},\,\delta\in(0,r]} = (\phi_{d,\delta})_{d\in\mathbb{N},\,\delta\in(0,r]}$ in the notation of Theorem 3.14) hence ensures that

(A) there exist unique continuous functions $u_d\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $u_d(0,x) = \varphi_d(x)$, which satisfy for all $d \in \mathbb{N}$ that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $u_d|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned} (\tfrac{\partial}{\partial t}u_d)(t,x) &= \operatorname{Trace}\bigl(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t,x)\bigr) \\ &\quad + (\tfrac{\partial}{\partial x}u_d)(t,x)\, \mu_d(x) \end{aligned} \tag{375}$$

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

(B) there exist $C \in (0,\infty)$, $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\,\varepsilon\in(0,1]} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq C\, d^{1/2(5+\mathfrak{c})m(\mathfrak{c})+\mathfrak{c}+\mathfrak{c}^2+4\mathfrak{c}}\, \varepsilon^{-4-\mathfrak{c}}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, and

$$\left[\int_{\mathbb{R}^d} |u_d(T,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^{m(p)}\ \nu_d|_{[0,1]^d}(dx)\right]^{1/m(p)} \leq \varepsilon. \tag{376}$$

This implies item (i). Moreover, note that (B) and the Hölder inequality demonstrate that for all $\mathfrak{C} \in [\max\{C, 4+\mathfrak{c}, 1/2(5+\mathfrak{c})m(\mathfrak{c})+\mathfrak{c}+\mathfrak{c}^2+4\mathfrak{c}\}, \infty)$ it holds that

$$\mathcal{P}(\psi_{d,\varepsilon}) \leq C\, d^{1/2(5+\mathfrak{c})m(\mathfrak{c})+\mathfrak{c}+\mathfrak{c}^2+4\mathfrak{c}}\, \varepsilon^{-4-\mathfrak{c}} \leq \mathfrak{C}\, d^{\mathfrak{C}}\, \varepsilon^{-(4+\mathfrak{c})} \leq \mathfrak{C}\, d^{\mathfrak{C}}\, \varepsilon^{-\mathfrak{C}} \tag{377}$$

and

$$\begin{aligned}
&\left[\int_{[0,1]^d} |u_d(T,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^{p}\ dx\right]^{1/p} \\
&\leq \left[\int_{[0,1]^d} |u_d(T,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^{m(p)}\ dx\right]^{1/m(p)} \\
&= \left[\int_{\mathbb{R}^d} |u_d(T,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^{m(p)}\ \nu_d|_{[0,1]^d}(dx)\right]^{1/m(p)} \leq \varepsilon.
\end{aligned} \tag{378}$$

This establishes item (ii). The proof of Corollary 3.16 is thus completed. ☐

4 Artifical neural network approximations for Black-Scholes partial differential equations

4.1 Elementary properties of the Black-Scholes model

In this subsection we establish in Lemma 4.2 below a few elementary properties of the coefficient functions in the Black-Scholes model. For the sake of completeness we also provide in this subsection a detailed proof of Lemma 4.2.

Setting 4.1. *Let $p \in [2,\infty)$, $T \in (0,\infty)$, $\theta \in [0,\infty)$, $(\alpha_{d,i})_{d\in\mathbb{N},i\in\{1,2,\ldots,d\}}$, $(\beta_{d,i})_{d\in\mathbb{N},i\in\{1,2,\ldots,d\}} \subseteq \mathbb{R}$ satisfy that $\sup_{d\in\mathbb{N},i\in\{1,2,\ldots,d\}}(|\alpha_{d,i}|+|\beta_{d,i}|) < \infty$, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, for every $d \in \mathbb{N}$ let $\langle\cdot,\cdot\rangle_{\mathbb{R}^d} \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean* *scalar product, for every* $d \in \mathbb{N}$ *let* $|||\cdot|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \colon \mathbb{R}^{d\times d} \to [0,\infty)$ *be the Hilbert-Schmidt norm on* $\mathbb{R}^{d\times d}$*, let* $e_{d,i} \in \mathbb{R}^d$*,* $d \in \mathbb{N}$*,* $i \in \{1,2,\ldots,d\}$*, satisfy for all* $d \in \mathbb{N}$ *that* $e_{d,1} = (1,0,\ldots,0)$*,* $e_{d,2} = (0,1,0,\ldots,0)$*,* $\ldots$*,* $e_{d,d} = (0,\ldots,0,1)$*, let* $\mathbf{B}_d = (\mathbf{B}_d^{(i,j)})_{i,j\in\{1,2,\ldots,d\}} \in \mathbb{R}^{d\times d}$*,* $d \in \mathbb{N}$*, satisfy for all* $d \in \mathbb{N}$*,* $i \in \{1,2,\ldots,d\}$ *that* $\langle e_{d,i}, \mathbf{B}_d\mathbf{B}_d^* e_{d,i}\rangle_{\mathbb{R}^d} = 1$*, let* $\mu_d \colon \mathbb{R}^d \to \mathbb{R}^d$*,* $d \in \mathbb{N}$*, and* $\sigma_d \colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$*,* $d \in \mathbb{N}$*, be the functions which satisfy for all* $d \in \mathbb{N}$*,* $x = (x_1,x_2,\ldots,x_d) \in \mathbb{R}^d$ *that*

$$\mu_d(x) = (\alpha_{d,1}x_1,\ldots,\alpha_{d,d}x_d) \quad \text{and} \quad \sigma_d(x) = \operatorname{diag}(\beta_{d,1}x_1,\ldots,\beta_{d,d}x_d)\mathbf{B}_d, \tag{379}$$

let $\nu_d \colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$*,* $d \in \mathbb{N}$*, be probability measures which satisfy for all* $q \in (0,\infty)$ *that*

$$\sup_{d\in\mathbb{N}} \left[d^{-\theta q} \textstyle\int_{\mathbb{R}^d} \|x\|_{\mathbb{R}^d}^q \, \nu_d(dx) \right] < \infty, \tag{380}$$

let

$$\mathcal{N} = \cup_{\mathcal{L}\in\{2,3,\ldots\}} \cup_{(l_0,l_1,\ldots,l_{\mathcal{L}})\in((\mathbb{N}^{\mathcal{L}})\times\{1\})} \left(\times_{k=1}^{\mathcal{L}} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}) \right), \tag{381}$$

let $\mathbf{A}_d \in C(\mathbb{R}^d,\mathbb{R}^d)$*,* $d \in \mathbb{N}$*, be the functions which satisfy for all* $d \in \mathbb{N}$*,* $x = (x_1,x_2,\ldots,x_d) \in \mathbb{R}^d$ *that*

$$\mathbf{A}_d(x) = (\max\{x_1,0\},\max\{x_2,0\},\ldots,\max\{x_d,0\}), \tag{382}$$

and let $\mathscr{P}, \mathcal{P} \colon \mathcal{N} \to \mathbb{N}$ *and* $\mathcal{R} \colon \mathcal{N} \to \cup_{d=1}^\infty C(\mathbb{R}^d,\mathbb{R})$ *be the functions which satisfy for all* $\mathcal{L} \in \{2,3,\ldots\}$*,* $(l_0,l_1,\ldots,l_{\mathcal{L}}) \in ((\mathbb{N}^{\mathcal{L}})\times\{1\})$*,* $\Phi = ((W_1,B_1),\ldots,(W_{\mathcal{L}},B_{\mathcal{L}})) = ((W_k^{(i,j)})_{i\in\{1,2,\ldots,l_k\},j\in\{1,2,\ldots,l_{k-1}\}}, (B_k^{(i)})_{i\in\{1,2,\ldots,l_k\}})_{k\in\{1,2,\ldots,\mathcal{L}\}} \in (\times_{k=1}^{\mathcal{L}} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$*,* $x_0 \in \mathbb{R}^{l_0}, x_1 \in \mathbb{R}^{l_1}, \ldots, x_{\mathcal{L}-1} \in \mathbb{R}^{l_{\mathcal{L}-1}}$ *with* $\forall\, k \in \mathbb{N}\cap(0,\mathcal{L}) \colon x_k = \mathbf{A}_{l_k}(W_k x_{k-1} + B_k)$ *that*

$$\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0},\mathbb{R}), \qquad (\mathcal{R}(\Phi))(x_0) = W_{\mathcal{L}}x_{\mathcal{L}-1} + B_{\mathcal{L}}, \tag{383}$$

$$\mathscr{P}(\Phi) = \sum_{k=1}^{\mathcal{L}} \sum_{i=1}^{l_k} \left(\mathbb{1}_{\mathbb{R}\setminus\{0\}}(B_k^{(i)}) + \sum_{j=1}^{l_{k-1}} \mathbb{1}_{\mathbb{R}\setminus\{0\}}(W_k^{(i,j)}) \right), \tag{384}$$

and $\mathcal{P}(\Phi) = \sum_{k=1}^{\mathcal{L}} l_k(l_{k-1}+1)$*.*

Lemma 4.2. *Assume Setting 4.1. Then*

(i) it holds for all $d \in \mathbb{N}$*,* $x,y \in \mathbb{R}^d$*,* $\lambda \in \mathbb{R}$ *that*

$$\mu_d(\lambda x + y) + \lambda\mu_d(0) = \lambda\mu_d(x) + \mu_d(y), \tag{385}$$

(ii) it holds for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ that

$$\sigma_d(\lambda x + y) + \lambda \sigma_d(0) = \lambda \sigma_d(x) + \sigma_d(y), \tag{386}$$

and

(iii) for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \|\mu_d(x)\|_{\mathbb{R}^d} + \|\!|\sigma_d(x)|\!\|_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} &\leq 2 \left[\sup_{d \in \mathbb{N}, i \in \{1,2,\ldots,d\}} (|\alpha_{d,i}| + |\beta_{d,i}|) \right] \|x\|_{\mathbb{R}^d} \\ &\leq 2 \left[\sup_{d \in \mathbb{N}, i \in \{1,2,\ldots,d\}} (|\alpha_{d,i}| + |\beta_{d,i}|) \right] (1 + \|x\|_{\mathbb{R}^d}) < \infty. \end{aligned} \tag{387}$$

Proof of Lemma 4.2. Throughout this proof let $L \in (0, \infty)$ be given by

$$L = \sup_{d \in \mathbb{N}} \sup_{i \in \{1,2,\ldots,d\}} (|\alpha_{d,i}| + |\beta_{d,i}|). \tag{388}$$

First, note that the fact that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $\mu_d(x) = \operatorname{diag}(\alpha_{d,1}, \ldots, \alpha_{d,d})x$ and Lemma 2.6 prove item (i). Moreover, observe that (379) implies that for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ it holds that $\sigma_d(0) = 0$ and

$$\sigma_d(\lambda x + y) = \lambda \sigma_d(x) + \sigma_d(y). \tag{389}$$

This establishes item (ii). In addition, note that for all $d \in \mathbb{N}$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \|\mu_d(x)\|_{\mathbb{R}^d} &= \|(\alpha_{d,1}x_1, \ldots, \alpha_{d,d}x_d)\|_{\mathbb{R}^d} = \left[\textstyle\sum_{i=1}^d |\alpha_{d,i}x_i|^2 \right]^{1/2} \\ &\leq \left[\left(\max\{|\alpha_{d,1}|, \ldots, |\alpha_{d,d}|\} \right)^2 \textstyle\sum_{i=1}^d |x_i|^2 \right]^{1/2} \\ &= \max\{|\alpha_{d,1}|, \ldots, |\alpha_{d,d}|\} \, \|x\|_{\mathbb{R}^d} \leq L \, \|x\|_{\mathbb{R}^d} \\ &\leq L(1 + \|x\|_{\mathbb{R}^d}) < \infty. \end{aligned} \tag{390}$$

Moreover, observe that the fact that for all $d \in \mathbb{N}$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that $\sigma_d(x) = (\beta_{d,i} x_i \mathbf{B}_d^{(i,j)})_{i,j \in \{1,2,\ldots,d\}} \in \mathbb{R}^{d \times d}$ assures that for all $d \in \mathbb{N}$,

$x = (x_1, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
|||\sigma_d(x)|||^2_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} &= \sum_{i,j=1}^{d} \big|\beta_{d,i} x_i \mathbf{B}_d^{(i,j)}\big|^2 \\
&= \sum_{i=1}^{d} \Big(|x_i|^2 |\beta_{d,i}|^2 \textstyle\sum_{j=1}^{d} \big|\mathbf{B}_d^{(i,j)}\big|^2 \Big) \\
&\leq \Big[\max_{i \in \{1,2,\dots,d\}} \Big(|\beta_{d,i}|^2 \textstyle\sum_{j=1}^{d} \big|\mathbf{B}_d^{(i,j)}\big|^2 \Big) \Big] \displaystyle\sum_{i=1}^{d} |x_i|^2 \\
&= \Big[\max_{i \in \{1,2,\dots,d\}} \Big(|\beta_{d,i}|^2 \textstyle\sum_{j=1}^{d} \big|\mathbf{B}_d^{(i,j)}\big|^2 \Big) \Big] \|x\|^2_{\mathbb{R}^d} .
\end{aligned} \tag{391}$$

The fact that for all $d \in \mathbb{N}$, $i \in \{1, 2, \dots, d\}$ it holds that

$$\textstyle\sum_{j=1}^{d} \big|\mathbf{B}_d^{(i,j)}\big|^2 = \langle \mathbf{B}_d^* e_{d,i}, \mathbf{B}_d^* e_{d,i} \rangle_{\mathbb{R}^d} = \langle e_{d,i}, \mathbf{B}_d \mathbf{B}_d^* e_{d,i} \rangle_{\mathbb{R}^d} = 1 \tag{392}$$

hence demonstrates that for all $d \in \mathbb{N}$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
|||\sigma_d(x)|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} &\leq \Big[\max_{i \in \{1,2,\dots,d\}} |\beta_{d,i}|^2 \Big]^{1/2} \|x\|_{\mathbb{R}^d} \\
&= \Big[\max_{i \in \{1,2,\dots,d\}} |\beta_{d,i}| \Big] \|x\|_{\mathbb{R}^d} \\
&\leq L \|x\|_{\mathbb{R}^d} \leq L(1 + \|x\|_{\mathbb{R}^d}) < \infty.
\end{aligned} \tag{393}$$

Combining this and (390) assures that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that

$$\|\mu_d(x)\|_{\mathbb{R}^d} + |||\sigma_d(x)|||_{\mathrm{HS}(\mathbb{R}^d,\mathbb{R}^d)} \leq 2L \|x\|_{\mathbb{R}^d} \leq 2L(1 + \|x\|_{\mathbb{R}^d}) < \infty. \tag{394}$$

This establishes item (iii). The proof of Lemma 4.2 is thus completed. ☐

4.2 Transformations of viscosity solutions

In this subsection we establish in Proposition 4.3, Corollary 4.4, and Corollary 4.5 a few elementary and essentially well-known transformation results for viscosity solutions of certain second-order PDEs.

