Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

Uniform error estimates for artificial neural network approximations for heat equations

L. Gonon and Ph. Grohs and A. Jentzen and D. Kofler and D. Šiška

Research Report No. 2019-61
November 2019

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

Uniform error estimates for artificial neural network approximations for heat equations

Lukas Gonon[1], Philipp Grohs[2], Arnulf Jentzen[3,4], David Kofler[5], and David Šiška[6]

[1] Faculty of Mathematics and Statistics, University of St. Gallen, Switzerland, e-mail: lukas.gonon@unisg.ch

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

[3] Department of Mathematics, ETH Zurich, Switzerland, e-mail: arnulf.jentzen@sam.math.ethz.ch

[4] Faculty of Mathematics and Computer Science, University of Münster, Germany, e-mail: ajentzen@uni-muenster.de

[5] Department of Mathematics, ETH Zurich, Switzerland, e-mail: david.kofler@zenai.ch

[6] School of Mathematics, University of Edinburgh, United Kingdom, e-mail: d.siska@ed.ac.uk

November 27, 2019

Abstract

Recently, artificial neural networks (ANNs) in conjunction with stochastic gradient descent optimization methods have been employed to approximately compute solutions of possibly rather high-dimensional partial differential equations (PDEs). Very recently, there have also been a number of rigorous mathematical results in the scientific literature which examine the approximation capabilities of such deep learning based approximation algorithms for PDEs. These mathematical results from the scientific literature prove in part that algorithms based on ANNs are capable of overcoming the curse of dimensionality in the numerical approximation of high-dimensional PDEs. In these mathematical results from the scientific literature usually the error between the solution of the PDE and the approximating ANN is measured in the L^p-sense with respect to some $p \in [1, \infty)$ and some probability measure. In many applications it is, however, also important to control the error in a uniform L^∞-sense. The key contribution of the main result of this article is to develop the techniques to obtain error estimates between solutions of PDEs and approximating ANNs in the uniform L^∞-sense. In particular, we prove that the number of parameters of an ANN to uniformly approximate the classical solution of the heat equation in a region $[a, b]^d$ for a fixed time point $T \in (0, \infty)$ grows at most polynomially in the dimension $d \in \mathbb{N}$ and the reciprocal of the approximation precision $\varepsilon > 0$. This shows that ANNs can overcome the curse of dimensionality in the numerical approximation of the heat equation when the error is measured in the uniform L^∞-norm.

Contents

1 Introduction

Artificial neural networks (ANNs) play a central role in machine learning applications such as computer vision (cf., e.g., [36,46,59]), speech recognition (cf., e.g., [27,35,62]), game intelligence (cf., e.g., [57,58]), and finance (cf., e.g., [7,12,60]). Recently, ANNs in conjunction with stochastic gradient descent optimization methods have also been employed to approximately compute solutions of possibly rather high-dimensional partial differential equations (PDEs); cf., for example, [4,5,6,7,8,10,13,14,17,18,19,22,23,26,32,33,34,37,41,48,49,50,53,54,61] and the references mentioned therein. The numerical simulation results in the above named references indicate that such deep learning based approximation methods for PDEs have the fundamental power to overcome the *curse of dimensionality* (cf., e.g., Bellman [9]) in the sense that the number of parameters of the approximating ANN grows at most polynomially in both the reciprocal of the prescribed approximation accuracy $\varepsilon > 0$ and the dimension $d \in \mathbb{N}$ of the considered PDE. Very recently, there have also been a number of rigorous mathematical results examining the approximation capabilities of these deep learning based approximation algorithms for PDEs (see, e.g., [11,20,28,30,33,38,43,47,55,61]). These works prove in part that algorithms based on ANNs are capable of overcoming the curse of dimensionality in the numerical approximation of high-dimensional PDEs. In particular, the works [11, 20, 28, 30, 38, 43, 47, 55] provide mathematical convergence results of such deep learning based numerical approximation methods for PDEs with dimension-independent error constants and convergence rates which depend on the dimension only polynomially.

Except of in the article Elbrächter et al. [20], in each of the approximation results in the above cited articles [11,28,30,33,38,43,47,55,61] the error between the solution of the PDE and the approximating ANN is measured in the L^p-sense with respect to some $p \in [1, \infty)$ and some probability measure. In many applications it is, however, also important to control the error in a uniform L^∞-sense. This is precisely the subject of this article. More specifically, it is the key contribution of Theorem 5.4 in Subsection 5.2 below, which is the main result of this article, to prove that ANNs can overcome the curse of dimensionality in the numerical approximation of the heat equation when the error is measured in the uniform L^∞-norm. The arguments used to prove the approximation results in the above cited articles, where the error between the solution of the PDE and the approximating ANN is measured in the L^p-sense with respect to some $p \in [1, \infty)$ and some probability measure, can not be employed for the uniform L^∞-norm approximation and the article Elbrächter et al. [20] is concerned with a specific class of PDEs so that the PDEs can essentially be solved analytically and the error analysis in [20] strongly exploits this explicit solution representation. The key contribution of the main result of this article, Theorem 5.4 in Subsection 5.2 below, is to develop the techniques to obtain error estimates between solutions of PDEs and approximating ANNs in the uniform L^∞-sense. To illustrate the findings of the main result of this article in more detail, we now present in the following theorem a special case of Theorem 5.4.

Theorem 1.1. *Let* $a \in \mathbb{R}$*,* $b \in (a, \infty)$*,* $c, T \in (0, \infty)$*,* $\mathbf{a} \in C^1(\mathbb{R}, \mathbb{R})$*, let* $\mathbf{A}_d \colon \mathbb{R}^d \to \mathbb{R}^d$*,* $d \in \mathbb{N}$*, satisfy for all* $d \in \mathbb{N}$*,* $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ *that* $\mathbf{A}_d(x) = (\mathbf{a}(x_1), \ldots, \mathbf{a}(x_d))$*, let* $\mathbf{N} = \cup_{L \in \mathbb{N} \cap [2,\infty)} \cup_{l_0,\ldots,l_L \in \mathbb{N}} \big(\times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \big)$*, let* $\mathcal{P} \colon \mathbf{N} \to \mathbb{N}$ *and* $\mathcal{R} \colon \mathbf{N} \to \cup_{m,n \in \mathbb{N}} C(\mathbb{R}^m, \mathbb{R}^n)$ *satisfy for all* $L \in \mathbb{N} \cap [2, \infty)$*,* $l_0, \ldots, l_L \in \mathbb{N}$*,* $\Phi = ((W_1, B_1), \ldots, (W_L, B_L)) \in (\times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$*,* $x_0 \in \mathbb{R}^{l_0}, \ldots, x_{L-1} \in \mathbb{R}^{l_{L-1}}$ *with* $\forall\, k \in \mathbb{N} \cap (0, L) \colon x_k = \mathbf{A}_{l_k}(W_k x_{k-1} + B_k)$ *that* $\mathcal{P}(\Phi) = \sum_{k=1}^{L} l_k(l_{k-1}+1)$*,* $(\mathcal{R}\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$*, and* $(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L$*, let* $\varphi_d \in C(\mathbb{R}^d, \mathbb{R})$*,* $d \in \mathbb{N}$*, let* $(\phi_{\varepsilon,d})_{(\varepsilon,d) \in (0,1] \times \mathbb{N}} \subseteq \mathbf{N}$*, let* $\|\cdot\| \colon \big(\cup_{d \in \mathbb{N}} \mathbb{R}^d \big) \to [0, \infty)$ *satisfy for all* $d \in \mathbb{N}$*,* $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ *that* $\|x\| = [\sum_{j=1}^{d} |x_j|^2]^{1/2}$*, and assume for all* $\varepsilon \in (0, 1]$*,* $d \in \mathbb{N}$*,* $x \in \mathbb{R}^d$ *that*

$$(\mathcal{R}\phi_{\varepsilon,d}) \in C(\mathbb{R}^d, \mathbb{R}), \qquad |(\mathcal{R}\phi_{\varepsilon,d})(x)| + \|(\nabla(\mathcal{R}\phi_{\varepsilon,d}))(x)\| \leq c d^c (1 + \|x\|^c), \quad (1)$$

and $\quad \mathcal{P}(\phi_{\varepsilon,d}) \leq cd^c\varepsilon^{-c}, \qquad |\varphi_d(x) - (\mathcal{R}\phi_{\varepsilon,d})(x)| \leq \varepsilon cd^c(1 + \|x\|^c)\,. \quad (2)$

Then

(i) there exist unique at most polynomially growing $u_d \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, *which satisfy for all* $d \in \mathbb{N}$, $t \in (0,T]$, $x \in \mathbb{R}^d$ *that* $u_d|_{(0,T]\times\mathbb{R}^d} \in C^{1,2}((0,T] \times \mathbb{R}^d, \mathbb{R})$, $u_d(0,x) = \varphi_d(x)$, *and*

$$(\tfrac{\partial}{\partial t}u_d)(t,x) = (\Delta_x u_d)(t,x) \tag{3}$$

and

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

$$\sup_{x\in[a,b]^d} |u_d(T,x) - (\mathcal{R}\psi_{\varepsilon,d})(x)| \leq \varepsilon. \tag{4}$$

Theorem 1.1 is an immediate consequence of Corollary 5.5 in Subsection 5.2 below. Corollary 5.5, in turn, follows from Theorem 5.4 in Subsection 5.2, the main result of the article. Let us add a few comments on some of the mathematical objects appearing in Theorem 1.1 above. The real number $T \in (0,\infty)$ denotes the time horizon on which we consider the heat equations in (3). The function $\mathbf{a} \in C^1(\mathbb{R},\mathbb{R})$ describes the activation function which we employ for the considered ANN approximations. In particular, Theorem 1.1 applies to ANNs with the standard logistic function as the activation function in which case the function $\mathbf{a} \in C^1(\mathbb{R},\mathbb{R})$ in Theorem 1.1 satisfies that for all $x \in \mathbb{R}$ it holds that $\mathbf{a}(x) = (1 + e^{-x})^{-1}$. The set $\mathbf{N}$ contains all possible ANNs, where each ANN is described abstractly in terms of the number of hidden layers, the number of nodes in each layer, and the values of the parameters (weights and biases in each layer), the function $\mathcal{P}\colon \mathbf{N} \to \mathbb{N}$ maps each ANN $\Phi \in \mathbf{N}$ to its total number of parameters $\mathcal{P}(\Phi)$, and the function $\mathcal{R}\colon \mathbf{N} \to \cup_{m,n\in\mathbb{N}} C(\mathbb{R}^m, \mathbb{R}^n)$ maps each ANN $\Phi \in \mathbf{N}$ to the actual function $(\mathcal{R}\Phi)$ (its realization) associated to Φ (cf., e.g., Grohs et al. [29, Section 2.1] and Petersen & Voigtlaender [52, Section 2]). Item (ii) in Theorem 1.1 above establishes under the hypotheses of Theorem 1.1 that the solution $u_d\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ of the heat equation can at time T be approximated by means of an ANN without the curse of dimensionality. To sum up, roughly speaking, Theorem 1.1 shows that if the initial conditions of the heat equations can be approximated well by ANNs, then the number of parameters of an ANN to uniformly approximate the classical solution of the heat equation in a region $[a,b]^d$ for a fixed time point $T \in (0,\infty)$ grows at most polynomially in the dimension $d \in \mathbb{N}$ and the reciprocal of the approximation precision $\varepsilon > 0$.

In our proof of Theorem 1.1 and Theorem 5.4, respectively, we employ several probabilistic and analytic arguments. In particular, we use the Feynman-Kac formula for viscosity solutions of Kolmogorov PDEs (cf., for example, Hairer et al. [31, Corollary 4.17]), Monte Carlo approximations (cf. Proposition 2.3), Sobolev type estimates for Monte Carlo approximations (cf. Lemma 2.16), the fact that stochastic differerential equations with affine linear coefficient functions are affine linear in the initial condition (cf. Grohs et al. [28, Proposition 2.20]), as well as an existence result for realizations of random variables (cf. Grohs et al. [28, Proposition 3.3]).

The remainder of this article is organized as follows. In Section 2 below preliminary results on Monte Carlo approximations together with Sobolev type estimates are established. Section 3 contains preliminary results on stochastic differential equations. In Section 4 we employ the results from Sections 2–3 to obtain uniform error estimates for ANN approximations. In Section 5 these uniform error estimates are used to prove Theorem 5.4 in Subsection 5.2 below, the main result of this article.

2 Sobolev and Monte Carlo estimates

2.1 Monte Carlo estimates

In this subsection we recall in Lemma 2.2 below an estimate for the $(p, 2)$-Kahane–Khintchine constant from the scientific literature (cf., for example, Cox et al. [16, Definition 5.4] or Grohs et al. [28, Definition 2.1]). Lemma 2.2, in particular, ensures that the $(p, 2)$-Kahane–Khintchine constant grows at most polynomially in p. Lemma 2.2 will be employed in the proof of Corollary 4.2 in Subsection 4.1 below. Our proof of Lemma 2.2 is based on an application of Hytönen et al. [40, Theorem 6.2.4] and is a slight extension of Grohs et al. [28, Lemma 2.2]. For completeness we also recall in Definition 2.1 below the notion of the Kahane–Khintchine constant (cf., e.g., Cox et al. [16, Definition 5.4]). Proposition 2.3 below is an L^p-approximation result for Monte-Carlo approximations. This L^p-approximation result for Monte-Carlo approximations is one of the main ingredients in our proof of Lemma 2.16 in Subsection 2.4 below. Proposition 2.3 is well-known in the literature and is proved, e.g., as Corollary 5.12 in Cox et al. [16].

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

$$\mathfrak{K}_p = \sup\left\{ c \in [0,\infty) \colon \left[\begin{array}{l} \exists\, \mathbb{R}\text{-Banach space } (E, \|\cdot\|_E) \colon \\ \exists \text{ probability space } (\Omega, \mathcal{F}, \mathbb{P}) \colon \\ \exists\, \mathbb{P}\text{-Rademacher family } r_j \colon \Omega \to \{-1, 1\},\, j \in \mathbb{N} \colon \\ \exists\, k \in \mathbb{N} \colon \exists x_1, \ldots, x_k \in E \setminus \{0\} \colon \\ \left(\mathbb{E}\left[\left\| \sum_{j=1}^k r_j x_j \right\|_E^p \right] \right)^{1/p} = c \left(\mathbb{E}\left[\left\| \sum_{j=1}^k r_j x_j \right\|_E^2 \right] \right)^{1/2} \end{array} \right] \right\} \tag{5}$$

and we call $\mathfrak{K}_p$ the $(p, 2)$-Kahane–Khintchine constant.