Proposition 4.3. *Let $d \in \mathbb{N}$, $a, \lambda \in \mathbb{R}$, $b \in (a, \infty)$, let $f \colon (a, b) \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{d \times d} \to \mathbb{R}$ be a function which satisfies for all $t \in (a, b)$, $x \in \mathbb{R}^d$, $\alpha \in \mathbb{R}$, $\eta \in \mathbb{R}^d$, $A, B \in \{C \in \mathbb{R}^{d \times d} \colon C^* = C\}$ with $A \leq B$ that*

$$f(t, x, \alpha, \eta, A) \leq f(t, x, \alpha, \eta, B), \tag{395}$$

let $u\colon (a,b)\times\mathbb{R}^d\to\mathbb{R}$ *be a continuous function which satisfies that* u *is a viscosity solution of*

$$\lambda(\tfrac{\partial}{\partial t}u)(t,x) = f\big(t,x,u(t,x),(\nabla_x u)(t,x),(\operatorname{Hess}_x u)(t,x)\big) \tag{396}$$

for $(t,x)\in(a,b)\times\mathbb{R}^d$*, let* $R\colon[a,b]\to[a,b]$ *be the function which satisfies for all* $t\in[a,b]$ *that* $R(t)=a+b-t$*, let* $U\colon(a,b)\times\mathbb{R}^d\to\mathbb{R}$ *be the function which satisfies for all* $t\in(a,b)$*,* $x\in\mathbb{R}^d$ *that* $U(t,x)=u(R(t),x)$*, and let* $F\colon(a,b)\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^{d\times d}\to\mathbb{R}$ *be a function which satisfies for all* $t\in(a,b)$*,* $x\in\mathbb{R}^d$*,* $\alpha\in\mathbb{R}$*,* $\eta\in\mathbb{R}^d$*,* $A\in\mathbb{R}^{d\times d}$ *that*

$$F(t,x,\alpha,\eta,A) = f(R(t),x,\alpha,\eta,A). \tag{397}$$

Then it holds that $U\colon(a,b)\times\mathbb{R}^d\to\mathbb{R}$ *is a continuous function which satisfies that* U *is a viscosity solution of*

$$-\lambda(\tfrac{\partial}{\partial t}U)(t,x) = F\big(t,x,U(t,x),(\nabla_x U)(t,x),(\operatorname{Hess}_x U)(t,x)\big) \tag{398}$$

for $(t,x)\in(a,b)\times\mathbb{R}^d$*.*

Proof of Proposition 4.3. Throughout this proof let $(s,y)\in(a,b)\times\mathbb{R}^d$, let

$$\Psi=(\Psi(t,x))_{(t,x)\in(a,b)\times\mathbb{R}^d},\Phi=(\Phi(t,x))_{(t,x)\in(a,b)\times\mathbb{R}^d}\colon(a,b)\times\mathbb{R}^d\to\mathbb{R} \tag{399}$$

be twice continuously differentiable functions which satisfy that $\Phi\geq U$, $\Phi(s,y)=U(s,y)$, $\Psi\leq U$, and $\Psi(s,y)=U(s,y)$, let

$$\psi=(\psi(t,x))_{(t,x)\in(a,b)\times\mathbb{R}^d},\varphi=(\varphi(t,x))_{(t,x)\in(a,b)\times\mathbb{R}^d}\colon(a,b)\times\mathbb{R}^d\to\mathbb{R} \tag{400}$$

be the functions which satisfy for all $(t,x)\in(a,b)\times\mathbb{R}^d$ that

$$\psi(t,x)=\Psi(R(t),x) \qquad\text{and}\qquad \varphi(t,x)=\Phi(R(t),x). \tag{401}$$

Observe that $R\colon[a,b]\to[a,b]$ is a bijective function which satisfies that $R|_{(a,b)}\colon(a,b)\to(a,b)$ is twice continuously differentiable and which satisfies for all $t\in[a,b]$, $r\in(a,b)$ that

$$R(R(t))=t \qquad\text{and}\qquad R'(r)=-1. \tag{402}$$

Combining this and (401) ensures for all $(t,x)\in(a,b)\times\mathbb{R}^d$ that

$$\Psi(t,x)=\psi(R(t),x) \qquad\text{and}\qquad \Phi(t,x)=\varphi(R(t),x). \tag{403}$$

Next note that (401), (402), and the hypothesis that for all $(t,x) \in (a,b) \times \mathbb{R}^d$ it holds that $U(t,x) = u(R(t),x)$ imply that for all $(t,x) \in (a,b) \times \mathbb{R}^d$ it holds that $\psi \in C^2((a,b) \times \mathbb{R}^d, \mathbb{R})$,

$$\psi(t,x) = \Psi(R(t),x) \leq U(R(t),x) = u(R(R(t)),x) = u(t,x), \tag{404}$$

and

$$\psi(R(s),y) = \Psi(s,y) = U(s,y) = u(R(s),y). \tag{405}$$

Moreover, observe that (401), (402), and the hypothesis that for all $(t,x) \in (a,b) \times \mathbb{R}^d$ it holds that $U(t,x) = u(R(t),x)$ demonstrate that for all $(t,x) \in (a,b) \times \mathbb{R}^d$ it holds that $\varphi \in C^2((a,b) \times \mathbb{R}^d, \mathbb{R})$,

$$\varphi(t,x) = \Phi(R(t),x) \geq U(R(t),x) = u(R(R(t)),x) = u(t,x), \tag{406}$$

and

$$\varphi(R(s),y) = \Phi(s,y) = U(s,y) = u(R(s),y). \tag{407}$$

Combining this, (404), and (405) with the hypothesis that u is a viscosity solution of

$$\lambda(\tfrac{\partial}{\partial t}u)(t,x) = f\big(t,x,u(t,x),(\nabla_x u)(t,x),(\operatorname{Hess}_x u)(t,x)\big) \tag{408}$$

for $(t,x) \in (a,b) \times \mathbb{R}^d$ implies that

$$\lambda(\tfrac{\partial}{\partial t}\varphi)(R(s),y) \leq f\big(R(s),y,\varphi(R(s),y),(\nabla_x\varphi)(R(s),y),(\operatorname{Hess}_x \varphi)(R(s),y)\big) \tag{409}$$

and

$$\lambda(\tfrac{\partial}{\partial t}\psi)(R(s),y) \geq f\big(R(s),y,\psi(R(s),y),(\nabla_x\psi)(R(s),y),(\operatorname{Hess}_x \psi)(R(s),y)\big). \tag{410}$$

This, (397), (402), and (403) ensure that

$$\begin{aligned}
&-\lambda(\tfrac{\partial}{\partial t}\Phi)(s,y) = \lambda(\tfrac{\partial}{\partial t}\varphi)(R(s),y) \\
&\leq f\big(R(s),y,\varphi(R(s),y),(\nabla_x\varphi)(R(s),y),(\operatorname{Hess}_x \varphi)(R(s),y)\big) \\
&= f\big(R(s),y,\Phi(s,y),(\nabla_x\Phi)(s,y),(\operatorname{Hess}_x \Phi)(s,y)\big) \\
&= F\big(s,y,\Phi(s,y),(\nabla_x\Phi)(s,y),(\operatorname{Hess}_x \Phi)(s,y)\big).
\end{aligned} \tag{411}$$

Moreover, observe that (397), (402), (403), and (410) demonstrate that

$$\begin{aligned}
&- \lambda(\tfrac{\partial}{\partial t}\Psi)(s,y) = \lambda(\tfrac{\partial}{\partial t}\psi)(R(s),y) \\
&\geq f\big(R(s), y, \psi(R(s),y), (\nabla_x \psi)(R(s),y), (\operatorname{Hess}_x \psi)(R(s),y)\big) \\
&= f\big(R(s), y, \Psi(s,y), (\nabla_x \Psi)(s,y), (\operatorname{Hess}_x \Psi)(s,y)\big) \\
&= F\big(s, y, \Psi(s,y), (\nabla_x \Psi)(s,y), (\operatorname{Hess}_x \Psi)(s,y)\big).
\end{aligned} \tag{412}$$

Next note that the hypothesis that $u\colon (a,b)\times\mathbb{R}^d \to \mathbb{R}$ is a continuous function and the hypothesis that for all $t \in (a,b)$, $x \in \mathbb{R}^d$ it holds that $U(t,x) = u(R(t),x)$ demonstrate that $U\colon (a,b)\times\mathbb{R}^d \to \mathbb{R}$ is a continuous function. Combining this with (411) and (412) assures that $U\colon (a,b)\times\mathbb{R}^d \to \mathbb{R}$ is a continuous function which satisfies that U is a viscosity subsolution and a viscosity supersolution of

$$- \lambda(\tfrac{\partial}{\partial t}U)(t,x) = F\big(t, x, U(t,x), (\nabla_x U)(t,x), (\operatorname{Hess}_x U)(t,x)\big) \tag{413}$$

for $(t,x) \in (a,b)\times\mathbb{R}^d$. This proves that $U\colon (a,b)\times\mathbb{R}^d \to \mathbb{R}$ is a continuous function which satisfies that U is a viscosity solution of

$$- \lambda(\tfrac{\partial}{\partial t}U)(t,x) = F\big(t, x, U(t,x), (\nabla_x U)(t,x), (\operatorname{Hess}_x U)(t,x)\big) \tag{414}$$

for $(t,x) \in (a,b)\times\mathbb{R}^d$. The proof of Proposition 4.3 is thus completed. ☐

Corollary 4.4. *Let $d \in \mathbb{N}$, $a, \mathfrak{a}, \mathfrak{b}, \lambda \in \mathbb{R}$, $b \in \mathbb{R}\backslash\{a\}$, let $\nu\colon \mathcal{B}(\mathbb{R}^d) \to [0,\infty]$ be a measure, let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ be a continuous function, let $\Phi\colon \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{B}(\mathbb{R}^d)\backslash\mathcal{B}(\mathbb{R})$-measurable function, let $f\colon \mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^{d\times d} \to \mathbb{R}$ be a function which satisfies for all $x \in \mathbb{R}^d$, $\alpha \in \mathbb{R}$, $\eta \in \mathbb{R}^d$, $A, B \in \{C \in \mathbb{R}^{d\times d}\colon C^* = C\}$ with $A \leq B$ that*

$$f(x,\alpha,\eta,A) \leq f(x,\alpha,\eta,B), \tag{415}$$

assume that $\mathfrak{a} = \min\{a,b\}$ and $\mathfrak{b} = \max\{a,b\}$, and assume that there exists a unique continuous function $u\colon [\mathfrak{a},\mathfrak{b}]\times\mathbb{R}^d \to \mathbb{R}$ which satisfies for all $x \in \mathbb{R}^d$ that $u(b,x) = \varphi(x)$, which satisfies that

$$\inf_{q\in(0,\infty)} \sup_{(t,x)\in[\mathfrak{a},\mathfrak{b}]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty, \tag{416}$$

and which satisfies that $u|_{(\mathfrak{a},\mathfrak{b})\times\mathbb{R}^d}$ is a viscosity solution of

$$\lambda(\tfrac{\partial}{\partial t}u)(t,x) = f\big(x, u(t,x), (\nabla_x u)(t,x), (\operatorname{Hess}_x u)(t,x)\big) \tag{417}$$

for $(t, x) \in (\mathfrak{a}, \mathfrak{b}) \times \mathbb{R}^d$ *and it holds that*

$$\left[\int_{\mathbb{R}^d} |u(a, x) - \Phi(x)|^p \, \nu(dx)\right]^{1/p} \leq \varepsilon. \tag{418}$$

Then there exists a unique continuous function $v \colon [\mathfrak{a}, \mathfrak{b}] \times \mathbb{R}^d \to \mathbb{R}$ *which satisfies for all* $x \in \mathbb{R}^d$ *that* $v(a, x) = \varphi(x)$*, which satisfies that*

$$\inf_{q \in (0,\infty)} \sup_{(t,x) \in [\mathfrak{a},\mathfrak{b}] \times \mathbb{R}^d} \frac{|v(t, x)|}{1 + \|x\|_{\mathbb{R}^d}^q} < \infty, \tag{419}$$

which satisfies that $v|_{(\mathfrak{a},\mathfrak{b}) \times \mathbb{R}^d}$ *is a viscosity solution of*

$$-\lambda(\tfrac{\partial}{\partial t} v)(t, x) = f\big(x, v(t, x), (\nabla_x v)(t, x), (\operatorname{Hess}_x v)(t, x)\big) \tag{420}$$

for $(t, x) \in (\mathfrak{a}, \mathfrak{b}) \times \mathbb{R}^d$ *and it holds that*

$$\left[\int_{\mathbb{R}^d} |v(b, x) - \Phi(x)|^p \, \nu(dx)\right]^{1/p} \leq \varepsilon. \tag{421}$$

Proof of Corollary 4.4. Throughout this proof let $v \colon [\mathfrak{a}, \mathfrak{b}] \times \mathbb{R}^d \to \mathbb{R}$ be the function which satisfies for all $t \in [\mathfrak{a}, \mathfrak{b}]$, $x \in \mathbb{R}^d$ that $v(t, x) = u(\mathfrak{a} + \mathfrak{b} - t, x)$. Note that for all $t \in [\mathfrak{a}, \mathfrak{b}]$, $x \in \mathbb{R}^d$ it holds that

$$v(t, x) = u(\mathfrak{a} + \mathfrak{b} - t, x) = u(a + b - t, x). \tag{422}$$

This and (418) ensure that

$$\left[\int_{\mathbb{R}^d} |v(b, x) - \Phi(x)|^p \, \nu(dx)\right]^{1/p} \leq \varepsilon. \tag{423}$$