Lemma 2.2. *For every $p \in [1,\infty)$ let $\mathfrak{K}_p$ be the $(p,2)$-Kahane–Khintchine constant (cf. Definition 2.1). Then it holds for all $p \in [1,\infty)$ that $\mathfrak{K}_p \leq \sqrt{\max\{1, p-1\}}$.*

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 $r_j : \Omega \to \{-1, 1\}$, $j \in \mathbb{N}$, be independent random variables which satisfy for all $j \in \mathbb{N}$ that

$$\mathbb{P}(r_j = -1) = \mathbb{P}(r_j = 1) = \frac{1}{2}, \tag{6}$$

and let $k \in \mathbb{N}$, $x_1, \ldots, x_k \in E \backslash \{0\}$. Observe that Jensen's inequality ensures that for all $p \in [1,2]$ it holds that

$$\left(\mathbb{E}\left[\left\|\sum_{j=1}^{k} r_j x_j\right\|_E^p\right]\right)^{1/p} = \left(\mathbb{E}\left[\left\|\sum_{j=1}^{k} r_j x_j\right\|_E^{2\frac{p}{2}}\right]\right)^{1/p} \leq \left(\mathbb{E}\left[\left\|\sum_{j=1}^{k} r_j x_j\right\|_E^2\right]\right)^{1/2}. \tag{7}$$

Next note that Hytönen et al. [40, Theorem 6.2.4] (with $q = p$, $p = 2$ for $p \in (2,\infty)$ in the notation of Hytönen et al. [40, Theorem 6.2.4]) implies that for all $p \in (2,\infty)$ it holds that

$$\left(\mathbb{E}\left[\left\|\sum_{j=1}^{k} r_j x_j\right\|_E^p\right]\right)^{1/p} \leq \sqrt{p-1}\left(\mathbb{E}\left[\left\|\sum_{j=1}^{k} r_j x_j\right\|_E^2\right]\right)^{1/2}. \tag{8}$$

Combining this with (7) demonstrates that for all $p \in [1,\infty)$ it holds that

$$\left(\mathbb{E}\left[\left\|\sum_{j=1}^{k} r_j x_j\right\|_E^p\right]\right)^{1/p} \leq \sqrt{\max\{1, p-1\}}\left(\mathbb{E}\left[\left\|\sum_{j=1}^{k} r_j x_j\right\|_E^2\right]\right)^{1/2}. \tag{9}$$

The proof of Lemma 2.2 is thus completed. □

Proposition 2.3. *Let $d, n \in \mathbb{N}$, $p \in [2,\infty)$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $\mathfrak{K}_p \in (0,\infty)$ be the $(p,2)$-Kahane–Khintchine constant (cf. Definition 2.1), let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X_i : \Omega \to \mathbb{R}^d$, $i \in \{1, \ldots, n\}$, be i.i.d. random variables with $\mathbb{E}\big[\|X_1\|\big] < \infty$. Then*

$$\left(\mathbb{E}\left[\left\|\mathbb{E}[X_1] - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)\right\|^p\right]\right)^{1/p} \leq \frac{2\,\mathfrak{K}_p\left(\mathbb{E}[\|X_1 - \mathbb{E}[X_1]\|^p]\right)^{1/p}}{\sqrt{n}}. \tag{10}$$

2.2 Volumes of the Euclidean unit balls

In this subsection we provide in Corollary 2.8 below an elementary and well-known upper bound for the volumes of the Euclidean unit balls. Corollary 2.8

will be used in our proof of Corollary 2.14 in Subsection 2.3 below. Our proof of Corollary 2.8 employs the elementary and well-known results in Lemmas 2.4–2.7 below. For completeness we also provide in this subsection detailed proofs for Lemmas 2.4–2.7 and Corollary 2.8.

Lemma 2.4. *Let* $\Gamma\colon (0,\infty) \to (0,\infty)$ *and* $\mathrm{B}\colon (0,\infty)^2 \to (0,\infty)$ *satisfy for all* $x, y \in (0,\infty)$ *that* $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$ *and* $\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$ *and let* $x, y \in (0,\infty)$. *Then*

(i) it holds that $\Gamma(x+1) = x\Gamma(x)$,

(ii) it holds that $\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$,

(iii) it holds that $\Gamma(1) = 1$ *and* $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, *and*

(iv) it holds that

$$\sqrt{\frac{2\pi}{x}}\left[\frac{x}{e}\right]^x \leq \Gamma(x) \leq \sqrt{\frac{2\pi}{x}}\left[\frac{x}{e}\right]^x e^{\frac{1}{12x}}. \tag{11}$$

Proof of Lemma 2.4. Throughout this proof let $\Phi\colon (0,\infty)\times(0,1) \to (0,\infty)^2$ satisfy for all $s \in (0,\infty)$, $t \in (0,1)$ that

$$\Phi(s,t) = (s(1-t), st) \tag{12}$$

and let $f\colon (0,\infty)^2 \to (0,\infty)$ satisfy for all $s, t \in (0,\infty)$ that

$$f(s,t) = s^{x-1}t^{y-1}e^{-(s+t)}. \tag{13}$$

Observe that integration by parts shows that

$$\Gamma(x+1) = \int_0^\infty t^x e^{-t}\,dt = \left[-t^x e^{-t}\right]_{t=0}^{t=\infty} + x\int_0^\infty t^{x-1}e^{-t}\,dt = x\Gamma(x). \tag{14}$$

This establishes item (i). Next note that Fubini's theorem ensures that

$$\begin{aligned}\Gamma(x)\Gamma(y) &= \int_0^\infty s^{x-1}e^{-s}\,ds \int_0^\infty t^{y-1}e^{-t}\,dt \\ &= \int_0^\infty\int_0^\infty s^{x-1}t^{y-1}e^{-(s+t)}\,ds\,dt = \int_0^\infty\int_0^\infty f(s,t)\,ds\,dt\,.\end{aligned} \tag{15}$$

Moreover, note that for all $s \in (0,\infty)$, $t \in (0,1)$ it holds that

$$\det(\Phi'(s,t)) = s \in (0,\infty). \tag{16}$$

This, (15), the integral transformation theorem (cf., for example, [15, Theorem 6.1.7]), and Fubini's theorem prove that

$$\begin{aligned}\Gamma(x)\Gamma(y) &= \int_0^\infty\int_0^1 f(\Phi(s,t))\,|\det(\Phi'(s,t))|\,dt\,ds \\ &= \int_0^\infty\int_0^1 (s(1-t))^{x-1}(st)^{y-1}e^{-(s(1-t)+st)}s\,dt\,ds \\ &= \int_0^\infty s^{x+y-1}e^{-s}\,ds\int_0^1 t^{y-1}(1-t)^{x-1}\,dt = \Gamma(x+y)\mathrm{B}(y,x).\end{aligned} \tag{17}$$

Next note that the integral transformation theorem with the diffeomorphism $(0,1) \ni t \mapsto (1-t) \in (0,1)$ ensures that

$$\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt = \int_0^1 (1-t)^{x-1}t^{y-1}\,dt = \mathrm{B}(y,x). \tag{18}$$

Combining this with (17) establishes item (ii). Next note that

$$\Gamma(1) = \int_0^\infty t^{1-1}e^{-t}\,dt = \int_0^\infty e^{-t}\,dt = 1. \tag{19}$$

Item (ii) and the integral transformation theorem with the diffeomorphism $(0,\frac{\pi}{2}) \ni t \mapsto [\sin(t)]^2 \in (0,1)$ therefore show that

$$\frac{\left[\Gamma\left(\frac{1}{2}\right)\right]^2}{\Gamma(1)} = \mathrm{B}\left(\tfrac{1}{2},\tfrac{1}{2}\right) = \int_0^1 t^{-1/2}(1-t)^{-1/2}\,dt = \int_0^{\frac{\pi}{2}} 2\,dt = \pi. \tag{20}$$

Combining this with (19) establishes item (iii). Next note that Artin [3, Chapter 3, (3.9)] ensures that there exists $\mu\colon(0,\infty)\to\mathbb{R}$ which satisfies for all $t\in(0,\infty)$ that $0<\mu(t)<\frac{1}{12t}$ and

$$\Gamma(t) = \sqrt{2\pi}t^{t-1/2}e^{-t}e^{\mu(t)}. \tag{21}$$

Hence, we obtain that

$$\sqrt{2\pi}x^{x-1/2}e^{-x} \le \Gamma(x) \le \sqrt{2\pi}x^{x-1/2}e^{-x}e^{\frac{1}{12x}}. \tag{22}$$

This establishes item (iv). The proof of Lemma 2.4 is thus completed. □

Lemma 2.5. *Let* $\mathrm{B}\colon(0,\infty)^2\to(0,\infty)$ *satisfy for all* $x,y\in(0,\infty)$ *that* $\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$. *Then it holds for all* $p\in[0,\infty)$ *that*

$$\int_0^{\frac{\pi}{2}} [\sin(t)]^p\,dt = \frac{\mathrm{B}\left(\frac{p+1}{2},\frac{1}{2}\right)}{2}. \tag{23}$$

Proof of Lemma 2.5. First, note that for all $t\in(0,1)$ it holds that

$$\arcsin'(t) = (1-t^2)^{-1/2}. \tag{24}$$

This and the integral transformation theorem with the diffeomorphism $(0,1) \ni t \mapsto \arcsin(t) \in (0,\frac{\pi}{2})$ ensure that for all $p\in[0,\infty)$ it holds that

$$\int_0^{\frac{\pi}{2}} [\sin(t)]^p\,dt = \int_0^1 t^p(1-t^2)^{-1/2}\,dt. \tag{25}$$

The integral transformation theorem with the diffeomorphism $(0,1)\ni t\mapsto\sqrt{t}\in(0,1)$ hence implies that for all $p\in[0,\infty)$ it holds that

$$\begin{aligned}\int_0^{\frac{\pi}{2}} [\sin(t)]^p\,dt &= \frac{1}{2}\int_0^1 t^{p/2-1/2}(1-t)^{-1/2}\,dt\\ &= \frac{1}{2}\int_0^1 t^{(p+1)/2-1}(1-t)^{1/2-1}\,dt = \frac{\mathrm{B}\left(\frac{p+1}{2},\frac{1}{2}\right)}{2}.\end{aligned} \tag{26}$$

The proof of Lemma 2.5 is thus completed. □

Lemma 2.6. *Let $R \in (0,\infty]$, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, for every $d \in \{2,3,\ldots\}$ let $B_d = \{x \in \mathbb{R}^d \colon \|x\|_{\mathbb{R}^d} < R\}$ and*

$$S_d = \begin{cases} (0,2\pi) & \text{for } d = 2 \\ (0,2\pi) \times (0,\pi)^{d-2} & \text{for } d \in \{3,4,\ldots\}, \end{cases} \tag{27}$$

and let $T_d \colon (0,R) \times S_d \to \mathbb{R}^d$, $d \in \{2,3,\ldots\}$, satisfy for all $d \in \{2,3,\ldots\}$, $r \in (0,R)$, $\varphi \in (0,2\pi)$, $\vartheta_1,\ldots,\vartheta_{d-2} \in (0,\pi)$ that if $d = 2$ then $T_2(r,\varphi) = r(\cos(\varphi),\sin(\varphi))$ and if $d \geq 3$ then

$$\begin{aligned} T_d(r,\varphi,\vartheta_1,\ldots,\vartheta_{d-2}) = r\Big(&\cos(\varphi)\left[\textstyle\prod_{i=1}^{d-2}\sin(\vartheta_i)\right], \sin(\varphi)\left[\textstyle\prod_{i=1}^{d-2}\sin(\vartheta_i)\right], \\ &\cos(\vartheta_1)\left[\textstyle\prod_{i=2}^{d-2}\sin(\vartheta_i)\right], \ldots, \cos(\vartheta_{d-3})\sin(\vartheta_{d-2}), \cos(\vartheta_{d-2})\Big). \end{aligned} \tag{28}$$

Then

(i) it holds for all $r \in (0,R)$, $\varphi \in (0,2\pi)$ that

$$|\det((T_2)'(r,\varphi))| = r, \tag{29}$$

(ii) it holds for all $d \in \{3,4,\ldots\}$, $r \in (0,R)$, $\varphi \in (0,2\pi)$, $\vartheta_1,\ldots,\vartheta_{d-2} \in (0,\pi)$ that

$$|\det((T_d)'(r,\varphi,\vartheta_1,\ldots,\vartheta_{d-2}))| = r^{d-1}\left[\textstyle\prod_{i=1}^{d-2}[\sin(\vartheta_i)]^i\right], \tag{30}$$

and

(iii) it holds for all $d \in \{2,3,\ldots\}$ and all $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}([0,\infty))$-measurable functions $f \colon \mathbb{R}^d \to [0,\infty)$ that

$$\int_{B_d} f(x)\,dx = \int_0^R \int_{S_d} f(T_d(r,\phi))\,|\det((T_d)'(r,\phi))|\,d\phi\,dr\,. \tag{31}$$

Proof of Lemma 2.6. Throughout this proof for every $d \in \{2,3,\ldots\}$ let $\lambda_{\mathbb{R}^d} : \mathcal{B}(\mathbb{R}^d) \to [0,\infty]$ be the Lebesgue–Borel measure on $\mathbb{R}^d$. Observe that for all $r \in (0,R)$, $\varphi \in (0,2\pi)$ it holds that

$$(T_2)'(r,\varphi) = \begin{pmatrix} \cos(\varphi) & -r\sin(\varphi) \\ \sin(\varphi) & r\cos(\varphi) \end{pmatrix}. \tag{32}$$

Hence, we obtain that for all $r \in (0,R)$, $\varphi \in (0,2\pi)$ it holds that

$$\big|\det\big((T_2)'(r,\varphi)\big)\big| = r[\cos(\varphi)]^2 + r[\sin(\varphi)]^2 = r. \tag{33}$$

This establishes item (i). Next observe that Amann & Escher [2, Ch. X, Lemma 8.8] establishes item (ii). To establish item (iii) we distinguish between the case $d = 2$ and the case $d \in \{3,4,\ldots\}$. First, we consider the case $d = 2$. Note that $T_2 \colon (0,R) \times (0,2\pi) \to T_2((0,R) \times (0,2\pi))$ is a bijective function. This and item

(i) show that $T_2\colon (0,R)\times(0,2\pi)\to T_2((0,R)\times(0,2\pi))$ is a diffeomorphism. Next observe that

$$T_2((0,R)\times(0,2\pi)) = B_2\backslash\{(x,0)\colon x\in[0,\infty)\}. \tag{34}$$

The fact that $T_2\colon (0,R)\times(0,2\pi)\to T_2((0,R)\times(0,2\pi))$ is a diffeomorphism, the fact that $\lambda_{\mathbb{R}^2}(\{(x,0)\colon x\in[0,\infty)\})=0$, and the integral transformation theorem hence demonstrate that for all $\mathcal{B}(\mathbb{R}^2)/\mathcal{B}([0,\infty))$-measurable functions $f\colon\mathbb{R}^2\to[0,\infty)$ it holds that

$$\int_{B_2} f(x)\,dx = \int_0^R\int_{S_2} f(T_2(r,\phi))\,\big|\det\big((T_2)'(r,\phi)\big)\big|\,d\phi\,dr\,. \tag{35}$$