Next note that (417) and Proposition 4.3 (with $d = d$, $a = \mathfrak{a}$, $\lambda = \lambda$, $b = \mathfrak{b}$, $f(t, x, \alpha, \eta, A) = f(x, \alpha, \eta, A)$, $u(t, x) = u(t, x)$, $U(t, x) = v(t, x)$ for $t \in (\mathfrak{a}, \mathfrak{b})$, $x \in \mathbb{R}^d$, $\alpha \in \mathbb{R}$, $\eta \in \mathbb{R}^d$, $A \in \mathbb{R}^{d \times d}$ in the notation of Proposition 4.3) demonstrate that $v|_{(\mathfrak{a},\mathfrak{b}) \times \mathbb{R}^d}$ is a viscosity solution of

$$-\lambda(\tfrac{\partial}{\partial t} v)(t, x) = f\big(x, v(t, x), (\nabla_x v)(t, x), (\operatorname{Hess}_x v)(t, x)\big) \tag{424}$$

for $(t, x) \in (\mathfrak{a}, \mathfrak{b}) \times \mathbb{R}^d$. Furthermore, observe that (416), (422), and the hypothesis that for all $x \in \mathbb{R}^d$ it holds that $u(b, x) = \varphi(x)$ imply that for all $x \in \mathbb{R}^d$ it holds that $v(a, x) = \varphi(x)$ and

$$\begin{aligned} \inf_{q\in(0,\infty)} \sup_{(t,x)\in[\mathfrak{a},\mathfrak{b}]\times\mathbb{R}^d} \frac{|v(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} &= \inf_{q\in(0,\infty)} \sup_{(t,x)\in[\mathfrak{a},\mathfrak{b}]\times\mathbb{R}^d} \frac{|u(\mathfrak{a}+\mathfrak{b}-t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} \\ &= \inf_{q\in(0,\infty)} \sup_{(t,x)\in[\mathfrak{a},\mathfrak{b}]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty. \end{aligned} \tag{425}$$

Next let $w\colon [\mathfrak{a}, \mathfrak{b}]\times\mathbb{R}^d \to \mathbb{R}$ be a continuous function which satisfies for all $x \in \mathbb{R}^d$ that $w(a, x) = \varphi(x)$, which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[\mathfrak{a},\mathfrak{b}]\times\mathbb{R}^d} \frac{|w(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, which satisfies that $w|_{(\mathfrak{a},\mathfrak{b})\times\mathbb{R}^d}$ is a viscosity solution of

$$-\lambda(\tfrac{\partial}{\partial t}w)(t,x) = f\big(x, w(t,x), (\nabla_x w)(t,x), (\operatorname{Hess}_x w)(t,x)\big) \tag{426}$$

for $(t, x) \in (\mathfrak{a}, \mathfrak{b}) \times \mathbb{R}^d$, and which satisfies that

$$\left[\int_{\mathbb{R}^d} |w(b,x) - \Phi(x)|^p \, \nu(dx)\right]^{1/p} \le \varepsilon, \tag{427}$$

and let $z\colon [\mathfrak{a}, \mathfrak{b}] \times \mathbb{R}^d \to \mathbb{R}$ be the function which satisfies for all $t \in [\mathfrak{a}, \mathfrak{b}]$, $x \in \mathbb{R}^d$ that

$$z(t,x) = w(\mathfrak{a}+\mathfrak{b}-t,x) = w(a+b-t,x). \tag{428}$$

Observe that z is a continuous function which satisfies that for all $x \in \mathbb{R}^d$ it holds that $z(b, x) = \varphi(x)$, which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[\mathfrak{a},\mathfrak{b}]\times\mathbb{R}^d} \frac{|z(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q}$ $< \infty$, which satisfies that $z|_{(\mathfrak{a},\mathfrak{b})\times\mathbb{R}^d}$ is a viscosity solution of

$$\lambda(\tfrac{\partial}{\partial t}z)(t,x) = f\big(x, z(t,x), (\nabla_x z)(t,x), (\operatorname{Hess}_x z)(t,x)\big) \tag{429}$$

for $(t, x) \in (\mathfrak{a}, \mathfrak{b})\times\mathbb{R}^d$ (cf. Proposition 4.3 (with $d = d$, $a = \mathfrak{a}$, $\lambda = -\lambda$, $b = \mathfrak{b}$, $f(t, x, \alpha, \eta, A) = f(x, \alpha, \eta, A)$, $u(t, x) = w(t, x)$, $U(t, x) = z(t, x)$ for $t \in (\mathfrak{a}, \mathfrak{b})$, $x \in \mathbb{R}^d$, $\alpha \in \mathbb{R}$, $\eta \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$ in the notation of Proposition 4.3)), and which satisfies that

$$\left[\int_{\mathbb{R}^d} |z(a,x) - \Phi(x)|^p \, \nu(dx)\right]^{1/p} \le \varepsilon. \tag{430}$$

Hence, we obtain that for all $t \in [\mathfrak{a}, \mathfrak{b}]$, $x \in \mathbb{R}^d$ it holds that $z(t, x) = u(t, x)$. The fact that for all $t \in [\mathfrak{a}, \mathfrak{b}]$, $x \in \mathbb{R}^d$ it holds that $v(t, x) = u(\mathfrak{a} + \mathfrak{b} - t, x)$

and $w(t,x) = z(\mathfrak{a}+\mathfrak{b}-t,x)$ therefore demonstrates that for all $t \in [\mathfrak{a},\mathfrak{b}]$, $x \in \mathbb{R}^d$ it holds that

$$w(t,x) = z(\mathfrak{a}+\mathfrak{b}-t,x) = u(\mathfrak{a}+\mathfrak{b}-t,x) = v(t,x). \tag{431}$$

Combining this, the fact that for all $x \in \mathbb{R}^d$ it holds that $v(a,x) = \varphi(x)$, (423), (424), and (425) completes the proof of Corollary 4.4. □

Corollary 4.5. *Assume Setting 4.1, let $d \in \mathbb{N}$, $\varepsilon, T \in (0,\infty)$, $\varphi \in C(\mathbb{R}^d,\mathbb{R})$, and $\psi \in \mathcal{N}$. Then the following two statements are equivalent:*

(i) There exists a unique continuous function $u\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ which satisfies for all $x \in \mathbb{R}^d$ that $u(T,x) = \varphi(x)$, which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, which satisfies that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned}(\tfrac{\partial}{\partial t}u)(t,x) + \big\langle (\nabla_x u)(t,x), \mu_d(x)\big\rangle_{\mathbb{R}^d} \\ + \tfrac{1}{2}\operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u)(t,x)\big) = 0\end{aligned} \tag{432}$$

for $(t,x) \in (0,T)\times\mathbb{R}^d$, and which satisfies that

$$\left[\int_{\mathbb{R}^d} |u(0,x) - (\mathcal{R}(\psi))(x)|^p \, \nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{433}$$

(ii) There exists a unique continuous function $v\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ which satisfies for all $x \in \mathbb{R}^d$ that $v(0,x) = \varphi(x)$, which satisfies that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|v(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, which satisfies that $v|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned}(\tfrac{\partial}{\partial t}v)(t,x) = \tfrac{1}{2}\operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x v)(t,x)\big) \\ + \big\langle (\nabla_x v)(t,x), \mu_d(x)\big\rangle_{\mathbb{R}^d}\end{aligned} \tag{434}$$

for $(t,x) \in (0,T)\times\mathbb{R}^d$, and which satisfies that

$$\left[\int_{\mathbb{R}^d} |v(T,x) - (\mathcal{R}(\psi))(x)|^p \, \nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{435}$$

Proof of Corollary 4.5. Observe that Corollary 4.4 (with $d = d$, $a = 0$, $\lambda = -1$, $b = T$, $\nu = \nu_d$, $\varphi(x) = \varphi(x)$, $\Phi(x) = (\mathcal{R}(\psi))(x)$, $f(x, \alpha, \eta, A) = \langle \eta, \mu_d(x) \rangle_{\mathbb{R}^d} + \frac{1}{2} \operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^* A\big)$, $u(t, x) = u(t, x)$ for $t \in [0, T]$, $x \in \mathbb{R}^d$, $\alpha \in \mathbb{R}$, $\eta \in \mathbb{R}^d$, $A \in \mathbb{R}^{d \times d}$ in the notation of Corollary 4.5) proves that item (i) implies item (ii). Next note that Corollary 4.4 (with $d = d$, $a = T$, $\lambda = 1$, $b = 0$, $\nu = \nu_d$, $\varphi(x) = \varphi(x)$, $\Phi(x) = (\mathcal{R}(\psi))(x)$, $f(x, \alpha, \eta, A) = \langle \eta, \mu_d(x) \rangle_{\mathbb{R}^d} + \frac{1}{2} \operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^* A\big)$, $u(t, x) = v(t, x)$ for $t \in [0, T]$, $x \in \mathbb{R}^d$, $\alpha \in \mathbb{R}$, $\eta \in \mathbb{R}^d$, $A \in \mathbb{R}^{d \times d}$ in the notation of Corollary 4.5) proves that item (ii) implies item (i). The proof of Corollary 4.5 is thus completed. □

4.3 Artificial neural network approximations for basket call options

In this subsection we establish in Proposition 4.7 below that ANN approximations overcome the curse of dimensionality in the numerical approximations of the Black-Scholes model in the case of basket call options. Our proof of Proposition 4.7 employs the elementary ANN representation result for the payoff functions associated to basket call options in Lemma 4.6 below. For the sake of completeness we also provide in this subsection a detailed proof of Lemma 4.6.

Lemma 4.6. *Assume Setting 4.1 and let* $(c_{d,i})_{d \in \mathbb{N}, i \in \{1,2,\ldots,d\}}, (K_d)_{d \in \mathbb{N}} \subseteq \mathbb{R}$. *Then there exists* $(\phi_d)_{d \in \mathbb{N}} \subseteq \mathcal{N}$ *such that for all* $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ *it holds that* $\mathcal{P}(\phi_d) \leq 4d$, $\mathcal{R}(\phi_d) \in C(\mathbb{R}^d, \mathbb{R})$, *and*

$$(\mathcal{R}(\phi_d))(x) = \max\{c_{d,1}x_1 + c_{d,2}x_2 + \ldots + c_{d,d}x_d - K_d, 0\}. \tag{436}$$

Proof of Lemma 4.6. Throughout this proof let $(\phi_d)_{d \in \mathbb{N}} \subseteq \mathcal{N}$ satisfy for all $d \in \mathbb{N}$ that

$$\phi_d = \big(((c_{d,1}, c_{d,2}, \ldots, c_{d,d}), -K_d), (1, 0)\big) \in (\mathbb{R}^{d \times 1} \times \mathbb{R}) \times (\mathbb{R} \times \mathbb{R}) \tag{437}$$

(i.e., ϕ_d corresponds to a fully connected feedforward artificial neural network with 3 layers with dimensions $(d, 1, 1)$). Note that (382) and (383) ensure that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ it holds that $\mathcal{R}(\phi_d) \in C(\mathbb{R}^d, \mathbb{R})$ and

$$\begin{aligned}(\mathcal{R}(\phi_d))(x) &= 1 \cdot \max\left\{\begin{pmatrix} c_{d,1} & c_{d,2} & \cdots & c_{d,d} \end{pmatrix} x + (-K_d), 0\right\} + 0 \\ &= \max\{c_{d,1}x_1 + c_{d,2}x_2 + \ldots + c_{d,d}x_d - K_d, 0\}.\end{aligned} \tag{438}$$

Moreover, observe that for all $d \in \mathbb{N}$ it holds that

$$\mathcal{P}(\phi_d) = 1(d+1) + 1(1+1) = d + 3 \leq 4d. \tag{439}$$

This and (438) complete the proof of Lemma 4.6. $\square$

Proposition 4.7. *Assume Setting 4.1, let* $(c_{d,i})_{d\in\mathbb{N},i\in\{1,2,\ldots,d\}} \subseteq [0,1]$, $(K_d)_{d\in\mathbb{N}} \in (0,\infty)$, *and assume for all* $d \in \mathbb{N}$ *that* $\sum_{i=1}^{d} c_{d,i} = 1$. *Then*

(i) there exist unique continuous functions $u_d \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, *which satisfy for all* $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ *that* $u_d(T,x) = \max\{c_{d,1}x_1 + c_{d,2}x_2 + \ldots + c_{d,d}x_d - K_d, 0\}$, *which satisfy for all* $d \in \mathbb{N}$ *that* $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, *and which satisfy for all* $d \in \mathbb{N}$ *that* $u_d|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$\begin{aligned}(\tfrac{\partial}{\partial t}u_d)(t,x) + \big\langle(\nabla_x u_d)(t,x), \mu_d(x)\big\rangle_{\mathbb{R}^d} \\ + \tfrac{1}{2}\operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t,x)\big) = 0\end{aligned} \tag{440}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ *and*

(ii) there exist $\mathfrak{C} \in (0,\infty)$, $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\,\varepsilon\in(0,1]} \subseteq \mathcal{N}$ *such that for all* $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ *it holds that* $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1}\, \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1}\, \varepsilon^{-2}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, *and*

$$\left[\int_{\mathbb{R}^d} |u_d(0,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \; \nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{441}$$

Proof of Proposition 4.7. Throughout this proof let $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that

$$\varphi_d(x) = \max\{c_{d,1}x_1 + c_{d,2}x_2 + \ldots + c_{d,d}x_d - K_d, 0\} \tag{442}$$

and let $(\chi_d)_{d\in\mathbb{N}}, (\phi_{d,\delta})_{d\in\mathbb{N},\delta\in(0,1]} \subseteq \mathcal{N}$ satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ that $\mathcal{P}(\chi_d) \leq 4d$, $\mathcal{R}(\chi_d) \in C(\mathbb{R}^d,\mathbb{R})$, $(\mathcal{R}(\chi_d))(x) = \varphi_d(x)$ (cf. Lemma 4.6), and $\phi_{d,\delta} = \chi_d$. Note that for all $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\mathcal{R}(\phi_{d,\delta}) = \mathcal{R}(\chi_d) = \varphi_d \in C(\mathbb{R}^d,\mathbb{R}). \tag{443}$$

This implies that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that

$$|\varphi_d(x) - (\mathcal{R}(\phi_{d,\delta}))(x)| = |\varphi_d(x) - \varphi_d(x)| = 0 \leq d^0\delta^0(1 + \|x\|_{\mathbb{R}^d}^2). \tag{444}$$

Moreover, observe that (443) and the hypothesis that for all $d \in \mathbb{N}$, $i \in \{1, 2, \ldots, d\}$ it holds that $c_{d,i} \geq 0$ and $\sum_{i=1}^{d} c_{d,i} = 1$ assure that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, $\delta \in (0, 1]$ it holds that

$$\begin{aligned} |(\mathcal{R}(\phi_{d,\delta}))(x)| &= |(\mathcal{R}(\chi_d))(x)| = |\varphi_d(x)| \\ &= \max\{c_{d,1}x_1 + c_{d,2}x_2 + \ldots + c_{d,d}x_d - K_d, 0\} \\ &\leq c_{d,1}|x_1| + c_{d,2}|x_2| + \ldots + c_{d,d}|x_d| \\ &\leq \left[\textstyle\sum_{i=1}^{d} c_{d,i}\right] \max\{|x_1|, |x_2|, \ldots, |x_d|\} \\ &\leq \|x\|_{\mathbb{R}^d} \leq d^0(1 + \|x\|_{\mathbb{R}^d}^2). \end{aligned} \tag{445}$$

In addition, observe that for all $d \in \mathbb{N}$, $\delta \in (0, 1]$ it holds that

$$\mathcal{P}(\phi_{d,\delta}) = \mathcal{P}(\chi_d) \leq 4d = 4d^1\delta^{-0}. \tag{446}$$

Combining this, (443), (444), (445), the hypothesis that for all $q \in (0, \infty)$ it holds that

$$\sup_{d \in \mathbb{N}} \left[d^{-\theta q} \textstyle\int_{\mathbb{R}^d} \|x\|_{\mathbb{R}^d}^q \, \nu_d(dx)\right] < \infty, \tag{447}$$

and Lemma 4.2 with Theorem 3.14 (with $T = T$, $r = 1$, $R = 1$, $v = 0$, $w = 0$, $z = 1$, $\mathbf{z} = 0$, $\theta = \theta$, $\mathfrak{c} = \max\{4, 2\left[\sup_{d \in \mathbb{N}, i \in \{1,2,\ldots,d\}}(|\alpha_{d,i}| + |\beta_{d,i}|)\right]\}$, $\mathbf{v} = 2$, $p = p$, $\nu_d = \nu_d$, $\varphi_d = \varphi_d$, $\mu_d = \mu_d$, $\sigma_d = \sigma_d$, $a(x) = \max\{x, 0\}$, $\phi_{d,\delta} = \phi_{d,\delta}$ for $d \in \mathbb{N}$, $x \in \mathbb{R}$, $\delta \in (0, 1]$ in the notation of Theorem 3.14) demonstrates that there exist unique continuous functions $v_d \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $v_d(0, x) = \varphi_d(x)$, which satisfy for all $d \in \mathbb{N}$ that $\inf_{q \in (0,\infty)} \sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \frac{|v_d(t,x)|}{1 + \|x\|_{\mathbb{R}^d}^q} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $v_d|_{(0,T) \times \mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned} (\tfrac{\partial}{\partial t} v_d)(t, x) &= \tfrac{1}{2} \operatorname{Trace}\bigl(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x v_d)(t, x)\bigr) \\ &\quad + \bigl\langle (\nabla_x v_d)(t, x), \mu_d(x) \bigr\rangle_{\mathbb{R}^d} \end{aligned} \tag{448}$$

for $(t, x) \in (0, T) \times \mathbb{R}^d$ and that there exist $\mathfrak{C} \in (0, \infty)$, $(\psi_{d,\varepsilon})_{d \in \mathbb{N}, \varepsilon \in (0,1]} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1} \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1} \varepsilon^{-2}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, and

$$\left[\int_{\mathbb{R}^d} |v_d(T, x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \ \nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{449}$$

Corollary 4.5 hence assures that there exist unique continuous functions $u_d \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$u_d(T,x) = \varphi_d(x) = \max\{c_{d,1}x_1 + c_{d,2}x_2 + \ldots + c_{d,d}x_d - K_d, 0\}, \tag{450}$$

which satisfy for all $d \in \mathbb{N}$ that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $u_d|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned} &(\tfrac{\partial}{\partial t} u_d)(t,x) + \big\langle (\nabla_x u_d)(t,x), \mu_d(x) \big\rangle_{\mathbb{R}^d} \\ &+ \tfrac{1}{2} \operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t,x)\big) = 0 \end{aligned} \tag{451}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ and that it holds for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that

$$\left[\int_{\mathbb{R}^d} |u_d(0,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \, \nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{452}$$

Combining this with the fact that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1}\, \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1}\, \varepsilon^{-2}$, and $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$ establishes items (i)–(ii). The proof of Proposition 4.7 is thus completed. □

4.4 Artificial neural network approximations for basket put options

In this subsection we establish in Proposition 4.9 below that ANN approximations overcome the curse of dimensionality in the numerical approximation of the Black-Scholes model in the case of basket put options. Our proof of Proposition 4.9 employs the elementary ANN representation result for the payoff functions associated to basket put options in Lemma 4.8 below. For the sake of completeness we also provide in this subsection a detailed proof of Lemma 4.8.

Lemma 4.8. *Assume Setting 4.1 and let $(c_{d,i})_{d\in\mathbb{N},i\in\{1,2,\ldots,d\}} \subseteq \mathbb{R}$, $K \in \mathbb{R}$. Then there exists $(\phi_d)_{d\in\mathbb{N}} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ it holds that $\mathcal{P}(\phi_d) \leq 4d$, $\mathcal{R}(\phi_d) \in C(\mathbb{R}^d, \mathbb{R})$, and*

$$(\mathcal{R}(\phi_d))(x) = \max\{K - (c_{d,1}x_1 + c_{d,2}x_2 + \ldots + c_{d,d}x_d), 0\}. \tag{453}$$

Proof of Lemma 4.8. Note that Lemma 4.6 (with $c_{d,i} = -c_{d,i}$, $K_d = -K$ for $d \in \mathbb{N}$, $i \in \{1, 2, \ldots, d\}$ in the notation of Lemma 4.6) demonstrates that there exists $(\phi_d)_{d\in\mathbb{N}} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ it holds that $\mathcal{P}(\phi_d) \leq 4d$, $\mathcal{R}(\phi_d) \in C(\mathbb{R}^d, \mathbb{R})$, and

$$\begin{aligned}(\mathcal{R}(\phi_d))(x) &= \max\{-c_{d,1}x_1 - c_{d,2}x_2 - \ldots - c_{d,d}x_d + K, 0\} \\ &= \max\{K - (c_{d,1}x_1 + c_{d,2}x_2 + \ldots + c_{d,d}x_d), 0\}.\end{aligned} \tag{454}$$

The proof of Lemma 4.8 is thus completed. □

Proposition 4.9. *Assume Setting 4.1 and let $(c_{d,i})_{d\in\mathbb{N}, i\in\{1,2,\ldots,d\}} \subseteq [0, 1]$, $K \in (0, \infty)$ satisfy for all $d \in \mathbb{N}$ that $\sum_{i=1}^{d} c_{d,i} = 1$. Then*

(i) there exist unique continuous functions $u_d \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $u_d(T, x) = \max\{K-(c_{d,1}x_1+c_{d,2}x_2+\ldots+c_{d,d}x_d), 0\}$, which satisfy for all $d \in \mathbb{N}$ that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u_d(t,x)|}{1+\|x\|^q_{\mathbb{R}^d}} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $u_d|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned}&(\tfrac{\partial}{\partial t}u_d)(t, x) + \big\langle(\nabla_x u_d)(t, x), \mu_d(x)\big\rangle_{\mathbb{R}^d} \\ &+ \tfrac{1}{2}\operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t, x)\big) = 0\end{aligned} \tag{455}$$

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

(ii) there exist $\mathfrak{C} \in (0, \infty)$, $(\psi_{d,\varepsilon})_{d\in\mathbb{N}, \varepsilon\in(0,1]} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1}\, \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1}\, \varepsilon^{-2}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, and

$$\left[\int_{\mathbb{R}^d} |u_d(0, x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \; \nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{456}$$

Proof of Proposition 4.9. Throughout this proof let $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that

$$\varphi_d(x) = \max\{K - (c_{d,1}x_1 + c_{d,2}x_2 + \ldots + c_{d,d}x_d), 0\} \tag{457}$$

and let $(\chi_d)_{d\in\mathbb{N}}, (\phi_{d,\delta})_{d\in\mathbb{N},\delta\in(0,1]} \subseteq \mathcal{N}$ satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0, 1]$ that $\mathcal{P}(\chi_d) \leq 4d$, $\mathcal{R}(\chi_d) \in C(\mathbb{R}^d, \mathbb{R})$, $(\mathcal{R}(\chi_d))(x) = \varphi_d(x)$ (cf. Lemma 4.8), and $\phi_{d,\delta} = \chi_d$. Note that for all $d \in \mathbb{N}$, $\delta \in (0, 1]$ it holds that

$$\mathcal{R}(\phi_{d,\delta}) = \mathcal{R}(\chi_d) = \varphi_d \in C(\mathbb{R}^d, \mathbb{R}). \tag{458}$$

This and the hypothesis that for all $d \in \mathbb{N}$, $i \in \{1, 2, \ldots, d\}$ it holds that $c_{d,i} \in [0, 1]$ and $\sum_{i=1}^{d} c_{d,i} = 1$ assure that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, $\delta \in (0, 1]$ it holds that

$$\begin{aligned} |(\mathcal{R}(\phi_{d,\delta}))(x)| = |(\mathcal{R}(\chi_d))(x)| &= |\varphi_d(x)| \\ &= \max\{K - (c_{d,1}x_1 + c_{d,2}x_2 + \ldots + c_{d,d}x_d), 0\} \\ &\leq K + c_{d,1}|x_1| + c_{d,2}|x_2| + \ldots + c_{d,d}|x_d| \\ &\leq K + \left[\textstyle\sum_{i=1}^{d} c_{d,i}\right] \max\{|x_1|, |x_2|, \ldots, |x_d|\} \\ &\leq K + \|x\|_{\mathbb{R}^d} \leq K + 1 + \|x\|_{\mathbb{R}^d}^2 \\ &\leq (K + 1)\, d^0 (1 + \|x\|_{\mathbb{R}^d}^2). \end{aligned} \tag{459}$$

In addition, observe that for all $d \in \mathbb{N}$, $\delta \in (0, 1]$ it holds that

$$\mathcal{P}(\phi_{d,\delta}) = \mathcal{P}(\chi_d) \leq 4d = 4d^1 \delta^{-0}. \tag{460}$$

Furthermore, note that (458) ensures that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0, 1]$ it holds that

$$\begin{aligned} |\varphi_d(x) - (\mathcal{R}(\phi_{d,\delta}))(x)| &= |\varphi_d(x) - (\mathcal{R}(\chi_d))(x)| \\ &= |\varphi_d(x) - \varphi_d(x)| = 0 \leq d^0 \delta^0 (1 + \|x\|_{\mathbb{R}^d}^2). \end{aligned} \tag{461}$$

Combining this, (458)–(460), the fact that $(\varphi_d)_{d \in \mathbb{N}}$ are continuous functions, the hypothesis that for all $q \in (0, \infty)$ it holds that

$$\sup_{d \in \mathbb{N}} \left[d^{-\theta q} \textstyle\int_{\mathbb{R}^d} \|x\|_{\mathbb{R}^d}^q \, \nu_d(dx) \right] < \infty, \tag{462}$$

and Lemma 4.2 with Theorem 3.14 (with $T = T$, $r = 1$, $R = 1$, $v = 0$, $w = 0$, $z = 1$, $\theta = \theta$, $\mathbf{z} = 0$, $\mathfrak{c} = \max\{4, K + 1, 2 \left[\sup_{d \in \mathbb{N}, i \in \{1,2,\ldots,d\}} (|\alpha_{d,i}| + |\beta_{d,i}|)\right]\}$, $\mathbf{v} = 2$, $p = p$, $\nu_d = \nu_d$, $\varphi_d = \varphi_d$, $\mu_d = \mu_d$, $\sigma_d = \sigma_d$, $a(x) = \max\{x, 0\}$, $\phi_{d,\delta} = \phi_{d,\delta}$ for $d \in \mathbb{N}$, $x \in \mathbb{R}$, $\delta \in (0, 1]$ in the notation of Theorem 3.14) demonstrate that there exist unique continuous functions $v_d \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $v_d(0, x) = \varphi_d(x)$, which satisfy for all $d \in \mathbb{N}$ that $\inf_{q \in (0,\infty)} \sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \frac{|v_d(t,x)|}{1 + \|x\|_{\mathbb{R}^d}^q} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $v_d|_{(0,T) \times \mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned} (\tfrac{\partial}{\partial t} v_d)(t, x) = \tfrac{1}{2} \operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^* (\operatorname{Hess}_x v_d)(t, x)\big) \\ + \big\langle (\nabla_x v_d)(t, x), \mu_d(x) \big\rangle_{\mathbb{R}^d} \end{aligned} \tag{463}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ and that there exist $\mathfrak{C} \in (0,\infty)$, $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\, \varepsilon\in(0,1]} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1}\, \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1}\, \varepsilon^{-2}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, and

$$\left[\int_{\mathbb{R}^d} |v_d(T,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \; \nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{464}$$

Corollary 4.5 hence assures that there exist unique continuous functions $u_d \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$u_d(T,x) = \varphi_d(x) = \max\{K - (c_{d,1}x_1 + c_{d,2}x_2 + \ldots + c_{d,d}x_d), 0\}, \tag{465}$$

which satisfy for all $d \in \mathbb{N}$ that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u_d(t,x)|}{1+\|x\|^q_{\mathbb{R}^d}} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $u_d|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned} &(\tfrac{\partial}{\partial t}u_d)(t,x) + \big\langle (\nabla_x u_d)(t,x), \mu_d(x)\big\rangle_{\mathbb{R}^d} \\ &+ \tfrac{1}{2}\operatorname{Trace}\!\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t,x)\big) = 0 \end{aligned} \tag{466}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ and that it holds for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that

$$\left[\int_{\mathbb{R}^d} |u_d(0,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \; \nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{467}$$

Combining this with the fact that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1}\, \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+1}\, \varepsilon^{-2}$, and $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$ establishes items (i)–(ii). The proof of Proposition 4.9 is thus completed. □

4.5 Artificial neural network approximations for call on max options

In this subsection we establish in Proposition 4.13 below that ANN approximations overcome the curse of dimensionality in the numerical approximation of the Black-Scholes model in the case of call on max options. Our proof of Proposition 4.13 employs the ANN representation result for the payoff functions associated to call on max options in Lemma 4.12 below. Our proof of Lemma 4.12, in turn, uses the elementary and essentially well-known facts in Lemmas 4.10–4.11. For the sake of completeness we also provide in this subsection detailed proofs of Lemmas 4.10–4.11.