This establishes item (iii) in the case $d=2$. Next we consider the case $d\in\{3,4,\ldots\}$. Note that Amann & Escher [2, Ch. X, Lemma 8.8] implies that $T_d\colon (0,R)\times(0,2\pi)\times(0,\pi)^{d-2}\to T_d((0,R)\times(0,2\pi)\times(0,\pi)^{d-2})$ is a diffeomorphism with

$$T_d((0,R)\times(0,2\pi)\times(0,\pi)^{d-2}) = B_d\backslash\big([0,\infty)\times\{0\}\times\mathbb{R}^{d-2}\big). \tag{36}$$

The fact that $\lambda_{\mathbb{R}^d}\big([0,\infty)\times\{0\}\times\mathbb{R}^{d-2}\big)=0$ and the integral transformation theorem hence show that for all $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}([0,\infty))$-measurable functions $f\colon\mathbb{R}^d\to[0,\infty)$ it holds that

$$\int_{B_d} f(x)\,dx = \int_0^R\int_{S_d} f(T_d(r,\phi))\,\big|\det\big((T_d)'(r,\phi)\big)\big|\,d\phi\,dr\,. \tag{37}$$

This establishes item (iii) in the case $d\in\{3,4,\ldots\}$. The proof of Lemma 2.6 is thus completed. □

Lemma 2.7. *Let $d\in\mathbb{N}$, let $\|\cdot\|:\mathbb{R}^d\to[0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $\lambda:\mathcal{B}(\mathbb{R}^d)\to[0,\infty]$ be the Lebesgue–Borel measure on $\mathbb{R}^d$, let $\mathbb{B}\subseteq\mathbb{R}^d$ be the set given by $\mathbb{B}=\{x\in\mathbb{R}^d\colon \|x\|<1\}$, and let $\Gamma\colon(0,\infty)\to(0,\infty)$ satisfy for all $x\in(0,\infty)$ that $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}\,dt$. Then*

(i) for $d\in\{2,3,\ldots\}$ it holds that

$$\lambda(\mathbb{B}) = \frac{2\pi}{d}\left[\prod_{i=1}^{d-2}\int_0^\pi [\sin(\vartheta_i)]^i\,d\vartheta_i\right] \tag{38}$$

and

(ii) it holds that

$$\lambda(\mathbb{B}) = \frac{\pi^{d/2}}{\Gamma\big(\frac{d}{2}+1\big)}. \tag{39}$$

Proof of Lemma 2.7. To establish (38) and (39) we distinguish between the case $d = 1$, the case $d = 2$, and the case $d \geq 3$. First, we consider the case $d = 1$. Note that items (i) and (iii) in Lemma 2.4 show that

$$\Gamma\big(\tfrac{1}{2} + 1\big) = \frac{\Gamma\big(\frac{1}{2}\big)}{2} = \frac{\pi^{1/2}}{2}. \tag{40}$$

This implies that

$$\frac{\pi^{1/2}}{\Gamma\big(\frac{1}{2} + 1\big)} = 2. \tag{41}$$

Combining this and the fact that $\lambda(\mathbb{B}) = \lambda((-1, 1)) = 2$ establishes (39) in the case $d = 1$. Next we consider the case $d = 2$. Note that items (i) and (iii) in Lemma 2.6 and Fubini's theorem prove that

$$\lambda(\mathbb{B}) = \int_{\mathbb{B}} dx = \int_0^{2\pi} \int_0^1 r \, dr \, d\varphi = \pi. \tag{42}$$

Next note that items (i) and (iii) in Lemma 2.4 show that

$$\Gamma(2) = \Gamma(1 + 1) = \Gamma(1) = 1. \tag{43}$$

This implies that

$$\frac{\pi}{\Gamma(1 + 1)} = \pi. \tag{44}$$

Combining this with (42) establishes (38) and (39) in the case $d = 2$. Next we consider the case $d \geq 3$. Note that items (ii)–(iii) in Lemma 2.6 and Fubini's theorem ensure that

$$\begin{aligned} \lambda(\mathbb{B}) &= \int_{\mathbb{B}} dx \\ &= \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} \int_0^1 r^{d-1} \left[\prod_{i=1}^{d-2} [\sin(\vartheta_i)]^i \right] dr \, d\varphi \, d\vartheta_1 \cdots d\vartheta_{d-2} \\ &= \frac{1}{d} \int_0^{2\pi} d\varphi \left[\prod_{i=1}^{d-2} \int_0^{\pi} [\sin(\vartheta_i)]^i \, d\vartheta_i \right]. \end{aligned} \tag{45}$$

This establishes (38) in the case $d \in \{3, 4, \ldots\}$. Moreover, note that (45) and the fact that for all $k \in \mathbb{N}$ it holds that $\int_0^{\pi} [\sin(t)]^k \, dt = 2 \int_0^{\frac{\pi}{2}} [\sin(t)]^k \, dt$ show that

$$\lambda(\mathbb{B}) = \frac{4}{d} \int_0^{\frac{\pi}{2}} d\varphi \left[\prod_{i=1}^{d-2} 2 \int_0^{\frac{\pi}{2}} [\sin(\vartheta_i)]^i \, d\vartheta_i \right]. \tag{46}$$

Combining this, Lemma 2.5 (with $p = i$ for $i \in \{0, \ldots, d-2\}$ in the notation of

Lemma 2.5), and item (ii) in Lemma 2.4 demonstrates that

$$\begin{aligned}\lambda(\mathbb{B}) &= \frac{2}{d}\mathrm{B}\big(\tfrac{1}{2},\tfrac{1}{2}\big)\left[\prod_{i=1}^{d-2}\mathrm{B}\big(\tfrac{i+1}{2},\tfrac{1}{2}\big)\right] \\ &= \frac{2}{d}\frac{\Gamma\big(\frac{1}{2}\big)\,\Gamma\big(\frac{1}{2}\big)}{\Gamma(1)}\left[\prod_{i=1}^{d-2}\frac{\Gamma\big(\frac{i+1}{2}\big)\,\Gamma\big(\frac{1}{2}\big)}{\Gamma\big(\frac{i+2}{2}\big)}\right] = \frac{2}{d}\frac{\big[\Gamma\big(\frac{1}{2}\big)\big]^d}{\Gamma\big(\frac{d}{2}\big)}.\end{aligned} \tag{47}$$

Items (i) and (iii) in Lemma 2.4 hence show that

$$\lambda(\mathbb{B}) = \frac{\big[\Gamma\big(\frac{1}{2}\big)\big]^d}{\frac{d}{2}\Gamma\big(\frac{d}{2}\big)} = \frac{\pi^{d/2}}{\Gamma\big(\frac{d}{2}+1\big)}. \tag{48}$$

This establishes (39) in the case $d \in \{3, 4, \ldots\}$. The proof of Lemma 2.7 is thus completed. ☐

Corollary 2.8. *Let $d \in \mathbb{N}$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, and let $\lambda : \mathcal{B}(\mathbb{R}^d) \to [0,\infty]$ be the Lebesgue–Borel measure on $\mathbb{R}^d$. Then*

$$\lambda\big(\{x \in \mathbb{R}^d \colon \|x\| < 1\}\big) \le \frac{1}{\sqrt{d\pi}}\left[\frac{2\pi e}{d}\right]^{d/2}. \tag{49}$$

Proof of Lemma 2.8. Throughout this proof let $\Gamma\colon (0,\infty) \to (0,\infty)$ satisfy for all $x \in (0,\infty)$ that $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$. Note that Lemma 2.7 shows that

$$\lambda\big(\{x \in \mathbb{R}^d \colon \|x\| < 1\}\big) = \frac{\pi^{\frac{d}{2}}}{\Gamma\big(\frac{d}{2}+1\big)}. \tag{50}$$

Moreover, note that items (i) and (iv) in Lemma 2.4 imply that for all $x \in (0,\infty)$ it holds that

$$\Gamma(x+1) = x\Gamma(x) \ge \sqrt{2\pi x}\left[\frac{x}{e}\right]^x. \tag{51}$$

Hence, we obtain that

$$\frac{1}{\Gamma\big(\frac{d}{2}+1\big)} \le \frac{1}{\sqrt{d\pi}}\left[\frac{2e}{d}\right]^{d/2}. \tag{52}$$

Combining this with (50) demonstrates that

$$\lambda\big(\{x \in \mathbb{R}^d \colon \|x\| < 1\}\big) \le \frac{1}{\sqrt{d\pi}}\left[\frac{2\pi e}{d}\right]^{d/2}. \tag{53}$$

The proof of Lemma 2.8 is thus completed. ☐

2.3 Sobolev type estimates for smooth functions

In this subsection we present in Corollary 2.15 below a Sobolev type estimate for smooth functions with an explicit and dimension-independent constant in the Sobolev type estimate. Corollary 2.15 will be employed in our proof of Corollary 4.2 in Subsection 4.1 below. Corollary 2.15 is a consequence of the elementary results in Lemma 2.9 and Corollary 2.14 below. Corollary 2.14, in turn, follows from Proposition 2.13 below. Proposition 2.13 is a special case of Mizuguchi et al. [51, Theorem 2.1]. For completeness we also provide in this subsection a proof for Proposition 2.13. Our proof of Proposition 2.13 employs the well-known results in Lemmas 2.10–2.12 below. Results similar to Lemmas 2.10–2.12 can, e.g., be found in Mizuguchi et al. [51, Lemma 3.1, Theorem 3.3, and Theorem 3.4] and Gilbarg & Trudinger [25, Lemma 7.16].

Lemma 2.9. *Let $\Phi \in C^1((0,1),\mathbb{R})$ satisfy that $\int_0^1 (|\Phi(x)| + |\Phi'(x)|)\,dx < \infty$. Then*

$$\sup_{x\in(0,1)} |\Phi(x)| \leq \int_0^1 (|\Phi(x)| + |\Phi'(x)|)\,dx\,. \tag{54}$$

Proof of Lemma 2.9. First, note that the fundamental theorem of calculus ensures that for all $x \in (0,1)$ it holds that

$$\Phi(x) = \int_0^1 \left[\Phi(s) - \int_x^s \Phi'(t)\,dt\right] ds. \tag{55}$$

The triangle inequality and the hypothesis that $\int_0^1 (|\Phi(x)| + |\Phi'(x)|)\,dx < \infty$ hence show that for all $x \in (0,1)$ it holds that

$$\begin{aligned} |\Phi(x)| &= \left|\int_0^1 \left[\Phi(s) - \int_x^s \Phi'(t)\,dt\right] ds\right| \\ &\leq \int_0^1 \left[|\Phi(s)| + \int_x^s |\Phi'(t)|\,dt\right] ds \\ &\leq \int_0^1 |\Phi(s)|\,ds + \int_0^1 |\Phi'(t)|\,dt < \infty. \end{aligned} \tag{56}$$

This implies (54). The proof of Lemma 2.9 is thus completed. ☐

Lemma 2.10. *Let $d \in \{2,3,\ldots\}$, $p \in (d,\infty)$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $\lambda : \mathcal{B}(\mathbb{R}^d) \to [0,\infty]$ be the Lebesgue–Borel measure on $\mathbb{R}^d$, and let $W \subseteq \mathbb{R}^d$ be a non-empty, bounded, and open set. Then*

$$\begin{aligned} &\int_{\cup_{x\in W}\{y-x:\, y\in W\}} \|z\|^{(1-d)\frac{p}{p-1}}\,dz \\ &\leq \frac{d(p-1)}{p-d}\left[\sup_{v,w\in W}\|v-w\|\right]^{\frac{p-d}{p-1}} \lambda\big(\{x\in\mathbb{R}^d\colon \|x\|<1\}\big). \end{aligned} \tag{57}$$

Proof of Lemma 2.10. Throughout this proof let $\rho \in (0,\infty)$ satisfy that $\rho = \sup_{v,w\in W} \|v-w\|$, let $V \subseteq \mathbb{R}^d$ be the set given by $V = \cup_{x\in W}\{y-x\colon y \in W\}$, let $S \subseteq \mathbb{R}^{d-1}$ be the set given by

$$S = \begin{cases} (0,2\pi) & : d = 2 \\ (0,2\pi)\times(0,\pi)^{d-2} & : d \in \{3,4,\ldots\}, \end{cases} \tag{58}$$

let $B_r \subseteq \mathbb{R}^d$, $r \in (0,\infty)$, be the sets which satisfy for all $r \in (0,\infty)$ that $B_r = \{x \in \mathbb{R}^d\colon \|x\| < r\}$, and let $T_R\colon (0,R)\times S \to \mathbb{R}^d$, $R \in (0,\infty]$, satisfy for all $R \in (0,\infty]$, $r \in (0,R)$, $\varphi \in (0,2\pi)$, $\vartheta_1,\ldots,\vartheta_{d-2} \in (0,\pi)$ that if $d = 2$ then $T_R(r,\varphi) = r(\cos(\varphi),\sin(\varphi))$ and if $d \in \{3,4,\ldots\}$ then

$$\begin{split} T_R(r,\varphi,\vartheta_1,\ldots,\vartheta_{d-2}) = r\Big(&\cos(\varphi)\left[\textstyle\prod_{i=1}^{d-2}\sin(\vartheta_i)\right],\sin(\varphi)\left[\textstyle\prod_{i=1}^{d-2}\sin(\vartheta_i)\right], \\ &\cos(\vartheta_1)\left[\textstyle\prod_{i=2}^{d-2}\sin(\vartheta_i)\right],\ldots,\cos(\vartheta_{d-3})\sin(\vartheta_{d-2}),\cos(\vartheta_{d-2})\Big). \end{split} \tag{59}$$

Observe that

$$(1-d)\frac{p}{p-1} + d = \frac{(1-d)p + d(p-1)}{p-1} = \frac{p-d}{p-1}\,. \tag{60}$$

Next note that items (i)–(iii) in Lemma 2.6 and the fact that for all $r \in (0,\rho)$, $\phi \in S$ it holds that $\|T_\rho(r,\phi)\| = r$ show that

$$\begin{split} \int_{B_\rho} \|x\|^{(1-d)\frac{p}{p-1}}\,dx &= \int_0^\rho \int_S r^{(1-d)\frac{p}{p-1}} \left|\det\left((T_\rho)'(r,\phi)\right)\right|\,d\phi\,dr \\ &= \int_0^\rho \int_S r^{(1-d)\frac{p}{p-1}} r^{d-1} \left|\det\left((T_\infty)'(1,\phi)\right)\right|\,d\phi\,dr\,. \end{split} \tag{61}$$

The fact that $V \subseteq B_\rho$, Fubini's theorem, and (60) hence demonstrate that

$$\begin{split} \int_V \|x\|^{(1-d)\frac{p}{p-1}}\,dx &\leq \int_{B_\rho} \|x\|^{(1-d)\frac{p}{p-1}}\,dx \\ &= \int_S \left(\int_0^\rho r^{(1-d)\frac{p}{p-1}} r^{d-1}\,dr\right) \left|\det\left((T_\infty)'(1,\phi)\right)\right|\,d\phi \\ &= \rho^{\frac{p-d}{p-1}}\frac{(p-1)}{p-d}\int_S \left|\det\left((T_\infty)'(1,\phi)\right)\right|\,d\phi\,. \end{split} \tag{62}$$