Lemma 4.10. *Let $x, y \in \mathbb{R}$ and let $(\cdot)^+ : \mathbb{R} \to [0, \infty)$ be the function which satisfies for all $q \in \mathbb{R}$ that $(q)^+ = \max\{q, 0\}$. Then*

(i) it holds that $x = (x)^+ - (-x)^+$ and

(ii) it holds that $\max\{x, y\} = (x - y)^+ + y$.

(iii) it holds that $\min\{x, y\} = -(x - y)^+ + x$.

Proof of Lemma 4.10. Observe that

$$\begin{aligned}(x)^+ - (-x)^+ &= \max\{x, 0\} - \max\{-x, 0\} \\ &= [x - 0]\mathbb{1}_{[0,\infty)}(x) + [0 - (-x)]\mathbb{1}_{(-\infty,0)}(x) = x.\end{aligned} \tag{468}$$

This establishes item (i). Next note that

$$(x - y)^+ + y = \max\{x - y, 0\} + y = \max\{x, y\}. \tag{469}$$

This proves item (ii). Moreover, observe that

$$\begin{aligned}-(x - y)^+ + x = -(\max\{x - y, 0\} - x) &= -\max\{-y, -x\} \\ &= \min\{y, x\}.\end{aligned} \tag{470}$$

This establishes item (iii). The proof of Lemma 4.10 is thus completed. ☐

Lemma 4.11. *Let $(a_n)_{n\in\mathbb{N}} \subseteq [0, \infty)$ be the sequence which satisfies for all $n \in \mathbb{N}$ that*

$$\begin{aligned}a_n &= (2(n - 1) + 1)(n + 1) \\ &\quad + \left[\textstyle\sum_{k=1}^{n-1}(2(n - (k + 1)) + 1)(2(n - k) + 1 + 1)\right] + 1(1 + 1).\end{aligned} \tag{471}$$

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

$$a_n \leq 6n^3. \tag{472}$$

Proof of Lemma 4.11. Observe that for all $n \in \mathbb{N}$ it holds that

$$\begin{aligned}a_n &= (2(n - 1) + 1)(n + 1) \\ &\quad + \left[\textstyle\sum_{k=1}^{n-1}(2(n - (k + 1)) + 1)(2(n - k) + 1 + 1)\right] + 1(1 + 1) \\ &= (2n - 1)(n + 1) + \left[\textstyle\sum_{k=1}^{n-1}(2(n - k) - 1)(2(n - k) + 2)\right] + 2 \\ &= 2n^2 + n - 1 + \left[\textstyle\sum_{k=1}^{n-1} 4(n - k)^2 + 2(n - k) - 2\right] + 2 \\ &= 2n^2 + n + 1 + \left[\textstyle\sum_{k=1}^{n-1} 4n^2 - 8nk + 4k^2 + 2n - 2k - 2\right] \\ &= 2n^2 + n + 1 + 4n^2(n - 1) - 8n\left[\textstyle\sum_{k=1}^{n-1} k\right] + 4\left[\textstyle\sum_{k=1}^{n-1} k^2\right] \\ &\quad + 2n(n - 1) - 2\left[\textstyle\sum_{k=1}^{n-1} k\right] - 2(n - 1).\end{aligned} \tag{473}$$

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

$$\textstyle\sum_{k=1}^{n-1} k = \frac{(n-1)n}{2} \qquad \text{and} \textstyle\sum_{k=1}^{n-1} k^2 = \frac{(n-1)n(2n-1)}{6} \tag{474}$$

therefore assures for all $n \in \mathbb{N}$ that

$$\begin{aligned} a_n &= 2n^2 + n + 1 + 4n^2(n-1) - 8n\left[\tfrac{(n-1)n}{2}\right] + 4\left[\tfrac{(n-1)n(2n-1)}{6}\right] \\ &\quad + 2n(n-1) - 2\left[\tfrac{(n-1)n}{2}\right] - 2(n-1) \\ &= 2n^2 + n + 1 + 4n^3 - 4n^2 - \left[4n^3 - 4n^2\right] \\ &\quad + \left[\tfrac{8}{6}n^3 + \tfrac{4(-3)}{6}n^2 + \tfrac{4}{6}n\right] + 2n^2 - 2n - \left[n^2 - n\right] - 2n + 2 \\ &= \left(4 - 4 + \tfrac{4}{3}\right)n^3 + (2 - 4 + 4 - 2 + 2 - 1)\,n^2 \\ &\quad + \left(1 + \tfrac{2}{3} - 2 + 1 - 2\right)n + 1 + 2 \\ &\leq \tfrac{4}{3}n^3 + n^2 + 3 \leq (\tfrac{4}{3} + 1 + 3)n^3 \leq 6n^3. \end{aligned} \tag{475}$$

The proof of Lemma 4.11 is thus completed. □

Lemma 4.12. *Assume Setting 4.1 and let* $(K_d)_{d\in\mathbb{N}}, (c_{d,i})_{d\in\mathbb{N}, i\in\{1,2,\ldots,d\}} \subseteq \mathbb{R}$*. Then there exists* $(\phi_d)_{d\in\mathbb{N}} \subseteq \mathcal{N}$ *such that for all* $d \in \mathbb{N}$*,* $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ *it holds that* $\mathcal{P}(\phi_d) \leq 6d^3$*,* $\mathcal{R}(\phi_d) \in C(\mathbb{R}^d, \mathbb{R})$*, and*

$$(\mathcal{R}(\phi_d))(x) = \max\{\max\{c_{d,1}x_1, c_{d,2}x_2, \ldots, c_{d,d}x_d\} - K_d, 0\}. \tag{476}$$

Proof of Lemma 4.12. Throughout this proof let $d \in \mathbb{N}$, let

$$\begin{aligned} \phi &= ((W_1, B_1), (W_2, B_2), \ldots, (W_d, B_d), (W_{d+1}, B_{d+1})) \\ &\in (\mathbb{R}^{(2(d-1)+1)\times d} \times \mathbb{R}^{2(d-1)+1}) \\ &\quad \times (\times_{k=1}^{d-1}(\mathbb{R}^{(2(d-k)-1)\times(2(d-k)+1)} \times \mathbb{R}^{2(d-k)-1})) \\ &\quad \times (\mathbb{R}^{1\times 1} \times \mathbb{R}^1) \end{aligned} \tag{477}$$

(i.e. ϕ corresponds to fully connected feedforward artificial neural network with $d + 2$ layers with dimensions $(d, 2(d-1)+1, 2(d-2)+1, 2(d-3)+$

$1, \ldots, 3, 1, 1))$ satisfy for all $k \in \{1, 2, \ldots, d-2\}$ that

$$W_1 = \begin{pmatrix} c_{d,1} & -c_{d,2} & 0 & \cdots & 0 \\ 0 & c_{d,2} & 0 & \cdots & 0 \\ 0 & -c_{d,2} & 0 & \cdots & 0 \\ 0 & 0 & c_{d,3} & \cdots & 0 \\ 0 & 0 & -c_{d,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & c_{d,d} \\ 0 & 0 & 0 & \cdots & -c_{d,d} \end{pmatrix} \in \mathbb{R}^{(2(d-1)+1) \times d}, \tag{478}$$

$$W_{k+1} = \begin{pmatrix} 1 & 1 & -1 & -1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix} \tag{479}$$
$$\in \mathbb{R}^{(2(d-k)-1) \times (2(d-k)+1)},$$

$$W_d = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}, \qquad W_{d+1} = \begin{pmatrix} 1 \end{pmatrix} \in \mathbb{R}^{1 \times 1}, \qquad B_1 = 0, \tag{480}$$

$$B_2 = 0, \quad \ldots, B_{d-1} = 0, \qquad B_d = -K_d, \qquad \text{and} \qquad B_{d+1} = 0, \tag{481}$$

let $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, let $z_0 \in \mathbb{R}^d, z_1 \in \mathbb{R}^{2(d-1)-1}, z_2 \in \mathbb{R}^{2(d-2)+1}, \ldots,$ $z_{d-1} \in \mathbb{R}^3, z_d \in \mathbb{R}, z_{d+1} \in \mathbb{R}$ satisfy for all $k \in \{1, 2, \ldots, d-1\}$ that $z_0 = x$, $z_k = \mathbf{A}_{2(d-k)+1}(W_k z_{k-1} + B_k)$, $z_d = \mathbf{A}_1(W_d z_{d-1} + B_d)$, and

$$z_{d+1} = W_{d+1} z_d + B_{d+1}, \tag{482}$$

and let $(\cdot)^+ \colon \mathbb{R} \to [0, \infty)$ be the function which satisfies for all $q \in \mathbb{R}$ that $(q)^+ = \max\{q, 0\}$. Note that (383), (477), and (482) imply that $(\mathcal{R}(\phi)) \in C(\mathbb{R}^d, \mathbb{R})$ and

$$z_{d+1} = (\mathcal{R}(\phi))(x). \tag{483}$$

Next we claim that for all $k \in \{1, 2, \dots, d-1\}$ it holds that

$$z_k = \begin{pmatrix} \big(\max\{c_{d,1}x_1, c_{d,2}x_2, \dots, c_{d,k}x_k\} - c_{d,k+1}x_{k+1}\big)^+ \\ (c_{d,k+1}x_{k+1})^+ \\ (-c_{d,k+1}x_{k+1})^+ \\ (c_{d,k+2}x_{k+2})^+ \\ (-c_{d,k+2}x_{k+2})^+ \\ \vdots \\ (c_{d,d}x_d)^+ \\ (-c_{d,d}x_d)^+ \end{pmatrix}. \tag{484}$$

We now prove (484) by induction on $k \in \{1, 2, \dots, d-1\}$. For the base case $k = 1$ note that (382), (478), (480), and (482) assure that

$$\begin{aligned} z_1 &= \mathbf{A}_{2(d-1)+1}(W_1 z_0 + B_1) = \mathbf{A}_{2(d-1)+1}(W_1 x) \\ &= \mathbf{A}_{2(d-1)+1} \begin{pmatrix} c_{d,1}x_1 - c_{d,2}x_2 \\ c_{d,2}x_2 \\ -c_{d,2}x_2 \\ c_{d,3}x_3 \\ -c_{d,3}x_3 \\ \vdots \\ c_{d,d}x_d \\ -c_{d,d}x_d \end{pmatrix} = \begin{pmatrix} \big(\max\{c_{d,1}x_1\} - c_{d,2}x_2\big)^+ \\ (c_{d,2}x_2)^+ \\ (-c_{d,2}x_2)^+ \\ (c_{d,3}x_3)^+ \\ (-c_{d,3}x_3)^+ \\ \vdots \\ (c_{d,d}x_d)^+ \\ (-c_{d,d}x_d)^+ \end{pmatrix}. \end{aligned} \tag{485}$$

This establishes (484) in the base case $k = 1$. Next note that item (ii) in Lemma 4.10 implies that for all $a, b, c \in \mathbb{R}$ it holds that

$$(b - a)^+ + a - c = \max\{a, b\} - c. \tag{486}$$

For the induction step $\{1, 2, \dots, d-2\} \ni k \to k+1 \in \{2, 3, \dots, d-1\}$ observe that (382), (479), (481), (482), (486) (with $a = c_{d,k+1}x_{k+1}$, $b = \max\{c_{d,1}x_1, c_{d,2}x_2, \dots, c_{d,k}x_k\}$, $c = c_{d,k+2}x_{k+2}$ in the notation of (486)), and item (i) in Lemma 4.10 demonstrate that for all $k \in \{1, 2, \ldots, d-2\}$ with

$$z_k = \begin{pmatrix} \big(\max\{c_{d,1}x_1, c_{d,2}x_2, \ldots, c_{d,k}x_k\} - c_{d,k+1}x_{k+1} \big)^+ \\ (c_{d,k+1}x_{k+1})^+ \\ (-c_{d,k+1}x_{k+1})^+ \\ (c_{d,k+2}x_{k+2})^+ \\ (-c_{d,k+2}x_{k+2})^+ \\ \vdots \\ (c_{d,d}x_d)^+ \\ (-c_{d,d}x_d)^+ \end{pmatrix} \tag{487}$$

it holds that

$$
\begin{split}
z_{k+1} &= \mathbf{A}_{2(d-(k+1))+1}(W_{k+1}z_k + B_{k+1}) = \mathbf{A}_{2(d-(k+1))+1}(W_{k+1}z_k) \\
&= \mathbf{A}_{2(d-(k+1))+1}\begin{pmatrix} \left[\substack{(\max\{c_{d,1}x_1,\ldots,c_{d,k}x_k\}-c_{d,k+1}x_{k+1})^+ \\ +(c_{d,k+1}x_{k+1})^+-(-c_{d,k+1}x_{k+1})^+-(c_{d,k+2}x_{k+2})^++(-c_{d,k+2}x_{k+2})^+}\right] \\ (c_{d,k+2}x_{k+2})^+ - (-c_{d,k+2}x_{k+2})^+ \\ -(c_{d,k+2}x_{k+2})^+ + (-c_{d,k+2}x_{k+2})^+ \\ (c_{d,k+3}x_{k+3})^+ - (-c_{d,k+3}x_{k+3})^+ \\ -(c_{d,k+3}x_{k+3})^+ + (-c_{d,k+3}x_{k+3})^+ \\ \vdots \\ (c_{d,d}x_d)^+ - (-c_{d,d}x_d)^+ \\ -(c_{d,d}x_d)^+ + (-c_{d,d}x_d)^+ \end{pmatrix} \\
&= \mathbf{A}_{2(d-(k+1))+1}\begin{pmatrix} (\max\{c_{d,1}x_1,\ldots,c_{d,k}x_k\}-c_{d,k+1}x_{k+1})^+ + c_{d,k+1}x_{k+1} - c_{d,k+2}x_{k+2} \\ c_{d,k+2}x_{k+2} \\ -c_{d,k+2}x_{k+2} \\ c_{d,k+3}x_{k+3} \\ -c_{d,k+3}x_{k+3} \\ \vdots \\ c_{d,d}x_d \\ -c_{d,d}x_d \end{pmatrix} \\
&= \begin{pmatrix} \big(\max\{c_{d,1}x_1,\ldots,c_{d,k}x_k,c_{d,k+1}x_{k+1}\} - c_{d,k+2}x_{k+2}\big)^+ \\ (c_{d,k+2}x_{k+2})^+ \\ (-c_{d,k+2}x_{k+2})^+ \\ (c_{d,k+3}x_{k+3})^+ \\ (-c_{d,k+3}x_{k+3})^+ \\ \vdots \\ (c_{d,d}x_d)^+ \\ (-c_{d,d}x_d)^+ \end{pmatrix}.
\end{split}
\tag{488}
$$

Induction thus proves (484). Next observe that (480), (481), (482), (484), and item (ii) in Lemma 4.10 imply that

$$
\begin{aligned}
z_d &= \mathbf{A}_1(W_d z_{d-1} + B_d) \\
&= \mathbf{A}_1\left(\begin{pmatrix}1 & 1 & -1\end{pmatrix}\begin{pmatrix}\big(\max\{c_{d,1}x_1,\ldots,c_{d,d-1}x_{d-1}\} - c_{d,d}x_d\big)^+ \\ (c_{d,d}x_d)^+ \\ (-c_{d,d}x_d)^+\end{pmatrix} - K_d\right) \\
&= \mathbf{A}_1\Big(\big(\max\{c_{d,1}x_1,\ldots,c_{d,d-1}x_{d-1}\} - c_{d,d}x_d\big)^+ \\
&\quad + (c_{d,d}x_d)^+ - (-c_{d,d}x_d)^+ - K_d\Big) \\
&= \mathbf{A}_1\Big(\big(\max\{c_{d,1}x_1,\ldots,c_{d,d-1}x_{d-1}\} - c_{d,d}x_d\big)^+ + c_{d,d}x_d - K_d\Big) \\
&= \mathbf{A}_1\big(\max\{c_{d,1}x_1,\ldots,c_{d,d-1}x_{d-1},c_{d,d}x_d\} - K_d\big) \\
&= \max\{\max\{c_{d,1}x_1,\ldots,c_{d,d}x_d\} - K_d, 0\}.
\end{aligned} \tag{489}
$$

Combining this with (480)–(483) establishes that

$$
\begin{aligned}
(\mathcal{R}(\phi))(x) = z_{d+1} &= W_{d+1}z_d + B_{d+1} \\
&= z_d = \max\{\max\{c_{d,1}x_1,\ldots,c_{d,d}x_d\} - K_d, 0\}.
\end{aligned} \tag{490}
$$

In addition, observe that Lemma 4.11 implies that

$$
\begin{aligned}
\mathcal{P}(\phi) &= (2(d-1)+1)(d+1) \\
&\quad + \Big[\textstyle\sum_{k=1}^{d-1}(2(d-(k+1))+1)(2(d-k)+1+1)\Big] + 1(1+1) \\
&\le 6d^3.
\end{aligned} \tag{491}
$$

Combining this, (483), and (490) completes the proof of Lemma 4.12. ☐

Proposition 4.13. *Assume Setting 4.1 and let* $(K_d)_{d\in\mathbb{N}}, (c_{d,i})_{d\in\mathbb{N}, i\in\{1,2,\ldots,d\}} \subseteq [0,\infty)$ *satisfy that* $\sup_{d\in\mathbb{N}, i\in\{1,2,\ldots,d\}} c_{d,i} < \infty$. *Then*

(i) there exist unique continuous functions $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$, $d\in\mathbb{N}$, *which satisfy for all* $d\in\mathbb{N}$, $x=(x_1,x_2,\ldots,x_d)\in\mathbb{R}^d$ *that* $u_d(T,x) = \max\{\max\{c_{d,1}x_1, c_{d,2}x_2,\ldots,c_{d,d}x_d\} - K_d, 0\}$, *which satisfy for all* $d\in\mathbb{N}$ *that* $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, *and which satisfy for all* $d\in\mathbb{N}$ *that* $u_d|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$
\begin{aligned}
&(\tfrac{\partial}{\partial t}u_d)(t,x) + \big\langle (\nabla_x u_d)(t,x), \mu_d(x)\big\rangle_{\mathbb{R}^d} \\
&+ \tfrac{1}{2}\operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t,x)\big) = 0
\end{aligned} \tag{492}
$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ *and*

(ii) there exist $\mathfrak{C} \in (0,\infty)$, $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\,\varepsilon\in(0,1]} \subseteq \mathcal{N}$ *such that for all* $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ *it holds that* $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-2}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, *and*

$$\left[\int_{\mathbb{R}^d} |u_d(0,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \;\nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{493}$$

Proof of Proposition 4.13. Throughout this proof let $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that

$$\varphi_d(x) = \max\{\max\{c_{d,1}x_1, c_{d,2}x_2, \ldots, c_{d,d}x_d\} - K_d, 0\}, \tag{494}$$

let $(\chi_d)_{d\in\mathbb{N}}, (\phi_{d,\delta})_{d\in\mathbb{N},\delta\in(0,1]} \subseteq \mathcal{N}$ satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ that $\mathcal{P}(\chi_d) \leq 6d^3$, $\mathcal{R}(\chi_d) \in C(\mathbb{R}^d,\mathbb{R})$, $(\mathcal{R}(\chi_d))(x) = \varphi_d(x)$ (cf. Lemma 4.12), and $\phi_{d,\delta} = \chi_d$, and let $C \in [0,\infty)$ be given by $C = \sup_{d\in\mathbb{N}, i\in\{1,2,\ldots,d\}} c_{d,i}$. Note that for all $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\mathcal{R}(\phi_{d,\delta}) = \mathcal{R}(\chi_d) = \varphi_d \in C(\mathbb{R}^d,\mathbb{R}). \tag{495}$$

This and the fact that for all $d \in \mathbb{N}, i \in \{1,2,\ldots,d\}$ it holds that $c_{d,i} \in [0,\infty)$ ensures that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that

$$\begin{aligned}
|(\mathcal{R}(\phi_{d,\delta}))(x)| &= |(\mathcal{R}(\chi_d))(x)| = |\varphi_d(x)| \\
&= \max\{\max\{c_{d,1}x_1, c_{d,2}x_2, \ldots, c_{d,d}x_d\} - K_d, 0\} \\
&\leq \max\{c_{d,1}|x_1|, c_{d,2}|x_2|, \ldots, c_{d,d}|x_d|\} \\
&\leq C\max\{|x_1|, |x_2|, \ldots, |x_d|\} \\
&\leq C\,\|x\|_{\mathbb{R}^d} \leq Cd^0(1 + \|x\|_{\mathbb{R}^d}^2).
\end{aligned} \tag{496}$$

In addition, observe that for all $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\mathcal{P}(\phi_{d,\delta}) = \mathcal{P}(\chi_d) \leq 6d^3 = 6d^3\delta^{-0}. \tag{497}$$

Furthermore, note that (495) ensures that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that

$$\begin{aligned}
|\varphi_d(x) - (\mathcal{R}(\phi_{d,\delta}))(x)| &= |\varphi_d(x) - (\mathcal{R}(\chi_d))(x)| \\
&= |\varphi_d(x) - \varphi_d(x)| = 0 \leq d^0\delta^0(1 + \|x\|_{\mathbb{R}^d}^2).
\end{aligned} \tag{498}$$

Combining this, (495), (496), (497), the fact that $(\varphi_d)_{d\in\mathbb{N}}$ are continuous functions, the hypothesis that for all $q \in (0,\infty)$ it holds that

$$\sup_{d\in\mathbb{N}} \left[d^{-\theta q} \int_{\mathbb{R}^d} \|x\|_{\mathbb{R}^d}^q \, \nu_d(dx) \right] < \infty, \tag{499}$$

and Lemma 4.2 with Theorem 3.14 (with $T = T$, $r = 1$, $R = 1$, $v = 0$, $w = 0$, $z = 3$, $\mathbf{z} = 0$, $\theta = \theta$, $\mathfrak{c} = \max\{6, C, 2\left[\sup_{d\in\mathbb{N}, i\in\{1,2,\ldots,d\}}(|\alpha_{d,i}| + |\beta_{d,i}|)\right]\}$, $\mathbf{v} = 2$, $p = p$, $\nu_d = \nu_d$, $\varphi_d = \varphi_d$, $\mu_d = \mu_d$, $\sigma_d = \sigma_d$, $a(x) = \max\{x, 0\}$, $\phi_{d,\delta} = \phi_{d,\delta}$ for $d \in \mathbb{N}$, $x \in \mathbb{R}$, $\delta \in (0,1]$ in the notation of Theorem 3.14) demonstrates that there exist unique continuous functions $v_d \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $v_d(0,x) = \varphi_d(x)$, which satisfy for all $d \in \mathbb{N}$ that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|v_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $v_d|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned}(\tfrac{\partial}{\partial t} v_d)(t,x) &= \tfrac{1}{2} \operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x v_d)(t,x)\big) \\ &\quad + \big\langle (\nabla_x v_d)(t,x), \mu_d(x) \big\rangle_{\mathbb{R}^d}\end{aligned} \tag{500}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ and that there exist $\mathfrak{C} \in (0,\infty)$, $(\psi_{d,\varepsilon})_{d\in\mathbb{N}, \varepsilon\in(0,1]} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-2}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, and

$$\left[\int_{\mathbb{R}^d} |v_d(T,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \; \nu_d(dx) \right]^{1/p} \leq \varepsilon. \tag{501}$$

Corollary 4.5 hence assures that there exist unique continuous functions $u_d \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$u_d(T,x) = \varphi_d(x) = \max\{\max\{c_{d,1}x_1, c_{d,2}x_2, \ldots, c_{d,d}x_d\} - K_d, 0\}, \tag{502}$$

which satisfy for all $d \in \mathbb{N}$ that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $u_d|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned}&(\tfrac{\partial}{\partial t} u_d)(t,x) + \big\langle (\nabla_x u_d)(t,x), \mu_d(x) \big\rangle_{\mathbb{R}^d} \\ &+ \tfrac{1}{2} \operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t,x)\big) = 0\end{aligned} \tag{503}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ and that it holds for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that

$$\left[\int_{\mathbb{R}^d} |u_d(0,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \; \nu_d(dx) \right]^{1/p} \leq \varepsilon. \tag{504}$$

Combining this with the fact that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-2}$, and $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$ establishes items (i)–(ii). The proof of Proposition 4.13 is thus completed. ☐

4.6 Artificial neural network approximations for call on min options

In this subsection we establish in Proposition 4.15 below that ANN approximations overcome the curse of dimensionality in the numerical approximation of the Black-Scholes model in the case of call on min options. Our proof of Proposition 4.15 employs the ANN representation result for the payoff functions associated to call on min options in Lemma 4.14 below.

Lemma 4.14. *Assume Setting 4.1 and let $(K_d)_{d\in\mathbb{N}}, (c_{d,i})_{d\in\mathbb{N}, i\in\{1,2,\dots,d\}} \subseteq \mathbb{R}$. Then there exists $(\phi_d)_{d\in\mathbb{N}} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that $\mathcal{P}(\phi_d) \leq 6d^3$, $\mathcal{R}(\phi_d) \in C(\mathbb{R}^d, \mathbb{R})$, and*

$$(\mathcal{R}(\phi_d))(x) = \max\{\min\{c_{d,1}x_1, c_{d,2}x_2, \dots, c_{d,d}x_d\} - K_d, 0\}. \tag{505}$$

Proof of Lemma 4.14. Throughout this proof let $d \in \mathbb{N}$, let

$$\begin{aligned} \phi &= ((W_1, B_1), (W_2, B_2), \dots, (W_d, B_d), (W_{d+1}, B_{d+1})) \\ &\in \big(\mathbb{R}^{(2(d-1)+1)\times d} \times \mathbb{R}^{2(d-1)+1}\big) \\ &\quad \times \big(\times_{k=1}^{d-1} (\mathbb{R}^{(2(d-k)-1)\times(2(d-k)+1)} \times \mathbb{R}^{2(d-k)-1})\big) \\ &\quad \times \big(\mathbb{R}^{1\times 1} \times \mathbb{R}^1\big) \end{aligned} \tag{506}$$