This, Lemma 2.6, and Lemma 2.7 hence prove that

$$\begin{split} &\int_V \|x\|^{(1-d)\frac{p}{p-1}}\,dx \leq \rho^{\frac{p-d}{p-1}}\frac{(p-1)}{p-d}\int_S \left|\det\left((T_\infty)'(1,\phi)\right)\right|\,d\phi \\ &= \rho^{\frac{p-d}{p-1}}\frac{(p-1)}{p-d}\int_0^{2\pi}\int_0^\pi \cdots \int_0^\pi \left[\textstyle\prod_{i=1}^{d-2}[\sin(\vartheta_i)]^i\right]\,d\vartheta_1\cdots d\vartheta_{d-2}\,d\varphi \\ &= \rho^{\frac{p-d}{p-1}}\frac{(p-1)}{p-d}d\frac{2\pi}{d}\prod_{i=1}^{d-2}\int_0^\pi [\sin(\vartheta_i)]^i\,d\vartheta_i \;= \rho^{\frac{p-d}{p-1}}\frac{d(p-1)}{p-d}\lambda(B_1)\,. \end{split} \tag{63}$$

The proof of Lemma 2.10 is thus completed. ∎

Lemma 2.11. *Let $d \in \{2, 3, \ldots\}$, $p \in (d, \infty)$, let $W \subseteq \mathbb{R}^d$ be a non-empty, open, bounded, and convex set, let $\Phi \in C^1(W, \mathbb{R})$, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the standard norm on $\mathbb{R}^d$, let $\lambda : \mathcal{B}(\mathbb{R}^d) \to [0, \infty]$ be the Lebesgue–Borel measure on $\mathbb{R}^d$, and assume that $\int_W (|\Phi(x)|^p + \|(\nabla\Phi)(x)\|^p)\, dx < \infty$. Then it holds for all $x \in W$ that*

$$\begin{aligned} &\left|\lambda(W)\Phi(x) - \int_W \Phi(y)\, dy\right| \\ &\leq \frac{1}{d}\left[\sup_{v,w\in W} \|v - w\|\right]^d \int_W \|(\nabla\Phi)(y)\| \, \|x - y\|^{1-d}\, dy. \end{aligned} \tag{64}$$

Proof of Lemma 2.11. Throughout this proof let $x \in W$, let $\rho \in (0, \infty)$ satisfy that $\rho = \sup_{v,w\in W} \|v - w\|$, let $\langle \cdot, \cdot \rangle : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, let $S \subseteq \mathbb{R}^{d-1}$ be the set given by

$$S = \begin{cases} (0, 2\pi) & : d = 2 \\ (0, 2\pi) \times (0, \pi)^{d-2} & : d \in \{3, 4, \ldots\}, \end{cases} \tag{65}$$

let $(\omega_{v,w})_{(v,w)\in W^2} \subseteq \mathbb{R}^d$ satisfy for all $v, w \in W$ that

$$\omega_{v,w} = \begin{cases} \frac{w-v}{\|w-v\|} & : v \neq w \\ 0 & : v = w, \end{cases} \tag{66}$$

let $B_r \subseteq \mathbb{R}^d$, $r \in (0, \infty)$, satisfy for all $r \in (0, \infty)$ that $B_r = \{y \in \mathbb{R}^d : \|y - x\| < r\}$, let $E : \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $y \in \mathbb{R}^d$ that

$$E(y) = \begin{cases} (\nabla\Phi)(y) & : y \in W \\ 0 & : y \in \mathbb{R}^d \backslash W, \end{cases} \tag{67}$$

and let $T_R : (0, R) \times S \to \mathbb{R}^d$, $R \in (0, \infty]$, satisfy for all $R \in (0, \infty]$, $r \in (0, R)$, $\varphi \in (0, 2\pi)$, $\vartheta_1, \ldots, \vartheta_{d-2} \in (0, \pi)$ that if $d = 2$ then $T_R(r, \varphi) = r(\cos(\varphi), \sin(\varphi))$ and if $d \geq 3$ then

$$\begin{aligned} &T_R(r, \varphi, \vartheta_1, \ldots, \vartheta_{d-2}) = r\Big(\cos(\varphi)\left[\textstyle\prod_{i=1}^{d-2} \sin(\vartheta_i)\right], \sin(\varphi)\left[\textstyle\prod_{i=1}^{d-2} \sin(\vartheta_i)\right], \\ &\cos(\vartheta_1)\left[\textstyle\prod_{i=2}^{d-2} \sin(\vartheta_i)\right], \ldots, \cos(\vartheta_{d-3})\sin(\vartheta_{d-2}), \cos(\vartheta_{d-2})\Big). \end{aligned} \tag{68}$$

Observe that the hypothesis that $\int_W (|\Phi(y)|^p + \|(\nabla\Phi)(y)\|^p)\, dy < \infty$, the hypothesis that W is bounded, and Hölder's inequality ensure that

$$\begin{aligned} &\int_W |\Phi(y)| + \|(\nabla\Phi)(y)\|\, dy \\ &\leq [\lambda(W)]^{(p-1)/p}\left(\left[\int_W |\Phi(y)|^p\, dy\right]^{1/p} + \left[\int_W \|(\nabla\Phi)(y)\|^p\, dy\right]^{1/p}\right) < \infty. \end{aligned} \tag{69}$$

Next note that the assumption that W is convex and the fundamental theorem of calculus demonstrate that for all $y \in W$ it holds that

$$\Phi(x) - \Phi(y) = -(\Phi(x + r\omega_{x,y}))|_{r=0}^{r=\|y-x\|} = -\int_0^{\|y-x\|} \frac{d}{dr}\Phi(x + r\omega_{x,y})\, dr. \tag{70}$$

The fact that $\lambda(W) < \infty$, (69), the Cauchy-Schwarz inequality, and the fact that for all $y \in W \setminus \{x\}$ it holds that $\|\omega_{x,y}\| = 1$ hence prove that

$$\begin{aligned}
&\left|\lambda(W)\Phi(x) - \int_W \Phi(y)\, dy\right| = \left|\int_W \int_0^{\|y-x\|} \frac{d}{dr}\Phi(x + r\omega_{x,y})\, dr\, dy\right| \\
&= \left|\int_W \int_0^{\|y-x\|} \langle (\nabla\Phi)(x + r\omega_{x,y}), \omega_{x,y}\rangle\, dr\, dy\right| \\
&\leq \int_W \int_0^{\|y-x\|} |\langle (\nabla\Phi)(x + r\omega_{x,y}), \omega_{x,y}\rangle|\, dr\, dy \\
&\leq \int_W \int_0^{\|y-x\|} \|(\nabla\Phi)(x + r\omega_{x,y})\|\, \|\omega_{x,y}\|\, dr\, dy \\
&= \int_W \int_0^{\|y-x\|} \|(\nabla\Phi)(x + r\omega_{x,y})\|\, dr\, dy.
\end{aligned} \tag{71}$$

The fact that $W \subseteq B_\rho$ and Fubini's theorem therefore show that

$$\left|\lambda(W)\Phi(x) - \int_W \Phi(y)\, dy\right| \leq \int_0^\infty \int_{B_\rho} \|E(x + r\omega_{x,y})\|\, dy\, dr. \tag{72}$$

Next observe that the integral transformation theorem with the diffeomorphism $\{v \in \mathbb{R}^d \colon \|v\| < \rho\} \ni y \mapsto y + x \in B_\rho$, items (i)–(iii) in Lemma 2.6, the fact that for all $r \in (0, \rho)$, $\phi \in S$ it holds that $\|T_\rho(r, \phi)\| = r$, and (68) imply that for all $r \in (0, \rho)$ it holds that

$$\begin{aligned}
&\int_{B_\rho} \|E(x + r\omega_{x,y})\|\, dy = \int_{\{v\in\mathbb{R}^d \colon \|v\|<\rho\}} \|E(x + r\omega_{0,y})\|\, dy \\
&= \int_0^\rho \int_S \left\|E\Big(x + r\tfrac{T_\rho(s,\phi)}{\|T_\rho(s,\phi)\|}\Big)\right\| \left|\det\big((T_\rho)'(s,\phi)\big)\right|\, d\phi\, ds \\
&= \int_0^\rho \int_S \|E(x + T_\infty(r,\phi))\| \left|\det\big((T_\infty)'(s,\phi)\big)\right|\, d\phi\, ds\,.
\end{aligned} \tag{73}$$

Items (i)–(ii) in Lemma 2.6, Fubini's theorem, and (72) therefore prove that

$$\begin{aligned}
&\left|\lambda(W)\Phi(x) - \int_W \Phi(y)\,dy\right| \\
&\leq \int_0^\infty \int_0^\rho \int_S \|E(x + T_\infty(r,\phi))\| \left|\det\big((T_\infty)'(s,\phi)\big)\right| d\phi\,ds\,dr \\
&= \int_0^\infty \int_S \int_0^\rho \|E(x + T_\infty(r,\phi))\| \left|\det\big((T_\infty)'(s,\phi)\big)\right| ds\,d\phi\,dr \\
&= \int_0^\infty \int_S \int_0^\rho \|E(x + T_\infty(r,\phi))\| \left|\det\big((T_\infty)'(1,\phi)\big)\right| s^{d-1}\,ds\,d\phi\,dr \\
&= \frac{\rho^d}{d} \int_0^\infty \int_S \|E(x + T_\infty(r,\phi))\| \left|\det\big((T_\infty)'(1,\phi)\big)\right| d\phi\,dr\,.
\end{aligned} \tag{74}$$

Combining this, (68), items (i)-(iii) in Lemma 2.6, the fact that for all $r \in (0,\infty)$, $\phi \in S$ it holds that $\|T_\infty(r,\phi)\| = r$, and (67) demonstrates that

$$\begin{aligned}
&\left|\lambda(W)\Phi(x) - \int_W \Phi(y)\,dy\right| \\
&\leq \frac{\rho^d}{d} \int_0^\infty \int_S \|E(x + T_\infty(r,\phi))\| \left|\det\big((T_\infty)'(1,\phi)\big)\right| r^{1-d} r^{d-1}\,d\phi\,dr \\
&= \frac{\rho^d}{d} \int_0^\infty \int_S \|E(x + T_\infty(r,\phi))\| \, \|T_\infty(r,\phi)\|^{1-d} \left|\det\big((T_\infty)'(r,\phi)\big)\right| \,d\phi\,dr \\
&= \frac{\rho^d}{d} \int_{\mathbb{R}^d} \|E(x+y)\| \, \|y\|^{1-d}\,dy \\
&= \frac{\rho^d}{d} \int_{\mathbb{R}^d} \|E(y)\| \, \|x-y\|^{1-d}\,dy = \frac{\rho^d}{d} \int_W \|(\nabla\Phi)(y)\| \, \|x-y\|^{1-d}\,dy.
\end{aligned} \tag{75}$$

The proof of Lemma 2.11 is thus completed. □

Lemma 2.12. *Let $d \in \{2,3,\ldots\}$, $p \in (d,\infty)$, let $\lambda\colon \mathcal{B}(\mathbb{R}^d) \to [0,\infty]$ be the Lebesgue–Borel measure on $\mathbb{R}^d$, let $W \subseteq \mathbb{R}^d$ be an open, bounded, and convex set with $\lambda(W) > 0$, let $\Phi \in C^1(W,\mathbb{R})$, let $\|\cdot\|\colon \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, and assume that $\int_W (|\Phi(x)|^p + \|(\nabla\Phi)(x)\|^p)\,dx < \infty$. Then*

$$\begin{aligned}
&\sup_{x\in W} \left|\Phi(x) - \frac{1}{\lambda(W)} \int_W \Phi(y)\,dy\right| \\
&\leq \frac{\big[\sup_{v,w\in W}\|v-w\|\big]^d}{\lambda(W)d} \left[\int_{\cup_{x\in W}\{x-y\colon y\in W\}} \|z\|^{(1-d)\frac{p}{p-1}}\,dz\right]^{(p-1)/p} \\
&\cdot \left[\int_W \|(\nabla\Phi)(y)\|^p\,dy\right]^{1/p} < \infty.
\end{aligned} \tag{76}$$

Proof of Lemma 2.12. Throughout this proof let $\rho \in [0,\infty)$ satisfy that $\rho = \sup_{v,w\in W}\|v-w\|$, let $W_x \subseteq \mathbb{R}^d$, $x \in W$, satisfy for all $x \in W$ that $W_x =$

$\{x - y \colon y \in W\}$, let $V \subseteq \mathbb{R}^d$ be the set given by $V = \cup_{x \in W} W_x$, let $\psi \colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$ that

$$\psi(x) = \begin{cases} \|x\|^{1-d} & : x \in V \backslash \{0\} \\ 0 & : x \in (\mathbb{R}^d \backslash V) \cup \{0\}, \end{cases} \tag{77}$$

and let $E \colon \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $x \in \mathbb{R}^d$ that

$$E(x) = \begin{cases} (\nabla\Phi)(x) & : x \in W \\ 0 & : x \in \mathbb{R}^d \backslash W. \end{cases} \tag{78}$$

Observe that the hypothesis that $\int_W (|\Phi(x)|^p + \|(\nabla\Phi)(x)\|^p)\, dx < \infty$ ensures that

$$\int_{\mathbb{R}^d} \|E(x)\|^p \, dx = \int_W \|(\nabla\Phi)(x)\|^p \, dx < \infty. \tag{79}$$

Moreover, the assumption that $\lambda(W) > 0$, Lemma 2.11, and the integral transformation theorem prove that for all $x \in W$ it holds that

$$\begin{aligned} &\left| \Phi(x) - \frac{1}{\lambda(W)} \int_W \Phi(y)\, dy \right| \leq \frac{\rho^d}{\lambda(W) d} \int_W \|(\nabla\Phi)(y)\| \, \|x - y\|^{1-d} \, dy \\ &= \frac{\rho^d}{\lambda(W) d} \int_{W_x} \|(\nabla\Phi)(x - y)\| \, \|y\|^{1-d} \, dy \\ &\leq \frac{\rho^d}{\lambda(W) d} \int_V \|E(x - y)\| \, \|y\|^{1-d} \, dy = \frac{\rho^d}{\lambda(W) d} \int_{\mathbb{R}^d} \|E(x - y)\| \, \psi(y)\, dy\,. \end{aligned} \tag{80}$$

Lemma 2.10, (79), and Hölder's inequality therefore demonstrate that for all $x \in W$ it holds that

$$\begin{aligned} &\left| \Phi(x) - \frac{1}{\lambda(W)} \int_W \Phi(y)\, dy \right| \\ &\leq \frac{\rho^d}{\lambda(W) d} \left[\int_{\mathbb{R}^d} \|E(x - y)\|^p \, dy \right]^{1/p} \left[\int_{\mathbb{R}^d} \psi(y)^{p/(p-1)} \, dy \right]^{(p-1)/p} \\ &= \frac{\rho^d}{\lambda(W) d} \left[\int_W \|(\nabla\Phi)(y)\|^p \, dy \right]^{1/p} \left[\int_V \|y\|^{(1-d)\frac{p}{p-1}} \, dy \right]^{(p-1)/p} < \infty\,. \end{aligned} \tag{81}$$

The proof of Lemma 2.12 is thus completed. □

Proposition 2.13. *Let $d \in \{2, 3, \ldots\}$, $p \in (d, \infty)$, let $W \subseteq \mathbb{R}^d$ be an open, bounded, and convex set, let $\Phi \in C^1(W, \mathbb{R})$, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the standard norm on $\mathbb{R}^d$, let $\lambda : \mathcal{B}(\mathbb{R}^d) \to [0, \infty]$ be the Lebesgue–Borel measure on $\mathbb{R}^d$, let I be a finite and non-empty set, let $W_i \subseteq \mathbb{R}^d$, $i \in I$, be open and convex sets, assume for all $i \in I$, $j \in I \backslash \{i\}$ that $\lambda(W_i) > 0$, $W_i \cap W_j = \emptyset$, and $\overline{W} = \cup_{i \in I} \overline{W_i}$,* *assume that* $\int_W (|\Phi(x)|^p + \|(\nabla\Phi)(x)\|^p)\,dx < \infty$*, and let* $(D_i)_{i\in I} \subseteq [0,\infty)$ *satisfy for all* $i \in I$ *that*

$$D_i = \frac{\left[\sup_{v,w\in W_i} \|v-w\|\right]^d}{\lambda(W_i)d} \left[\int_{\cup_{x\in W_i}\{x-y\colon y\in W_i\}} \|z\|^{(1-d)\frac{p}{p-1}}\,dz\right]^{(p-1)/p}. \tag{82}$$

Then

$$\begin{aligned}\sup_{x\in W} |\Phi(x)| \leq 2^{(p-1)/p} \max\left\{\max_{i\in I}\Big([\lambda(W_i)]^{-1/p}\Big), \max_{i\in I}(D_i)\right\}\\ \cdot \left[\int_W (|\Phi(x)|^p + \|(\nabla\Phi)(x)\|^p)\,dx\right]^{1/p}.\end{aligned} \tag{83}$$

Proof of Proposition 2.13. Note that the hypothesis that $W \subseteq \mathbb{R}^d$ is open and convex and [56, Theorem 6.3] imply that for all $i \in I$ it holds that $W_i \subseteq W$. Next observe that the hypothesis that $\int_W |\Phi(x)|^p + \|(\nabla\Phi)(x)\|^p\,dx < \infty$, the hypothesis that for all $i \in I$ it holds that W_i is bounded, and Hölder's inequality demonstrate that for all $i \in I$ it holds that

$$\begin{aligned}\frac{1}{\lambda(W_i)}\int_{W_i} |\Phi(y)|\,dy &= \int_{W_i} \left|\frac{1}{\lambda(W_i)}\Phi(y)\right| dy\\ &\leq \left[\int_{W_i} [\lambda(W_i)]^{-p/(p-1)}\,dy\right]^{(p-1)/p} \left[\int_{W_i} |\Phi(y)|^p\,dy\right]^{1/p}\\ &\leq [\lambda(W_i)]^{-1/p}\left[\int_{W_i} |\Phi(y)|^p\,dy\right]^{1/p} < \infty.\end{aligned} \tag{84}$$

The hypothesis that $\overline{W} = \cup_{i\in I}\overline{W_i}$, the triangle inequality, and Lemma 2.12 (with $W = W_i$ for $i \in I$ in the notation of Lemma 2.12) hence show that

$$\begin{aligned}&\sup_{x\in W}|\Phi(x)| = \max_{i\in I}\bigg(\sup_{x\in W_i}|\Phi(x)|\bigg)\\ &\leq \max_{i\in I}\left(\sup_{x\in W_i}\left|\Phi(x) - \frac{1}{\lambda(W_i)}\int_{W_i}\Phi(y)\,dy\right| + \left|\frac{1}{\lambda(W_i)}\int_{W_i}\Phi(y)\,dy\right|\right)\\ &\leq \max_{i\in I}\left(D_i\left[\int_{W_i}\|(\nabla\Phi)(y)\|^p\,dy\right]^{1/p} + [\lambda(W_i)]^{-1/p}\left[\int_{W_i}|\Phi(y)|^p\,dy\right]^{1/p}\right)\\ &\leq \max\left\{\max_{i\in I}\Big([\lambda(W_i)]^{-1/p}\Big), \max_{i\in I}(D_i)\right\}\\ &\cdot \max_{i\in I}\left(\left[\int_{W_i}|\Phi(y)|^p\,dy\right]^{1/p} + \left[\int_{W_i}\|(\nabla\Phi)(y)\|^p\,dy\right]^{1/p}\right).\end{aligned} \tag{85}$$

Next note that for all $(x_i)_{i\in I} \subseteq \mathbb{R}$ it holds that

$$\max_{i\in I}|x_i| \leq \left[\sum_{i\in I}|x_i|^p\right]^{1/p}. \tag{86}$$

Combining this, the hypothesis that for all $i \in I$, $j \in I\backslash\{i\}$ it holds that $W_i \cap W_j = \emptyset$ and $\overline{W} = \cup_{i\in I}\overline{W_i}$, and the fact that for all $a, b \in [0,\infty)$ it holds that $(a+b)^p \leq 2^{p-1}(a^p + b^p)$ with (85) demonstrates that

$$\begin{aligned}
\sup_{x\in W} |\Phi(x)| \leq \max&\left\{\max_{i\in I}\Big([\lambda(W_i)]^{-1/p}\Big), \max_{i\in I}(D_i)\right\} \\
&\cdot \left[\sum_{i\in I}\left|\left[\int_{W_i}|\Phi(y)|^p\,dy\right]^{1/p} + \left[\int_{W_i}\|(\nabla\Phi)(y)\|^p\,dy\right]^{1/p}\right|^p\right]^{1/p} \\
&\leq 2^{\frac{p-1}{p}}\max\left\{\max_{i\in I}\Big([\lambda(W_i)]^{-1/p}\Big), \max_{i\in I}(D_i)\right\} \\
&\cdot\left[\sum_{i\in I}\int_{W_i}(|\Phi(y)|^p + \|(\nabla\Phi)(y)\|^p)\,dy\right]^{1/p} \\
&= 2^{\frac{p-1}{p}}\max\left\{\max_{i\in I}\Big([\lambda(W_i)]^{-1/p}\Big), \max_{i\in I}(D_i)\right\} \\
&\cdot\left[\int_{W}(|\Phi(y)|^p + \|(\nabla\Phi)(y)\|^p)\,dy\right]^{1/p}.
\end{aligned} \tag{87}$$

The proof of Proposition 2.13 is thus completed. □

Corollary 2.14. *Let $d \in \{2,3,\ldots\}$, $\Phi \in C^1((0,1)^d,\mathbb{R})$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, and assume that $\int_{(0,1)^d}(|\Phi(x)|^{d^2} + \|(\nabla\Phi)(x)\|^{d^2})\,dx < \infty$. Then*

$$\sup_{x\in(0,1)^d}|\Phi(x)| \leq 8\sqrt{e}\left[\int_{(0,1)^d}\Big(|\Phi(x)|^{d^2} + \|(\nabla\Phi)(x)\|^{d^2}\Big)\,dx\right]^{1/d^2}. \tag{88}$$

Proof of Corollary 2.14. Throughout this proof let $\lambda : \mathcal{B}(\mathbb{R}^d) \to [0,\infty]$ be the Lebesgue–Borel measure on $\mathbb{R}^d$, let $m \in \mathbb{N}$ satisfy that

$$m = \min\left(\mathbb{N}\cap\left[\frac{d}{2\ln(2)}(\ln(2)+\ln(\pi)+1),\infty\right)\right), \tag{89}$$

let $B \subseteq \mathbb{R}^d$ be the set given by $B = \{x \in \mathbb{R}^d \colon \|x\| < 1\}$, let I be the set given by $I = \{1,\ldots,2^m\}^d$, let $W_i, V_i \subseteq \mathbb{R}^d$, $i \in I$, be the sets which satisfy for all $i = (i_1,\ldots,i_d) \in I$ that

$$W_i = (\times_{j=1}^d((i_j-1)2^{-m}, i_j 2^{-m})) \quad\text{and}\quad V_i = \cup_{x\in W_i}\{x - y \colon y \in W_i\}, \tag{90}$$

and let $(D_i)_{i\in I}, (\rho_i)_{i\in I} \subseteq (0,\infty)$ satisfy for all $i \in I$ that

$$\rho_i = \sup_{v,w\in W_i}\|v-w\| \quad\text{and}\quad D_i = \frac{\rho_i^d}{\lambda(W_i)d}\left[\int_{V_i}\|x\|^{(1-d)\frac{d^2}{d^2-1}}\,dx\right]^{\frac{d^2-1}{d^2}}. \tag{91}$$

Observe that (89) ensures that

$$[2\pi e]^{d/2} = e^{\frac{d}{2}[\ln(2)+\ln(\pi)+1]} = 2^{\frac{d}{2\ln(2)}[\ln(2)+\ln(\pi)+1]} \leq 2^m. \tag{92}$$

Hence, we obtain that

$$[2\pi e]^{d/2} 2^{-m} \leq 1. \tag{93}$$

Next note that (90) shows that for all $i \in I$ it holds that

$$\lambda(W_i) = 2^{-dm} \qquad \text{and} \qquad \rho_i = \sqrt{d}2^{-m}. \tag{94}$$

Moreover, note that (89) implies that

$$\frac{d}{2\ln(2)}(\ln(2) + \ln(\pi) + 1) \leq m \leq \frac{d}{2\ln(2)}(\ln(2) + \ln(\pi) + 1) + 1. \tag{95}$$

This, (94), and (92) hence show that

$$\max_{i\in I}\Big([\lambda(W_i)]^{-1/d^2}\Big) = 2^{m/d} \leq 2^{\frac{\ln(2)+\ln(\pi)+1}{2\ln(2)}+\frac{1}{d}} = 2^{\frac{1}{d}}\sqrt{2\pi e}. \tag{96}$$

Next note that Lemma 2.10 (with $p = d^2$, $W = W_i$ for $i \in I$ in the notation of Lemma 2.10) assures that for all $i \in I$ it holds that

$$\begin{aligned} \left[\int_{V_i} \|x\|^{(1-d)\frac{d^2}{d^2-1}}\,dx\right]^{\frac{d^2-1}{d^2}} &\leq \rho_i^{\frac{d-1}{d}} \left(\frac{d(d^2-1)}{d^2-d}\lambda(B)\right)^{\frac{d^2-1}{d^2}} \\ &= \rho_i^{\frac{d-1}{d}} \left((d+1)\lambda(B)\right)^{\frac{d^2-1}{d^2}}. \end{aligned} \tag{97}$$

Corollary 2.8, the fact that $(d+1)\frac{1}{\sqrt{d\pi}}(2\pi e)^{\frac{d}{2}} \geq 1$, the fact that $(d+1) \leq 2d$, the fact that $\frac{d^2-1}{d^2} \leq 1$, and (94) hence prove that for all $i \in I$ it holds that

$$\begin{aligned} &\left[\int_{V_i} \|x\|^{(1-d)\frac{d^2}{d^2-1}}\,dx\right]^{\frac{d^2-1}{d^2}} \\ &\leq (\sqrt{d}2^{-m})^{\frac{d-1}{d}} \left((d+1)\frac{1}{\sqrt{d\pi}}(2\pi e)^{\frac{d}{2}}\right)^{\frac{d^2-1}{d^2}} d^{-\frac{d}{2}\frac{d^2-1}{d^2}} \\ &\leq (\sqrt{d}2^{-m})^{\frac{d-1}{d}} 2d\frac{1}{\sqrt{d\pi}}(2\pi e)^{\frac{d}{2}} d^{-\frac{d}{2}\frac{d^2-1}{d^2}} \\ &= 2^{1+\frac{m}{d}-m} d^{1+\frac{d-1}{2d}} \frac{1}{\sqrt{d\pi}}(2\pi e)^{\frac{d}{2}} d^{-\frac{d}{2}\frac{d^2-1}{d^2}} \\ &= 2^{1+\frac{m}{d}-m} d^{1+\frac{1}{2}-\frac{1}{2d}+\frac{1}{2d}-\frac{d}{2}} \frac{1}{\sqrt{d\pi}}(2\pi e)^{\frac{d}{2}} \\ &= 2^{1+\frac{m}{d}-m} d^{1-\frac{d}{2}} \frac{1}{\sqrt{\pi}}(2\pi e)^{\frac{d}{2}} = \frac{2^{1+\frac{m}{d}-m} d^{1-\frac{d}{2}} (2\pi e)^{d/2}}{\sqrt{\pi}}. \end{aligned} \tag{98}$$

This, (95), (92), and (93) therefore demonstrate that for all $i \in I$ it holds that

$$\left[\int_{V_i} \|x\|^{(1-d)\frac{d^2}{d^2-1}}\,dx\right]^{\frac{d^2-1}{d^2}} \leq \frac{2^{1+\frac{m}{d}} d^{1-\frac{d}{2}}}{\sqrt{\pi}} \leq \frac{2^{1+\frac{\ln(2)+\ln(\pi)+1}{2\ln(2)}+\frac{1}{d}} d^{1-\frac{d}{2}}}{\sqrt{\pi}} = 2^{1+\frac{1}{d}}\sqrt{2e}d^{1-\frac{d}{2}}. \tag{99}$$

Next note that (94) ensures that for all $i \in I$ it holds that

$$\frac{\rho_i^d}{\lambda(W_i)d} = d^{\frac{d}{2}-1} 2^{-dm} 2^{dm} = d^{\frac{d}{2}-1}. \tag{100}$$

Combining this with (99) demonstrates that

$$\max_{i\in I}(D_i) = \max_{i\in I}\left(\frac{\rho_i^d}{\lambda(W_i)d}\left[\int_{V_i} \|x\|^{(1-d)\frac{d^2}{d^2-1}}\,dx\right]^{\frac{d^2-1}{d^2}}\right) \leq 2^{1+\frac{1}{d}}\sqrt{2e}. \tag{101}$$

Combining this and (96) with the hypothesis that $d \in \{2, 3, \ldots\}$ demonstrates that

$$\begin{aligned} &\max\left\{\max_{i\in I}\Big([\lambda(W_i)]^{-\frac{1}{d^2}}\Big), \max_{i\in I}(D_i)\right\} \\ &= \max\left\{2^{\frac{1}{d}}\sqrt{2\pi e}, 2^{1+\frac{1}{d}}\sqrt{2e}\right\} \\ &= 2^{1+\frac{1}{d}}\sqrt{2e} \leq 2\sqrt{2}\sqrt{2e} = 4\sqrt{e}. \end{aligned} \tag{102}$$

Next note that (90) ensures that for all $i \in I$, $j \in I\backslash\{i\}$ it holds that

$$W_i \cap W_j = \emptyset \quad \text{and} \quad [0,1]^d = \cup_{i\in I}\overline{W_i}. \tag{103}$$