(i.e., ϕ corresponds to fully connected feedforward artificial neural network with $d+2$ layers with dimensions $(d, 2(d-1)+1, 2(d-2)+1, 2(d-3)+1, \dots, 3, 1, 1))$ satisfy for all $k \in \{1, 2, \dots, d-2\}$ that

$$W_1 = \begin{pmatrix} -c_{d,1} & c_{d,2} & 0 & \cdots & 0 \\ 0 & c_{d,2} & 0 & \cdots & 0 \\ 0 & -c_{d,2} & 0 & \cdots & 0 \\ 0 & 0 & c_{d,3} & \cdots & 0 \\ 0 & 0 & -c_{d,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & c_{d,d} \\ 0 & 0 & 0 & \cdots & -c_{d,d} \end{pmatrix} \in \mathbb{R}^{(2(d-1)+1)\times d}, \tag{507}$$

$$W_{k+1} = \begin{pmatrix} 1 & -1 & 1 & 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix} \in \mathbb{R}^{(2(d-k)-1)\times(2(d-k)+1)}, \tag{508}$$

$$W_d = \begin{pmatrix} -1 & 1 & -1 \end{pmatrix}, \qquad W_{d+1} = \begin{pmatrix} 1 \end{pmatrix} \in \mathbb{R}^{1\times 1}, \qquad B_1 = 0, \tag{509}$$

$$B_2 = 0, \quad \dots, B_{d-1} = 0, \qquad B_d = -K_d, \qquad \text{and} \qquad B_{d+1} = 0, \tag{510}$$

let $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, let $z_0 \in \mathbb{R}^d, z_1 \in \mathbb{R}^{2(d-1)-1}, z_2 \in \mathbb{R}^{2(d-2)+1}, \dots, z_{d-1} \in \mathbb{R}^3, z_d \in \mathbb{R}, z_{d+1} \in \mathbb{R}$ satisfy for all $k \in \{1, 2, \dots, d-1\}$ that $z_0 = x$, $z_k = \mathbf{A}_{2(d-k)+1}(W_k z_{k-1} + B_k)$, $z_d = \mathbf{A}_1(W_d z_{d-1} + B_d)$, and

$$z_{d+1} = W_{d+1} z_d + B_{d+1}, \tag{511}$$

and let $(\cdot)^+ \colon \mathbb{R} \to [0, \infty)$ be the function which satisfies for all $q \in \mathbb{R}$ that $(q)^+ = \max\{q, 0\}$. Note that (383), (506), and (511) imply that $(\mathcal{R}(\phi)) \in C(\mathbb{R}^d, \mathbb{R})$ and

$$z_{d+1} = (\mathcal{R}(\phi))(x). \tag{512}$$

Next we claim that for all $k \in \{1, 2, \dots, d-1\}$ it holds that

$$z_k = \begin{pmatrix} \big(c_{d,k+1}x_{k+1} - \min\{c_{d,1}x_1, c_{d,2}x_2, \dots, c_{d,k}x_k\}\big)^+ \\ (c_{d,k+1}x_{k+1})^+ \\ (-c_{d,k+1}x_{k+1})^+ \\ (c_{d,k+2}x_{k+2})^+ \\ (-c_{d,k+2}x_{k+2})^+ \\ \vdots \\ (c_{d,d}x_d)^+ \\ (-c_{d,d}x_d)^+ \end{pmatrix}. \tag{513}$$

We now prove (513) by induction on $k \in \{1, 2, \dots, d-1\}$. For the base case

$k = 1$ note that (382), (507), (509), and (511) assure that

$$z_1 = \mathbf{A}_{2(d-1)+1}(W_1 z_0 + B_1) = \mathbf{A}_{2(d-1)+1}(W_1 x)$$
$$= \mathbf{A}_{2(d-1)+1}\begin{pmatrix} -c_{d,1}x_1 + c_{d,2}x_2 \\ c_{d,2}x_2 \\ -c_{d,2}x_2 \\ c_{d,3}x_3 \\ -c_{d,3}x_3 \\ \vdots \\ c_{d,d}x_d \\ -c_{d,d}x_d \end{pmatrix} = \begin{pmatrix} \big(c_{d,2}x_2 - \min\{c_{d,1}x_1\}\big)^+ \\ (c_{d,2}x_2)^+ \\ (-c_{d,2}x_2)^+ \\ (c_{d,3}x_3)^+ \\ (-c_{d,3}x_3)^+ \\ \vdots \\ (c_{d,d}x_d)^+ \\ (-c_{d,d}x_d)^+ \end{pmatrix}. \tag{514}$$

This establishes (513) in the base case $k = 1$. Next note that item (iii) in Lemma 4.10 implies that for all $a, b, c \in \mathbb{R}$ it holds that

$$(a - b)^+ - a + c = c - \min\{a, b\}. \tag{515}$$

For the induction step $\{1, 2, \ldots, d-2\} \ni k \to k+1 \in \{2, 3, \ldots, d-1\}$ observe that (382), (508), (510), (511), (515) (with $a = c_{d,k+1}x_{k+1}$, $b = \min\{c_{d,1}x_1, c_{d,2}x_2, \ldots, c_{d,k}x_k\}$, $c = c_{d,k+2}x_{k+2}$ in the notation of (515)), and item (i) in Lemma 4.10 demonstrate that for all $k \in \{1, 2, \ldots, d-2\}$ with

$$z_k = \begin{pmatrix} \big(c_{d,k+1}x_{k+1} - \min\{c_{d,1}x_1, c_{d,2}x_2, \ldots, c_{d,k}x_k\}\big)^+ \\ (c_{d,k+1}x_{k+1})^+ \\ (-c_{d,k+1}x_{k+1})^+ \\ (c_{d,k+2}x_{k+2})^+ \\ (-c_{d,k+2}x_{k+2})^+ \\ \vdots \\ (c_{d,d}x_d)^+ \\ (-c_{d,d}x_d)^+ \end{pmatrix} \tag{516}$$

it holds that

$$
\begin{aligned}
z_{k+1} &= \mathbf{A}_{2(d-(k+1))+1}(W_{k+1}z_k + B_{k+1}) = \mathbf{A}_{2(d-(k+1))+1}(W_{k+1}z_k) \\
&= \mathbf{A}_{2(d-(k+1))+1}\begin{pmatrix} \begin{bmatrix} (c_{d,k+1}x_{k+1}-\min\{c_{d,1}x_1,c_{d,2}x_2,\dots,c_{d,k}x_k\})^+ \\ -(c_{d,k+1}x_{k+1})^+ + (-c_{d,k+1}x_{k+1})^+ + (c_{d,k+2}x_{k+2})^+ - (-c_{d,k+2}x_{k+2})^+ \end{bmatrix} \\ (c_{d,k+2}x_{k+2})^+ - (-c_{d,k+2}x_{k+2})^+ \\ -(c_{d,k+2}x_{k+2})^+ + (-c_{d,k+2}x_{k+2})^+ \\ (c_{d,k+3}x_{k+3})^+ - (-c_{d,k+3}x_{k+3})^+ \\ -(c_{d,k+3}x_{k+3})^+ + (-c_{d,k+3}x_{k+3})^+ \\ \vdots \\ (c_{d,d}x_d)^+ - (-c_{d,d}x_d)^+ \\ -(c_{d,d}x_d)^+ + (-c_{d,d}x_d)^+ \end{pmatrix} \\
&= \mathbf{A}_{2(d-(k+1))+1}\begin{pmatrix} (c_{d,k+1}x_{k+1}-\min\{c_{d,1}x_1,c_{d,2}x_2,\dots,c_{d,k}x_k\})^+ - c_{d,k+1}x_{k+1} + c_{d,k+2}x_{k+2} \\ c_{d,k+2}x_{k+2} \\ -c_{d,k+2}x_{k+2} \\ c_{d,k+3}x_{k+3} \\ -c_{d,k+3}x_{k+3} \\ \vdots \\ c_{d,d}x_d \\ -c_{d,d}x_d \end{pmatrix} \\
&= \begin{pmatrix} \big(c_{d,k+2}x_{k+2} - \min\{c_{d,1}x_1,\dots,c_{d,k}x_k,c_{d,k+1}x_{k+1}\}\big)^+ \\ (c_{d,k+2}x_{k+2})^+ \\ (-c_{d,k+2}x_{k+2})^+ \\ (c_{d,k+3}x_{k+3})^+ \\ (-c_{d,k+3}x_{k+3})^+ \\ \vdots \\ (c_{d,d}x_d)^+ \\ (-c_{d,d}x_d)^+ \end{pmatrix}.
\end{aligned} \tag{517}
$$

Induction thus proves (513). Next observe that (508), (510), (511), (513), and item (iii) in Lemma 4.10 imply that

$$
\begin{aligned}
z_d &= \mathbf{A}_1(W_d z_{d-1} + B_d) \\
&= \mathbf{A}_1\left(\begin{pmatrix}-1 & 1 & -1\end{pmatrix}\begin{pmatrix}\big(c_{d,d}x_d - \min\{c_{d,1}x_1, \dots, c_{d,d-1}x_{d-1}\}\big)^+ \\ (c_{d,d}x_d)^+ \\ (-c_{d,d}x_d)^+\end{pmatrix} - K_d\right) \\
&= \mathbf{A}_1\Big(- \big(c_{d,d}x_d - \min\{c_{d,1}x_1, \dots, c_{d,d-1}x_{d-1}\}\big)^+ \\
&\quad + (c_{d,d}x_d)^+ - (-c_{d,d}x_d)^+ - K_d\Big) \\
&= \mathbf{A}_1\Big(- \big(c_{d,d}x_d - \min\{c_{d,1}x_1, \dots, c_{d,d-1}x_{d-1}\}\big)^+ + c_{d,d}x_d - K_d\Big) \\
&= \mathbf{A}_1\big(\min\{c_{d,1}x_1, \dots, c_{d,d-1}x_{d-1}, c_{d,d}x_d\} - K_d\big) \\
&= \max\{\min\{c_{d,1}x_1, \dots, c_{d,d}x_d\} - K_d, 0\}.
\end{aligned} \tag{518}
$$

Combining this with (509)–(512) establishes that

$$
\begin{aligned}
(\mathcal{R}(\phi))(x) &= z_{d+1} = W_{d+1}z_d + B_{d+1} \\
&= z_d = \max\{\min\{c_{d,1}x_1, \dots, c_{d,d}x_d\} - K_d, 0\}.
\end{aligned} \tag{519}
$$

In addition, observe that Lemma 4.11 implies that

$$
\begin{aligned}
\mathcal{P}(\phi) &= (2(d-1)+1)(d+1) \\
&\quad + \Big[\textstyle\sum_{k=1}^{d-1}(2(d-(k+1))+1)(2(d-k)+1+1)\Big] + 1(1+1) \\
&\le 6d^3.
\end{aligned} \tag{520}
$$

Combining this, (512), and (519) completes the proof of Lemma 4.14. ☐

Proposition 4.15. *Assume Setting 4.1 and let* $(K_d)_{d\in\mathbb{N}}, (c_{d,i})_{d\in\mathbb{N}, i\in\{1,2,\dots,d\}} \subseteq [0,\infty)$ *satisfy that* $\sup_{d\in\mathbb{N}, i\in\{1,2,\dots,d\}} c_{d,i} < \infty$*. Then*

(i) there exist unique continuous functions $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$*,* $d\in\mathbb{N}$*, which satisfy for all* $d\in\mathbb{N}$*,* $x=(x_1,x_2,\dots,x_d)\in\mathbb{R}^d$ *that* $u_d(T,x) = \max\{\min\{c_{d,1}x_1, c_{d,2}x_2, \dots, c_{d,d}x_d\} - K_d, 0\}$*, which satisfy for all* $d\in\mathbb{N}$ *that* $\inf_{q\in(0,\infty)}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$*, and which satisfy for all* $d\in\mathbb{N}$ *that* $u_d|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$
\begin{aligned}
&(\tfrac{\partial}{\partial t}u_d)(t,x) + \big\langle (\nabla_x u_d)(t,x), \mu_d(x)\big\rangle_{\mathbb{R}^d} \\
&+ \tfrac{1}{2}\operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t,x)\big) = 0
\end{aligned} \tag{521}
$$

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

(ii) there exist $\mathfrak{C} \in (0,\infty)$, $(\psi_{d,\varepsilon})_{d\in\mathbb{N},\,\varepsilon\in(0,1]} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-2}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, and

$$\left[\int_{\mathbb{R}^d} |u_d(0,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \; \nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{522}$$

Proof of Proposition 4.15. Throughout this proof let $\varphi_d\colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that

$$\varphi_d(x) = \max\{\min\{c_{d,1}x_1, c_{d,2}x_2, \ldots, c_{d,d}x_d\} - K_d, 0\}, \tag{523}$$

let $(\chi_d)_{d\in\mathbb{N}}, (\phi_{d,\delta})_{d\in\mathbb{N},\delta\in(0,1]} \subseteq \mathcal{N}$ satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ that $\mathcal{P}(\chi_d) \leq 6d^3$, $\mathcal{R}(\chi_d) \in C(\mathbb{R}^d,\mathbb{R})$, $(\mathcal{R}(\chi_d))(x) = \varphi_d(x)$ (cf. Lemma 4.14), and $\phi_{d,\delta} = \chi_d$, and let $C \in [0,\infty)$ be given by $C = \sup_{d\in\mathbb{N},i\in\{1,2,\ldots,d\}} c_{d,i}$. Note that for all $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\mathcal{R}(\phi_{d,\delta}) = \mathcal{R}(\chi_d) = \varphi_d \in C(\mathbb{R}^d,\mathbb{R}). \tag{524}$$

This ensures that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that

$$\begin{aligned} |(\mathcal{R}(\phi_{d,\delta}))(x)| &= |(\mathcal{R}(\chi_d))(x)| = |\varphi_d(x)| \\ &= \max\{\min\{c_{d,1}x_1, c_{d,2}x_2, \ldots, c_{d,d}x_d\} - K_d, 0\} \\ &\leq \max\{c_{d,1}|x_1|, c_{d,2}|x_2|, \ldots, c_{d,d}|x_d|\} \\ &\leq C \max\{|x_1|, |x_2|, \ldots, |x_d|\} \\ &\leq C\,\|x\|_{\mathbb{R}^d} \leq C d^0 (1 + \|x\|_{\mathbb{R}^d}^2). \end{aligned} \tag{525}$$

In addition, observe that for all $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\mathcal{P}(\phi_{d,\delta}) = \mathcal{P}(\chi_d) \leq 6d^3 = 6d^3\delta^{-0}. \tag{526}$$

Furthermore, note that (524) ensures that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that

$$\begin{aligned} |\varphi_d(x) - (\mathcal{R}(\phi_{d,\delta}))(x)| &= |\varphi_d(x) - (\mathcal{R}(\chi_d))(x)| \\ &= |\varphi_d(x) - \varphi_d(x)| = 0 \leq d^0\delta^0(1 + \|x\|_{\mathbb{R}^d}^2). \end{aligned} \tag{527}$$