This, (102), (94), Proposition 2.13 (with $p = d^2$, $I = I$, $W = (0,1)^d$, $W_i = W_i$ for $i \in I$ in the notation of Proposition 2.13), and the hypothesis that $d \in \{2, 3, \ldots\}$ hence imply that

$$\begin{aligned} \sup_{x\in(0,1)^d} |\Phi(x)| &\leq 2^{1-\frac{1}{d^2}} \max\left\{\max_{i\in I}\Big([\lambda(W_i)]^{-\frac{1}{d^2}}\Big), \max_{i\in I}(D_i)\right\} \\ &\quad \cdot \left[\int_{(0,1)^d} \Big(|\Phi(x)|^{d^2} + \|(\nabla\Phi)(x)\|^{d^2}\Big)\,dx\right]^{1/d^2} \\ &\leq 8\sqrt{e}\left[\int_{(0,1)^d} \Big(|\Phi(x)|^{d^2} + \|(\nabla\Phi)(x)\|^{d^2}\Big)\,dx\right]^{1/d^2}. \end{aligned} \tag{104}$$

The proof of Corollary 2.14 is thus completed. □

Corollary 2.15. *Let $d \in \mathbb{N}$, $\Phi \in C^1((0,1)^d, \mathbb{R})$, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the standard norm on $\mathbb{R}^d$, and assume that $\int_{(0,1)^d} (|\Phi(x)|^{\max\{2,d^2\}} + \|(\nabla\Phi)(x)\|^{\max\{2,d^2\}})\, dx < \infty$. Then*

$$\sup_{x\in(0,1)^d} |\Phi(x)| \leq 8\sqrt{e} \left[\int_{(0,1)^d} \left(|\Phi(x)|^{\max\{2,d^2\}} + \|(\nabla\Phi)(x)\|^{\max\{2,d^2\}} \right) dx \right]^{1/\max\{2,d^2\}}. \tag{105}$$

Proof of Corollary 2.15. To establish (105) we distinguish between the case $d = 1$ and the case $d \in \{2, 3, \ldots\}$. First, we consider the case $d = 1$. Note that Lemma 2.9 ensures that

$$\begin{aligned} \sup_{x\in(0,1)} |\Phi(x)| &\leq \int_0^1 (|\Phi(x)| + |\Phi'(x)|)\, dx \\ &\leq 8\sqrt{e} \left[\int_0^1 \left(|\Phi(x)|^2 + |\Phi'(x)|^2 \right) dx \right]^{1/2}. \end{aligned} \tag{106}$$

This establishes (105) in the case $d = 1$. Next we consider the case $d \in \{2, 3, \ldots\}$. Note that Corollary 2.14 shows that

$$\sup_{x\in(0,1)^d} |\Phi(x)| \leq 8\sqrt{e} \left[\int_{(0,1)^d} \left(|\Phi(x)|^{d^2} + \|(\nabla\Phi)(x)\|^{d^2} \right) dx \right]^{1/d^2}. \tag{107}$$

This establishes (105) in the case $d \in \{2, 3, \ldots\}$. The proof of Corollary 2.15 is thus completed. □

2.4 Sobolev type estimates for Monte Carlo approximations

In this subsection we provide in Lemma 2.16 below a Sobolev type estimate for Monte Carlo approximations. Lemma 2.16 is one of the main ingredients in our proof of Lemma 4.1 in Subsection 4.1 below.

Lemma 2.16. *Let $d, n \in \mathbb{N}$, $\zeta, a \in \mathbb{R}$, $b \in (a, \infty)$, $p \in [1, \infty)$, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the standard norm on $\mathbb{R}^d$, let $\mathfrak{K} \in (0, \infty)$ be the $(\max\{2, p\}, 2)$-Kahane–Khintchine constant (cf. Definition 2.1 and Lemma 2.2), let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\xi_i \colon \mathbb{R}^d \times \Omega \to \mathbb{R}$, $i \in \{1, \ldots, n\}$, be i.i.d. random fields satisfying for all $i \in \{1, \ldots, n\}$, $\omega \in \Omega$ that $\xi_i(\cdot, \omega) \in C^1(\mathbb{R}^d, \mathbb{R})$, let $\xi \colon \mathbb{R}^d \times \Omega \to \mathbb{R}$ be the random field satisfying for all $x \in \mathbb{R}^d$, $\omega \in \Omega$ that $\xi(x, \omega) = \xi_1(x, \omega)$, assume for all $\Phi \in C^1((0,1)^d, \mathbb{R})$ with $\int_{(0,1)^d} (|\Phi(x)|^{\max\{2,p\}} + \|(\nabla\Phi)(x)\|^{\max\{2,p\}})\, dx < \infty$ that*

$$\sup_{x\in(0,1)^d} |\Phi(x)| \leq \zeta \left[\int_{(0,1)^d} \left(|\Phi(x)|^{\max\{2,p\}} + \|(\nabla\Phi)(x)\|^{\max\{2,p\}} \right) dx \right]^{1/\max\{2,p\}}, \tag{108}$$

and assume that for all $x \in [a,b]^d$ it holds that

$$\inf_{\delta\in(0,\infty)} \sup_{v\in[-\delta,\delta]^d} \mathbb{E}\left[|\xi(x+v)|^{1+\delta} + \|(\nabla\xi)(x+v)\|^{1+\delta} \right] < \infty. \tag{109}$$

Then

(i) it holds that

$$\sup_{x\in[a,b]^d} \left| \mathbb{E}\big[\xi(x)\big] - \frac{1}{n}\left(\sum_{i=1}^n \xi_i(x)\right)\right|^p \tag{110}$$

is a random variable and

(ii) it holds that

$$\begin{aligned}\left(\mathbb{E}\left[\sup_{x\in[a,b]^d} \left| \mathbb{E}\big[\xi(x)\big] - \frac{1}{n}\left(\sum_{i=1}^n \xi_i(x)\right)\right|^p\right]\right)^{1/p} \\ \leq \frac{4\mathfrak{K}\zeta}{\sqrt{n}}\left(\sup_{x\in[a,b]^d}\left[\big(\mathbb{E}\big[|\xi(x)|^{\max\{2,p\}}\big]\big)^{1/\max\{2,p\}}\right.\right. \\ \left.\left. + (b-a)\left|\mathbb{E}\left[\,\|(\nabla\xi)(x)\|^{\max\{2,p\}}\right]\right|^{1/\max\{2,p\}}\right]\right).\end{aligned} \tag{111}$$

Proof of Lemma 2.16. 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, let $q\in[2,\infty)$ satisfy that $q=\max\{2,p\}$, let $\rho\colon\mathbb{R}^d\to\mathbb{R}^d$ satisfy for all $x=(x_1,\dots,x_d)\in\mathbb{R}^d$ that

$$\rho(x) = \big((b-a)x_1+a, (b-a)x_2+a, \dots, (b-a)x_d+a\big), \tag{112}$$

let $e_1,\dots,e_d\in\mathbb{R}^d$ satisfy that $e_1=(1,0,\dots,0),\dots,e_d=(0,\dots,0,1)$, let $Y\colon[0,1]^d\times\Omega\to\mathbb{R}$ be the random field which satisfies for all $x\in[0,1]^d$ that

$$Y(x) = \mathbb{E}\big[\xi(\rho(x))\big] - \frac{1}{n}\left(\sum_{i=1}^n \xi_i(\rho(x))\right), \tag{113}$$

let $Z\colon[0,1]^d\times\Omega\to\mathbb{R}^d$ be the random field which satisfies for all $x\in[0,1]^d$ that

$$Z(x) = (b-a)\left[\mathbb{E}\left[(\nabla\xi)(\rho(x))\right] - \frac{1}{n}\left(\sum_{i=1}^n (\nabla\xi_i)(\rho(x))\right)\right], \tag{114}$$

and let $E\colon\Omega\to[0,\infty)$ be the random variable given by

$$E = \sup_{x\in[a,b]^d\cap\mathbb{Q}^d}\left|\mathbb{E}\big[\xi(x)\big] - \frac{1}{n}\left(\sum_{i=1}^n \xi_i(x)\right)\right|. \tag{115}$$

Note that (112) ensures that $\rho([0,1]^d) = [a,b]^d$. Furthermore, note that (109) ensures that for all $x \in [a,b]^d$ there exists $\delta_x \in (0,\infty)$ such that

$$\sup_{v\in[-\delta_x,\delta_x]^d} \mathbb{E}\left[\|(\nabla\xi)(x+v)\|^{1+\delta_x}\right] < \infty. \tag{116}$$

Hölder's inequality therefore shows that for all $x \in [a,b]^d$ there exists $\delta_x \in (0,\infty)$ which satisfies that

$$\begin{aligned} &\sup_{v\in[-\delta_x,\delta_x]^d} \mathbb{E}\left[\|(\nabla\xi)(x+v)\|\right] \\ &\qquad \le \sup_{v\in[-\delta_x,\delta_x]^d} \left(\mathbb{E}\left[\|(\nabla\xi)(x+v)\|^{1+\delta_x}\right]\right)^{1/(1+\delta_x)} < \infty. \end{aligned} \tag{117}$$

Next note that the collection $S_x = \{y \in [a,b]^d : y - x \in (-\delta_x,\delta_x)^d\}$, $x \in [a,b]^d$, is an open cover of $[a,b]^d$. The fact that $[a,b]^d$ is compact hence ensures that there exists $N \in \mathbb{N}$ and $x_k \in [a,b]^d$, $k \in \{1,\ldots,N\}$ which satisfies that the collection S_{x_k}, $k \in \{1,\ldots,N\}$ is a finite open cover of $[a,b]^d$. Combining this with (117) demonstrates that

$$\begin{aligned} &\sup_{x\in[a,b]^d} \mathbb{E}\left[\|(\nabla\xi)(x)\|\right] \\ &\qquad \le \max_{k\in\{1,\ldots,N\}} \sup_{v\in(-\delta_{x_k},\delta_{x_k})^d} \mathbb{E}\left[\|(\nabla\xi)(x_k+v)\|\right] < \infty\,. \end{aligned} \tag{118}$$

Moreover, note that the fact that for all $\omega \in \Omega$ it holds that the functions $[a,b]^d \ni x \mapsto \xi(x,\omega) \in \mathbb{R}$ and $[a,b]^d \ni x \mapsto (\nabla\xi)(x,\omega) \in \mathbb{R}^d$ are continuous ensures that $[a,b]^d \times \Omega \ni (x,\omega) \mapsto \xi(x,\omega) \in \mathbb{R}$ and $[a,b]^d \times \Omega \ni (x,\omega) \mapsto (\nabla\xi)(x,\omega) \in \mathbb{R}^d$ are Carathéodory functions. This implies that $[a,b]^d \times \Omega \ni (x,\omega) \mapsto \xi(x,\omega) \in \mathbb{R}$ is $(\mathcal{B}([a,b]^d) \otimes \mathcal{F})/\mathcal{B}(\mathbb{R})$-measurable and $[a,b]^d \times \Omega \ni (x,\omega) \mapsto (\nabla\xi)(x,\omega) \in \mathbb{R}^d$ is $(\mathcal{B}([a,b]^d) \otimes \mathcal{F})/\mathcal{B}(\mathbb{R}^d)$-measurable, see, e.g., Aliprantis and Border [1, Lemma 4.51]). Next note that the fundamental theorem of calculus ensures that for all $x, y \in \mathbb{R}^d$ it holds that

$$\xi(x) - \xi(y) = \int_0^1 \langle (\nabla\xi)\left(y + t(x-y)\right), x - y\rangle \, dt. \tag{119}$$

Hence, we obtain that for all $x, y \in [a,b]^d$ it holds that

$$|\xi(x) - \xi(y)| \le \|x - y\| \int_0^1 \|(\nabla\xi)\left(y + t(x-y)\right)\| \, dt. \tag{120}$$

Combining this with Fubini's theorem shows that for all $x, y \in [a,b]^d$ it holds that

$$
\begin{aligned}
\big|\mathbb{E}[\xi(x)] - \mathbb{E}[\xi(y)]\big| &\leq \mathbb{E}\big[|\xi(x) - \xi(y)|\big] \\
&\leq \|x-y\|\,\mathbb{E}\left[\int_0^1 \|(\nabla\xi)\,(y+t(x-y))\|\;dt\right] \\
&= \|x-y\| \int_0^1 \mathbb{E}\left[\|(\nabla\xi)\,(y+t(x-y))\|\right] dt \\
&\leq \|x-y\| \sup_{v\in[a,b]^d} \mathbb{E}\left[\|(\nabla\xi)\,(v)\|\right].
\end{aligned}
\tag{121}
$$

This and (118) prove that $[a,b]^d \ni x \mapsto \mathbb{E}[\xi(x)] \in \mathbb{R}$ is a Lipschitz continuous function. Hence, we obtain for all $\omega \in \Omega$ that $[0,1]^d \ni x \mapsto Y(x,\omega) \in \mathbb{R}$ is a continuous function. Combining this with (115) implies that

$$
E = \sup_{x\in[0,1]^d} |Y(x)|\,. \tag{122}
$$

This establishes item (i). Next note that (112) implies that for all $j \in \{1,\dots,d\}$, $x \in [0,1]^d$, $h \in \mathbb{R}$ it holds that

$$
\rho(x+he_j) - \rho(x) = (b-a)he_j\,. \tag{123}
$$

This and (119) show that for all $j \in \{1,\dots,d\}$, $x \in [0,1]^d$, $h \in \mathbb{R}\setminus\{0\}$ it holds that

$$
\frac{\xi(\rho(x+he_j)) - \xi(\rho(x))}{h} = (b-a)\int_0^1 \big\langle(\nabla\xi)\big(\rho(x)+t(b-a)he_j\big), e_j\big\rangle\; dt\,. \tag{124}
$$

Moreover, note that (109) implies that for all $x \in [0,1]^d$ there exists $\delta_x \in (0,\infty)$ such that

$$
\sup_{v\in[-\delta_x,\delta_x]^d} \mathbb{E}\left[\|(\nabla\xi)(\rho(x)+v)\|^{1+\delta_x}\right] < \infty. \tag{125}
$$

This, Hölder's inequality, and Fubini's theorem show that for all $j \in \{1,\dots,d\}$, $x \in [0,1]^d$ there exists $\delta_x \in (0,\infty)$ such that for all $h \in \{h' \in \mathbb{R} : |(b-a)h'| < \delta_x\}$ it holds that

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\int_0^1 (\nabla\xi)\big(\rho(x)+t(b-a)he_j\big)\,dt\right\|^{1+\delta_x}\right] \\
&\leq \mathbb{E}\left[\int_0^1 \big\|(\nabla\xi)\big(\rho(x)+t(b-a)he_j\big)\big\|^{1+\delta_x}\;dt\right] \\
&= \int_0^1 \mathbb{E}\left[\big\|(\nabla\xi)\big(\rho(x)+t(b-a)he_j\big)\big\|^{1+\delta_x}\right] dt \\
&\leq \sup_{v\in[-\delta_x,\delta_x]^d} \mathbb{E}\left[\big\|(\nabla\xi)\big(\rho(x)+v\big)\big\|^{1+\delta_x}\right] < \infty\,.
\end{aligned}
\tag{126}
$$

This and (124) show that for all $j \in \{1, \dots, d\}$, $x \in [0,1]^d$ there exists $\delta_x \in (0, \infty)$ such that for all $h \in \{h' \in \mathbb{R} \setminus \{0\} : |(b-a)h'| < \delta_x\}$ it holds that

$$\begin{aligned} &\mathbb{E}\left[\left|\frac{\xi(\rho(x+he_j)) - \xi(\rho(x))}{h}\right|^{1+\delta_x}\right] \\ &\leq (b-a)^{1+\delta_x}\mathbb{E}\left[\left\|\int_0^1 (\nabla\xi)\big(\rho(x) + t(b-a)he_j\big)\,dt\right\|^{1+\delta_x}\right] \\ &\leq (b-a)^{1+\delta_x} \sup_{v \in [-\delta_x, \delta_x]^d} \mathbb{E}\left[\big\|(\nabla\xi)\big(\rho(x)+v\big)\big\|^{1+\delta_x}\right] < \infty. \end{aligned} \tag{127}$$

This, the theorem of de la Vallée–Poussin (see, e.g., [45, Theorem 6.19]), and Vitali's convergence theorem (see, e.g., [45, Theorem 6.25]) show that for all $x \in [0,1]^d$, $j \in \{1, \dots, d\}$ there exists $\delta_x \in (0, \infty)$ such that for all $(h_m)_{m \in \mathbb{N}} \subseteq \{h' \in \mathbb{R} \setminus \{0\} : |(b-a)h'| < \delta_x\}$ with $\lim_{m\to\infty} h_m = 0$ it holds that

$$\begin{aligned} &\lim_{m\to\infty} \mathbb{E}\left[\frac{\xi(\rho(x+h_m e_j)) - \xi(\rho(x))}{h_m}\right] \\ &= \mathbb{E}\left[\lim_{m\to\infty} \frac{\xi(\rho(x+h_m e_j)) - \xi(\rho(x))}{h_m}\right]. \end{aligned} \tag{128}$$

Therefore, we obtain that for all $x \in [0,1]^d$, $j \in \{1, \dots, d\}$ there exists $\delta_x \in (0, \infty)$ such that for all $(h_m)_{m \in \mathbb{N}} \subseteq \{h' \in \mathbb{R} \setminus \{0\} : |(b-a)h'| < \delta_x\}$ with $\lim_{m\to\infty} h_m = 0$ it holds that

$$\lim_{m\to\infty} \mathbb{E}\left[\frac{\xi(\rho(x+h_m e_j)) - \xi(\rho(x))}{h_m}\right] = (b-a)\mathbb{E}\Big[\langle(\nabla\xi)(\rho(x)), e_j\rangle\Big]. \tag{129}$$

Furthermore, the theorem of de la Valleé–Poussin, Vitali's convergence theorem, and (125) prove that for all $x \in [0,1]^d$, $j \in \{1, \dots, d\}$ it holds that

$$\limsup_{\mathbb{R}^d \setminus \{0\} \ni h \to 0} |\mathbb{E}[\langle(\nabla\xi)(\rho(x)+h), e_j\rangle] - \mathbb{E}[\langle(\nabla\xi)(\rho(x)), e_j\rangle]| = 0. \tag{130}$$

This and (129) imply that for all $\omega \in \Omega$, $x \in (0,1)^d$ it holds that $((0,1)^d \ni y \mapsto Y(y,\omega) \in \mathbb{R}) \in C^1((0,1)^d, \mathbb{R})$ and $(\nabla Y)(x,\omega) = Z(x,\omega)$. Combining this, (108), and (122) demonstrates that

$$E = \sup_{x \in (0,1)^d} |Y(x)| \leq \zeta \left[\int_{(0,1)^d} (|Y(x)|^q + \|Z(x)\|^q)\,dx\right]^{1/q}. \tag{131}$$

Next observe that (108) ensures that $\zeta \in [0, \infty)$. Hölder's inequality, (131), and

Fubini's theorem hence show that

$$\begin{aligned}\big(\mathbb{E}\big[|E|^p\big]\big)^{1/p} \le \big(\mathbb{E}\big[|E|^q\big]\big)^{1/q} &\le \zeta \left(\mathbb{E}\left[\int_{(0,1)^d} |Y(x)|^q + \|Z(x)\|^q \, dx\right]\right)^{1/q} \\ &= \zeta \left[\int_{(0,1)^d} \mathbb{E}[|Y(x)|^q + \|Z(x)\|^q] \, dx\right]^{1/q} \\ &\le \zeta \left[\sup_{x\in[0,1]^d} \mathbb{E}[|Y(x)|^q + \|Z(x)\|^q]\right]^{1/q}.\end{aligned} \tag{132}$$

Next note that (109) and Proposition 2.3 prove that for all $x \in [0,1]^d$ it holds that

$$\big(\mathbb{E}\big[|Y(x)|^q\big]\big)^{1/q} \le \frac{2\,\mathfrak{K}}{\sqrt{n}}\big(\mathbb{E}\big[|\xi(\rho(x)) - \mathbb{E}[\xi(\rho(x))]|^q\big]\big)^{1/q} \tag{133}$$

and

$$\big(\mathbb{E}\big[\|Z(x)\|^q\big]\big)^{1/q} \le \frac{2\,\mathfrak{K}\,(b-a)}{\sqrt{n}} \left(\mathbb{E}\left[\left\|(\nabla\xi)(\rho(x)) - \mathbb{E}\Big[(\nabla\xi)(\rho(x))\Big]\right\|^q\right]\right)^{1/q}. \tag{134}$$

This and (132) imply that

$$\begin{aligned}\big(\mathbb{E}\big[|E|^p\big]\big)^{1/p} \le &\frac{2\mathfrak{K}\zeta}{\sqrt{n}}\Bigg[\sup_{x\in[0,1]^d}\Big(\mathbb{E}\Big[\big|\xi(\rho(x)) - \mathbb{E}\big[\xi(\rho(x))\big]\big|^q \\ &+ (b-a)^q \,\big\|(\nabla\xi)(\rho(x)) - \mathbb{E}\big[(\nabla\xi)(\rho(x))\big]\big\|^q\Big]\Big)\Bigg]^{1/q}.\end{aligned} \tag{135}$$

Hence, we obtain that

$$\begin{aligned}\big(\mathbb{E}\big[|E|^p\big]\big)^{1/p} \le &\frac{2\mathfrak{K}\zeta}{\sqrt{n}}\left[\sup_{x\in[a,b]^d}\left(\mathbb{E}\Big[|\xi(x) - \mathbb{E}[\xi(x)]|^q \right.\right. \\ &\left.\left. + (b-a)^q \;\|(\nabla\xi)(x) - \mathbb{E}[(\nabla\xi)(x)]\|^q\Big]\right)\right]^{1/q}.\end{aligned} \tag{136}$$

The fact that for all $r, s \in [0,\infty)$ it holds that $(r+s)^{1/q} \le r^{1/q} + s^{1/q}$ and the triangle inequality therefore demonstrate that

$$\begin{aligned}\big(\mathbb{E}\big[|E|^p\big]\big)^{1/p} \leq& \frac{2\mathfrak{K}\zeta}{\sqrt{n}}\Bigg(\sup_{x\in[a,b]^d}\Bigg[\Big(\mathbb{E}\Big[|\xi(x)-\mathbb{E}[\xi(x)]|^q\Big]\Big)^{1/q} \\ &+ (b-a)\big(\mathbb{E}\big[\|(\nabla\xi)(x)-\mathbb{E}\big[(\nabla\xi)(x)\big]\|^q\big]\big)^{1/q}\Bigg]\Bigg) \\ \leq& \frac{4\mathfrak{K}\zeta}{\sqrt{n}}\Bigg(\sup_{x\in[a,b]^d}\Big[\big(\mathbb{E}\big[|\xi(x)|^q\big]\big)^{1/q} + (b-a)\big(\mathbb{E}\big[\|(\nabla\xi)(x)\|^q\big]\big)^{1/q}\Big]\Bigg).\end{aligned} \tag{137}$$

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

3 Stochastic differential equations with affine coefficient functions

3.1 A priori estimates for Brownian motions

In this subsection we provide in Lemma 3.1 below essentially well-known a priori estimates for standard Brownian motions. Lemma 3.1 will be employed in our proof of Corollary 3.5 in Subsection 3.2 below. Our proof of Lemma 3.1 is a slight adaption of the proof of Lemma 2.5 in Hutzenthaler et al. [39].

Lemma 3.1. *Let $d, m \in \mathbb{N}$, $T \in [0,\infty)$, $p \in (0,\infty)$, $A \in \mathbb{R}^{d\times m}$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be a standard Brownian motion. Then it holds for all $t \in [0,T]$ that*

$$\big(\mathbb{E}\big[\|AW_t\|^p\big]\big)^{1/p} \leq \sqrt{\max\{1, p-1\}\operatorname{Trace}(A^*A)\, t}. \tag{138}$$

Proof of Lemma 3.1. Throughout this proof for every $n \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^n} : \mathbb{R}^n \to [0,\infty)$ be the standard norm on $\mathbb{R}^n$, let $(q_r)_{r\in[0,\infty)} \subseteq \mathbb{N}_0$ satisfy for all $r \in [0,\infty)$ that $q_r = \max(\mathbb{N}_0 \cap [0, r/2])$, let $f_r\colon \mathbb{R}^m \to \mathbb{R}$, $r \in [0,\infty)$, satisfy for all $r \in [0,\infty)$, $x \in \mathbb{R}^m$ that

$$f_r(x) = \|Ax\|_{\mathbb{R}^d}^r, \tag{139}$$

and let $\beta^{(i)}\colon [0,T]\times\Omega \to \mathbb{R}$, $i \in \{1,\ldots,m\}$, be the stochastic processes which satisfy for all $i \in \{1,\ldots,m\}$, $t \in [0,T]$ that

$$W_t = \big(\beta^{(1)}_t, \ldots, \beta^{(m)}_t\big). \tag{140}$$

Note that for all $r \in [2,\infty)$, $x \in \mathbb{R}^m$ it holds that

$$(\nabla f_r)(x) = r\|Ax\|_{\mathbb{R}^d}^{r-2} A^*Ax. \tag{141}$$

This implies that for all $r \in [2,\infty)$, $x \in \mathbb{R}^m$ it holds that

$$(\operatorname{Hess} f_r)(x) = r\,\|Ax\|_{\mathbb{R}^d}^{(r-2)} A^*A + \mathbb{1}_{\{x\neq 0\}} r(r-2)\|Ax\|_{\mathbb{R}^d}^{(r-4)}\,(A^*Ax)\,(A^*Ax)^*\,. \tag{142}$$

The fact that for all $B \in \mathbb{R}^{m\times d}$, $x \in \mathbb{R}^d$ it holds that $\|Bx\|_{\mathbb{R}^m}^2 \leq \operatorname{Trace}(B^*B)\|x\|_{\mathbb{R}^d}^2$ and $\operatorname{Trace}(B^*B) = \operatorname{Trace}(BB^*)$ hence shows that for all $r \in [2,\infty)$, $x \in \mathbb{R}^m$ it holds that

$$\begin{aligned}
&\operatorname{Trace}((\operatorname{Hess} f_r)(x))\\
&= \operatorname{Trace}\Big(r\,\|Ax\|_{\mathbb{R}^d}^{(r-2)} A^*A + \mathbb{1}_{\{x\neq 0\}}\,r\,(r-2)\,\|Ax\|_{\mathbb{R}^d}^{(r-4)}\,(A^*Ax)\,(A^*Ax)^*\Big)\\
&= r\,\|Ax\|_{\mathbb{R}^d}^{(r-2)}\operatorname{Trace}(A^*A) + \mathbb{1}_{\{x\neq 0\}}\,r\,(r-2)\,\|Ax\|_{\mathbb{R}^d}^{(r-4)}\,\|A^*Ax\|_{\mathbb{R}^m}^2\\
&\leq r\,\|Ax\|_{\mathbb{R}^d}^{(r-2)}\operatorname{Trace}(A^*A) + \mathbb{1}_{\{x\neq 0\}} r\,(r-2)\,\|Ax\|_{\mathbb{R}^d}^{(r-4)}\operatorname{Trace}(AA^*)\,\|Ax\|_{\mathbb{R}^d}^2\\
&= r\,\|Ax\|_{\mathbb{R}^d}^{(r-2)}\operatorname{Trace}(A^*A) + r\,(r-2)\,\|Ax\|_{\mathbb{R}^d}^{(r-2)}\operatorname{Trace}(A^*A)\\
&= r\,(r-1)\operatorname{Trace}(A^*A)\,f_{r-2}(x)\,.
\end{aligned} \tag{143}$$

Moreover, note that the fact that $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ is a stochastic process with continuous sample paths ensures that $W\colon [0,T]\times\Omega\to\mathbb{R}^m$ is a $(\mathcal{B}([0,T])\otimes \mathcal{F})/\mathcal{B}(\mathbb{R}^m)$-measurable function. The fact for all $r \in [2,\infty)$ it holds that $f_r \in C^2(\mathbb{R}^m,\mathbb{R})$ hence implies that for all $r \in [2,\infty)$, $i \in \{1,\dots,m\}$ it holds that

$$[0,T]\times\Omega \ni (t,\omega) \mapsto (\tfrac{\partial}{\partial x_i} f_r)(W_t(\omega)) \in \mathbb{R} \tag{144}$$

is a $(\mathcal{B}([0,T])\otimes\mathcal{F})/\mathcal{B}(\mathbb{R})$-measurable function. Combining this and (141) demonstrates that for all $r \in [2,\infty)$, $i \in \{1,\dots,m\}$ it holds that

$$\begin{aligned}
\int_0^T \mathbb{E}\Big[\big|(\tfrac{\partial}{\partial x_i} f_r)(W_t)\big|^2\Big]\,dt &\leq \int_0^T \mathbb{E}\Big[\|(\nabla f_r)(W_t)\|_{\mathbb{R}^m}^2\Big]\,dt\\
&= \int_0^T \mathbb{E}\Big[r^2\,\|AW_t\|_{\mathbb{R}^d}^{2r-4}\,\|A^*AW_t\|_{\mathbb{R}^m}^2\Big]\,dt\\
&\leq \int_0^T \mathbb{E}\Big[r^2\,\|A^*\|_{L(\mathbb{R}^d,\mathbb{R}^m)}^2\,\|AW_t\|_{\mathbb{R}^d}^{2r-2}\Big]\,dt\\
&\leq \int_0^T \mathbb{E}\Big[r^2\operatorname{Trace}(A^*A)\,\|AW_t\|_{\mathbb{R}^d}^{2r-2}\Big]\,dt\\
&\leq r^2\operatorname{Trace}(A^*A)\,T\left(\sup_{t\in[0,T]}\mathbb{E}\Big[\|AW_t\|_{\mathbb{R}^d}^{2r-2}\Big]\right).
\end{aligned} \tag{145}$$