Combining this, (524), (525), (526), the fact that $(\varphi_d)_{d\in\mathbb{N}}$ are continuous functions, the hypothesis that for all $q \in (0,\infty)$ it holds that

$$\sup_{d\in\mathbb{N}} \left[d^{-\theta q} \int_{\mathbb{R}^d} \|x\|_{\mathbb{R}^d}^q \, \nu_d(dx) \right] < \infty, \tag{528}$$

and Lemma 4.2 with Theorem 3.14 (with $T = T$, $r = 1$, $R = 1$, $v = 0$, $w = 0$, $z = 3$, $\mathbf{z} = 0$, $\theta = \theta$, $\mathfrak{c} = \max\{6, C, 2\left[\sup_{d\in\mathbb{N}, i\in\{1,2,\dots,d\}}(|\alpha_{d,i}| + |\beta_{d,i}|)\right]\}$, $\mathbf{v} = 2$, $p = p$, $\nu_d = \nu_d$, $\varphi_d = \varphi_d$, $\mu_d = \mu_d$, $\sigma_d = \sigma_d$, $a(x) = \max\{x, 0\}$, $\phi_{d,\delta} = \phi_{d,\delta}$ for $d \in \mathbb{N}$, $x \in \mathbb{R}$, $\delta \in (0,1]$ in the notation of Theorem 3.14) demonstrates that there exist unique continuous functions $v_d \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $v_d(0,x) = \varphi_d(x)$, which satisfy for all $d \in \mathbb{N}$ that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|v_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $v_d|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned}(\tfrac{\partial}{\partial t} v_d)(t,x) &= \tfrac{1}{2} \operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x v_d)(t,x)\big) \\ &\quad + \big\langle (\nabla_x v_d)(t,x), \mu_d(x) \big\rangle_{\mathbb{R}^d}\end{aligned} \tag{529}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ and that there exist $\mathfrak{C} \in (0,\infty)$, $(\psi_{d,\varepsilon})_{d\in\mathbb{N}, \varepsilon\in(0,1]} \subseteq \mathcal{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-2}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, and

$$\left[\int_{\mathbb{R}^d} |v_d(T,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \; \nu_d(dx) \right]^{1/p} \leq \varepsilon. \tag{530}$$

Corollary 4.5 hence assures that there exist unique continuous functions $u_d \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, which satisfy that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$u_d(T,x) = \varphi_d(x) = \max\{\min\{c_{d,1}x_1, c_{d,2}x_2, \dots, c_{d,d}x_d\} - K_d, 0\}, \tag{531}$$

which satisfy for all $d \in \mathbb{N}$ that $\inf_{q\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^q} < \infty$, and which satisfy for all $d \in \mathbb{N}$ that $u_d|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$\begin{aligned}&(\tfrac{\partial}{\partial t} u_d)(t,x) + \big\langle (\nabla_x u_d)(t,x), \mu_d(x) \big\rangle_{\mathbb{R}^d} \\ &+ \tfrac{1}{2} \operatorname{Trace}\big(\sigma_d(x)[\sigma_d(x)]^*(\operatorname{Hess}_x u_d)(t,x)\big) = 0\end{aligned} \tag{532}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ and that it holds for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that

$$\left[\int_{\mathbb{R}^d} |u_d(0,x) - (\mathcal{R}(\psi_{d,\varepsilon}))(x)|^p \; \nu_d(dx) \right]^{1/p} \leq \varepsilon. \tag{533}$$

Combining this with the fact that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-4}$, $\mathscr{P}(\psi_{d,\varepsilon}) \leq \mathfrak{C}\, d^{5\theta+3}\, \varepsilon^{-2}$, and $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$ establishes items (i)–(ii). The proof of Proposition 4.15 is thus completed. $\square$

Acknowledgements

We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG) through CRC 1173 and we gratefully acknowledge a Research Travel Grant by the Karlsruhe House of Young Scientists (KHYS) supporting the stay of the second author at ETH Zurich.

References

[1] Bach, F. Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research 18*, 19 (2017), 1–53.

[2] Barron, A. R. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory 39*, 3 (1993), 930–945.

[3] Barron, A. R. Approximation and estimation bounds for artificial neural networks. *Mach. Learn. 14*, 1 (1994), 115–133.

[4] Beck, C., Becker, S., Grohs, P., Jaafari, N., and Jentzen, A. Solving stochastic differential equations and Kolmogorov equations by means of deep learning. *arXiv:1806.00421* (2018), 56 pages.

[5] Beck, C., Jentzen, A., and E, W. Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *arXiv:1709.05963* (2017), 56 pages.

[6] Becker, S., Cheridito, P., and Jentzen, A. Deep optimal stopping. *arXiv:1804.05394* (2018), 18 pages.

[7] Bellman, R. *Dynamic programming*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 2010. Reprint of the 1957 edition, With a new introduction by Stuart Dreyfus.

[8] Blum, E. K., and Li, L. K. Approximation theory and feedforward networks. *Neural networks 4*, 4 (1991), 511–515.

[9] Bölcskei, H., Grohs, P., Kutyniok, G., and Petersen, P. Optimal approximation with sparsely connected deep neural networks. *arXiv:1705.01714* (2017), 36 pages.

[10] Burger, M., and Neubauer, A. Error Bounds for Approximation with Neural Networks . *Journal of Approximation Theory 112*, 2 (2001), 235–250.

[11] Candès, E. J. Ridgelets: Theory and Applications, 1998. Ph.D. thesis, Stanford University.

[12] Chen, T., and Chen, H. Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks. *IEEE Transactions on Neural Networks 6*, 4 (1995), 904–910.

[13] Chui, C. K., Li, X., and Mhaskar, H. N. Neural networks for localized approximation. *Math. Comp. 63*, 208 (1994), 607–623.

[14] Cox, S., Hutzenthaler, M., and Jentzen, A. Local Lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations. *arXiv:1309.5595* (2013), 84 pages.

[15] Cox, S., Hutzenthaler, M., Jentzen, A., van Neerven, J., and Welti, T. Convergence in Hölder norms with applications to Monte Carlo methods in infinite dimensions. *arXiv:1605.00856* (2017), 48 pages.

[16] Crandall, M. G., Ishii, H., and Lions, P.-L. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.) 27*, 1 (1992), 1–67.

[17] Crandall, M. G., and Lions, P.-L. Viscosity solutions of Hamilton-Jacobi equations. *Trans. Amer. Math. Soc. 277*, 1 (1983), 1–42.

[18] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signal 2*, 4 (1989), 303–314.

[19] Da Prato, G., and Zabczyk, J. *Stochastic equations in infinite dimensions*, vol. 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[20] Dahl, G. E., Yu, D., Deng, L., and Acero, A. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on audio, speech, and language processing 20*, 1 (2012), 30–42.

[21] DeVore, R., Oskolkov, K., and Petrushev, P. Approximation by feed-forward neural networks. *Ann. Numer. Math. 4* (1996), 261–287.

[22] E, W., Han, J., and Jentzen, A. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun. Math. Stat. 5*, 4 (2017), 349–380.

[23] E, W., and Yu, B. The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat. 6*, 1 (2018), 1–12.

[24] Eldan, R., and Shamir, O. The power of depth for feedforward neural networks. *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA* (2016), 907–940.

[25] Ellacott, S. Aspects of the numerical analysis of neural networks. *Acta Numer. 3* (1994), 145–202.

[26] Fujii, M., Takahashi, A., and Takahashi, M. Asymptotic Expansion as Prior Knowledge in Deep Learning Method for high dimensional BSDEs. *arXiv:1710.07030* (2017), 16 pages.

[27] Funahashi, K.-I. On the approximate realization of continuous mappings by neural networks. *Neural Networks 2*, 3 (1989), 183–192.

[28] Garling, D. J. H. *Inequalities: A journey into linear analysis.* Cambridge University Press, Cambridge, 2007.

[29] Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning.* MIT Press, 2016. `http://www.deeplearningbook.org`.

[30] Graves, A., Mohamed, A.-r., and Hinton, G. Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on* (2013), IEEE, pp. 6645–6649.

[31] Hairer, M., Hutzenthaler, M., and Jentzen, A. Loss of regularity for Kolmogorov equations. *Ann. Probab. 43*, 2 (03 2015), 468–527.

[32] Han, J., Jentzen, A., and E, W. Solving high-dimensional partial differential equations using deep learning. *arXiv:1707.02568* (2017), 13 pages.

[33] Hartman, E. J., Keeler, J. D., and Kowalski, J. M. Layered neural networks with gaussian hidden units as universal approximations. *Neural computation 2*, 2 (1990), 210–215.

[34] Henry-Labordere, P. Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM.

[35] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine 29*, 6 (2012), 82–97.

[36] Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks 4*, 2 (1991), 251 – 257.

[37] Hornik, K. Some new results on neural network approximation. *Neural networks 6*, 8 (1993), 1069–1072.

[38] Hornik, K., Stinchcombe, M., and White, H. Multilayer feedforward networks are universal approximators. *Neural Networks 2*, 5 (1989), 359–366.

[39] Hornik, K., Stinchcombe, M., and White, H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural networks 3*, 5 (1990), 551–560.

[40] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. Densely connected convolutional networks. In *CVPR* (2017), vol. 1, p. 3.

[41] Hytönen, T., van Neerven, J., Veraar, M., and Weis, L. *Analysis in Banach spaces. Vol. II*, vol. 67 of *Ergebnisse der Mathematik und*

ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]. Springer, Cham, 2017. Probabilistic methods and operator theory.

[42] Jentzen, A., and Kloeden, P. E. *Taylor approximations for stochastic partial differential equations*, vol. 83 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.

[43] Jones, F. *Lebesgue integration on Euclidean space*. Jones and Bartlett Publishers, Boston, MA, 1993.

[44] Kallenberg, O. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, 1997.

[45] Khoo, Y., Lu, J., and Ying, L. Solving parametric PDE problems with artificial neural networks. *arXiv:1707.03351* (2017), 17 pages.

[46] Klenke, A. *Probability theory*, second ed. Universitext. Springer, London, 2014. A comprehensive course.

[47] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105.

[48] LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. *nature 521*, 7553 (2015), 436.

[49] Leshno, M., Lin, V. Y., Pinkus, A., and Schocken, S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks 6*, 6 (1993), 861–867.

[50] McCulloch, W. S., and Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys. 5* (1943), 115–133.

[51] Mhaskar, H., and Micchelli, C. Degree of approximation by neural and translation networks with a single hidden layer. *Adv. Appl. Math. 16*, 2 (1995), 151–183.

[52] Mhaskar, H. N. Neural networks for optimal approximation of smooth and analytic functions. *Neural Comput. 8*, 1 (1996), 164–177.

[53] Mhaskar, H. N., and Poggio, T. Deep vs. shallow networks: An approximation theory perspective. *Analysis and Applications 14*, 06 (2016), 829–848.

[54] Mishra, S. A machine learning framework for data driven acceleration of computations of differential equations. *arXiv:1807.09519* (2018), 23 pages.

[55] Mittmann, K., and Steinwart, I. On the existence of continuous modifications of vector-valued random fields. *Georgian Math. J. 10*, 2 (2003), 311–317. Dedicated to the memory of Professor Revaz Chitashvili.

[56] Nabian, M. A., and Meidani, H. A Deep Neural Network Surrogate for High-Dimensional Random Partial Differential Equations. *arXiv:1806.02957* (2018), 23 pages.

[57] Nguyen-Thien, T., and Tran-Cong, T. Approximation of functions and their derivatives: A neural network implementation with applications. *Appl. Math. Model. 23*, 9 (1999), 687–704.

[58] Park, J., and Sandberg, I. W. Universal approximation using radial-basis-function networks. *Neural computation 3*, 2 (1991), 246–257.

[59] Perekrestenko, D., Grohs, P., Elbrächter, D., and Bölcskei, H. The universal approximation power of finite-width deep ReLU networks. *arXiv:1806.01528* (2018), 16 pages.

[60] Petersen, P., Raslan, M., and Voigtlaender, F. Topological properties of the set of functions generated by neural networks of fixed size. *arXiv:1806.08459* (2018), 45 pages.

[61] Petersen, P., and Voigtlaender, F. Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *arXiv:1709.05289* (2017), 54 pages.

[62] Pinkus, A. Approximation theory of the MLP model in neural networks. *Acta Numer. 8* (1999), 143–195.

[63] Priddy, K. L., and Keller, P. E. *Artificial neural networks: an introduction*, vol. 68. SPIE press, 2005.

[64] Raissi, M. Forward-Backward Stochastic Neural Networks: Deep Learning of High-dimensional Partial Differential Equations. *arXiv:1804.07010* (2018), 17 pages.

[65] Revuz, D., and Yor, M. *Continuous martingales and Brownian motion*, third ed., vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[66] Schmidhuber, J. Deep learning in neural networks: An overview. *Neural networks 61* (2015), 85–117.

[67] Schmitt, M. Lower bounds on the complexity of approximating continuous functions by sigmoidal neural networks. In *Proceedings of the 12th International Conference on Neural Information Processing Systems* (Cambridge, MA, USA, 1999), NIPS'99, MIT Press, pp. 328–334.

[68] Shaham, U., Cloninger, A., and Coifman, R. R. Provable approximation properties for deep neural networks. *Appl. Comput. Harmon. Anal.* (2015). to appear.

[69] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. Mastering the game of go with deep neural networks and tree search. *nature 529*, 7587 (2016), 484.

[70] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv:1712.01815* (2017), 19 pages.

[71] Simonyan, K., and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556* (2014), 14 pages.

[72] Sirignano, J., and Spiliopoulos, K. DGM: A deep learning algorithm for solving partial differential equations. *arXiv:1708.07469* (2017), 31 pages.

[73] Wu, C., Karanasou, P., Gales, M. J., and Sim, K. C. Stimulated deep neural network for speech recognition. Tech. rep., University of Cambridge Cambridge, 2016.

[74] Yarotsky, D. Error bounds for approximations with deep ReLU networks. *Neural Networks 94* (2017), 103–114.

[75] Yarotsky, D. Universal approximations of invariant maps by neural networks. *arXiv:1804.10306* (2018), 64 pages.