Next note that the fact that for all $r \in [2,\infty)$ it holds that $2r-2 \in [2,\infty)$ ensures that for all $r \in [2,\infty)$ it holds that

$$\sup_{t\in[0,T]}\mathbb{E}\Big[\|AW_t\|_{\mathbb{R}^d}^{2r-2}\Big] = \Big(\sup_{t\in[0,T]} t^{r-1}\Big)\mathbb{E}\Big[\|AW_1\|_{\mathbb{R}^d}^{2r-2}\Big] < \infty. \tag{146}$$

Combining this with (145) demonstrates that for all $r \in [2, \infty)$, $i \in \{1, \ldots, m\}$ it holds that

$$\int_0^T \mathbb{E}\Big[\big|(\tfrac{\partial}{\partial x_i} f_r)(W_t)\big|^2\Big] \, dt < \infty. \tag{147}$$

This proves that for all $r \in [2, \infty)$, $i \in \{1, \ldots, m\}$, $t \in [0, T]$ it holds that

$$\mathbb{E}\left[\int_0^t \big(\tfrac{\partial}{\partial x_i} f_r\big)(W_s) \, d\beta_s^{(i)}\right] = 0. \tag{148}$$

Itô's formula, Fubini's theorem, (139), and (143) hence show that for all $r \in [2, \infty)$, $t \in [0, T]$ it holds that

$$\begin{aligned}
&\mathbb{E}[f_r(W_t)] \\
&= \mathbb{E}\left[f_r(W_0) + \sum_{i=1}^{m} \left(\int_0^t \big(\tfrac{\partial}{\partial x_i} f_r\big)(W_s) \, d\beta_s^{(i)} + \frac{1}{2} \int_0^t \big(\tfrac{\partial^2}{\partial x_i^2} f_r\big)(W_s) \, ds\right)\right] \\
&= \frac{1}{2} \int_0^t \mathbb{E}\big[\operatorname{Trace}((\operatorname{Hess} f_r)(W_s)) \big] \, ds \\
&\leq \frac{r(r-1) \operatorname{Trace}(A^* A)}{2} \int_0^t \mathbb{E}\big[f_{r-2}(W_s)\big] \, ds.
\end{aligned} \tag{149}$$

This and (139) demonstrate that for all $t \in [0, T]$ it holds that

$$\begin{aligned}
\mathbb{E}\Big[\|A W_t\|_{\mathbb{R}^d}^2\Big] &\leq \frac{2(2-1)}{2} \operatorname{Trace}(A^* A) \int_0^t \mathbb{E}\big[f_0(W_s)\big] \, ds \\
&= \operatorname{Trace}(A^* A) \, t.
\end{aligned} \tag{150}$$

Hölder's inequality therefore proves that for all $r \in [0, 2)$, $t \in [0, T]$ it holds that

$$\mathbb{E}\big[f_r(W_t)\big] = \mathbb{E}\big[\|A W_t\|_{\mathbb{R}^d}^r\big] \leq \big(\mathbb{E}\big[\|A W_t\|_{\mathbb{R}^d}^2\big]\big)^{r/2} \leq (\operatorname{Trace}(A^* A) \, t)^{r/2}. \tag{151}$$

Hence, we obtain that for all $r \in (0, 2]$, $t \in [0, T]$ it holds that

$$\big(\mathbb{E}\big[\|A W_t\|_{\mathbb{R}^d}^r\big]\big)^{1/r} \leq \sqrt{\operatorname{Trace}(A^* A) \, t}. \tag{152}$$

Next note that (149), the fact that for all $r \in (2, \infty)$ it holds that $r - 2q_r \in [0, 2)$, and (151) imply that for all $r \in (2,\infty)$, $s_0 \in [0,T]$ it holds that

$$\begin{aligned}
&\mathbb{E}[\|AW_{s_0}\|_{\mathbb{R}^d}^r] \leq \frac{\left[\prod_{i=0}^{q_r-1}(r-2i)(r-1-2i)\right]}{2^{q_r}} \, [\mathrm{Trace}(A^*A)]^{q_r} \\
&\cdot \int_0^{s_0} \cdots \int_0^{s_{q_r-1}} \mathbb{E}\big[f_{r-2q_r}(W_{s_{q_r}})\big] \, ds_{q_r} \cdots ds_1 \\
&\leq \frac{\left[\prod_{i=0}^{q_r-1}(r-2i)(r-1-2i)\right]}{2^{q_r}} \, [\mathrm{Trace}(A^*A)]^{q_r+\frac{r-2q_r}{2}} \\
&\cdot \int_0^{s_0} \cdots \int_0^{s_{q_r-1}} (s_{q_r})^{\frac{r-2q_r}{2}} \, ds_{q_r} \cdots ds_1 \\
&= \frac{\left[\prod_{i=0}^{q_r-1}(r-2i)(r-1-2i)\right]}{2^{q_r}} \frac{2^{q_r}}{\left[\prod_{i=0}^{q_r-1}(r-2i)\right]} \, [\mathrm{Trace}(A^*A)]^{r/2} \, s_0^{r/2} \\
&= \left[\prod_{i=0}^{q_r-1}(r-1-2i)\right] [\mathrm{Trace}(A^*A)]^{r/2} \, s_0^{r/2}.
\end{aligned} \tag{153}$$

The fact that for all $r \in (2,\infty)$ it holds that $q_r \leq \frac{r}{2}$ hence demonstrates that for all $r \in (2,\infty)$, $t \in [0,T]$ it holds that

$$\begin{aligned}
\big(\mathbb{E}[\|AW_t\|_{\mathbb{R}^d}^r]\big)^{1/r} &\leq \left[\prod_{i=0}^{q_r-1}(r-1-2i)\right]^{1/r} \sqrt{\mathrm{Trace}(A^*A)\,t} \\
&\leq (r-1)^{\frac{q_r}{r}} \sqrt{\mathrm{Trace}(A^*A)\,t} \\
&\leq (r-1)^{\frac{r}{2r}} \sqrt{\mathrm{Trace}(A^*A)\,t} \\
&= \sqrt{(r-1)\,\mathrm{Trace}(A^*A)\,t}.
\end{aligned} \tag{154}$$

Combining this with (152) establishes (138). The proof of Lemma 3.1 is thus completed. □

3.2 A priori estimates for solutions

In this subsection we present in Lemma 3.4 and Corollary 3.5 below essentially well-known a priori estimates for solutions of stochastic differential equations with at most linearly growing drift coefficient functions and constant diffusion coefficient functions. Corollary 3.5 is one of the main ingredients in our proof of Lemma 4.1 in Subsection 4.1 below and is an immediate consequence of Lemma 3.1 above and Lemma 3.4 below. Our proof of Lemma 3.4 is a slight adaption of the proof of Lemma 2.6 in Beck et al. [5]. In our formulation of the statements of Lemma 3.4 and Corollary 3.5 below we employ the elementary result in Lemma 3.3 below. In our proof of Lemma 3.3 we employ the elementary result in Lemma 3.2 below. Lemma 3.2 and Lemma 3.3 study measurability properties for time-integrals of suitable stochastic processes.

Lemma 3.2. *Let $T \in [0,\infty)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $Y : [0,T]\times\Omega \to \mathbb{R}$ be $(\mathcal{B}([0,T])\otimes\mathcal{F})/\mathcal{B}(\mathbb{R})$-measurable, and assume that for all $\omega\in\Omega$ it holds that $\int_0^T |Y_t(\omega)|\,dt < \infty$. Then the function $\Omega \ni \omega \mapsto \int_0^T Y_t(\omega)\,dt \in \mathbb{R}$ is $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable.*

Proof. Throughout this proof let $Y^+ : [0,T]\times\Omega\to\mathbb{R}$ and $Y^- : [0,T]\times\Omega\to\mathbb{R}$ be the functions which satisfy for all $t\in[0,T]$, $\omega\in\Omega$ that $Y_t^+(\omega) = \max\{Y_t(\omega),0\}$ and $Y_t^-(\omega) = -\min\{Y_t(\omega),0\}$. Note that for all $\omega\in\Omega$ it holds that

$$\int_0^T Y_t(\omega)\,dt = \int_0^T Y_t^+(\omega)\,dt - \int_0^T Y_t^-(\omega)\,dt\,. \tag{155}$$

Moreover, observe that Tonelli's theorem implies that $\Omega\ni\omega\mapsto\int_0^T Y_t^+(\omega)\,dt \in \mathbb{R}$ is $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable and $\Omega\ni\omega\mapsto\int_0^T Y_t^-(\omega)\,dt\in\mathbb{R}$ is $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable. This and (155) prove that $\Omega\ni\omega\mapsto\int_0^T Y_t(\omega)\,dt\in\mathbb{R}$ is $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable. This completes the proof of Lemma 3.2. ☐

Lemma 3.3. *Let $d\in\mathbb{N}$, $c,C,T\in[0,\infty)$, let $\|\cdot\| : \mathbb{R}^d\to[0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $\mu\colon\mathbb{R}^d\to\mathbb{R}^d$ be a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable function which satisfies for all $x\in\mathbb{R}^d$ that $\|\mu(x)\|\le C + c\|x\|$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, and let $X\colon[0,T]\times\Omega\to\mathbb{R}^d$ be a stochastic process with continuous sample paths. Then for all $t\in[0,T]$ the function $\Omega\ni\omega\mapsto\int_0^t\mu(X_s(\omega))\,ds\in\mathbb{R}^d$ is $\mathcal{F}/\mathcal{B}(\mathbb{R}^d)$-measurable.*

Proof. The fact that $X\colon[0,T]\times\Omega\to\mathbb{R}^d$ is a stochastic process with continuous sample paths and Aliprantis and Border [1, Lemma 4.51] ensure that for all $t\in[0,T]$ the function $[0,t]\times\Omega\ni(s,\omega)\mapsto X_s(\omega)\in\mathbb{R}^d$ is $(\mathcal{B}([0,t])\otimes\mathcal{F})/\mathcal{B}(\mathbb{R}^d)$-measurable. This and the hypothesis that $\mu\colon\mathbb{R}^d\to\mathbb{R}^d$ is $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable imply that for all $i\in\{1,\dots,d\}$, $t\in[0,T]$ it holds that the function $[0,t]\times\Omega\ni(s,\omega)\mapsto\mu_i(X_s(\omega))\in\mathbb{R}$ is $(\mathcal{B}([0,t])\otimes\mathcal{F})/\mathcal{B}(\mathbb{R})$-measurable. Moreover, note that the hypothesis that for all $x\in\mathbb{R}^d$ it holds that $\|\mu(x)\|\le C+c\|x\|$ and the fact that for all $\omega\in\Omega$ the function $[0,T]\ni s\mapsto X_s(\omega)\in\mathbb{R}^d$ is continuous imply that for all $t\in[0,T]$, $\omega\in\Omega$, $i\in\{1,\dots,d\}$ it holds that $\int_0^t|\mu_i(X_s(\omega))|\,ds<\infty$. Lemma 3.2 (with $T=t$, $Y=\mu_i(X)$ for $t\in[0,T]$, $i\in\{1,\dots,d\}$ in the notation of Lemma 3.2) hence proves that for all $i\in\{1,\dots,d\}$, $t\in[0,T]$ the function $\Omega\ni\omega\mapsto\int_0^t\mu_i(X_s(\omega))\,ds\in\mathbb{R}$ is $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable. This implies that for all $t\in[0,T]$ the function $\Omega\ni\omega\mapsto\int_0^t\mu(X_s(\omega))\,ds\in\mathbb{R}^d$ is $\mathcal{F}/\mathcal{B}(\mathbb{R}^d)$-measurable. This completes the proof of Lemma 3.3. ☐

Lemma 3.4. *Let $d,m\in\mathbb{N}$, $x\in\mathbb{R}^d$, $p\in[1,\infty)$, $c,C,T\in[0,\infty)$, $A\in\mathbb{R}^{d\times m}$, let $\|\cdot\| : \mathbb{R}^d\to[0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $W\colon[0,T]\times\Omega\to\mathbb{R}^m$ be a standard Brownian motion, let $\mu\colon\mathbb{R}^d\to\mathbb{R}^d$ be a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable function which satisfies for all $y\in\mathbb{R}^d$ that $\|\mu(y)\|\le C+c\|y\|$, and let $X\colon[0,T]\times\Omega\to\mathbb{R}^d$ be a stochastic process with continuous sample paths which satisfies for all $t\in[0,T]$ that*

$$\mathbb{P}\bigg(X_t = x + \int_0^t \mu(X_s)\,ds + AW_t\bigg) = 1 \tag{156}$$

(cf. Lemma 3.3). Then

$$\big(\mathbb{E}[\|X_T\|^p]\big)^{1/p} \leq \Big(\|x\| + CT + \big(\mathbb{E}\big[\|AW_T\|^p\big]\big)^{1/p}\Big)\, e^{cT}. \tag{157}$$

Proof of Lemma 3.4. Throughout this proof for every $n \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^n} : \mathbb{R}^n \to [0,\infty)$ be the standard norm on $\mathbb{R}^n$, let $\beta^{(i)} \colon [0,T] \times \Omega \to \mathbb{R}$, $i \in \{1,\dots,m\}$, be the stochastic processes which satisfy for all $i \in \{1,\dots,m\}$, $t \in [0,T]$ that

$$W_t = \big(\beta_t^{(1)}, \dots, \beta_t^{(m)}\big) \tag{158}$$

and let $B \subseteq \Omega$ be the set given by

$$\begin{aligned} B &= \bigcap_{t\in[0,T]} \left\{X_t = x + \int_0^t \mu\big(X_s\big)\, ds + AW_t\right\} \\ &= \left\{\omega \in \Omega \colon \left(\forall\, t \in [0,T] \colon X_t(\omega) = x + \int_0^t \mu\big(X_s(\omega)\big)\, ds + AW_t(\omega)\right)\right\}. \end{aligned} \tag{159}$$

Observe that the fact that $X \colon [0,T] \times \Omega \to \mathbb{R}^d$ and $W \colon [0,T] \times \Omega \to \mathbb{R}^m$ are stochastic processes with continuous sample paths demonstrates that

$$B = \left(\bigcap_{t\in[0,T]\cap\mathbb{Q}} \left\{X_t = x + \int_0^t \mu\big(X_s\big)\, ds + AW_t\right\}\right) \in \mathcal{F}. \tag{160}$$

Combining this and (156) proves that

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}\left(\bigcap_{t\in[0,T]\cap\mathbb{Q}} \left\{X_t = x + \int_0^t \mu\big(X_s\big)\, ds + AW_t\right\}\right) \\ &= 1 - \mathbb{P}\left(\Omega \setminus \left[\bigcap_{t\in[0,T]\cap\mathbb{Q}} \left\{X_t = x + \int_0^t \mu\big(X_s\big)\, ds + AW_t\right\}\right]\right) \\ &= 1 - \mathbb{P}\left(\bigcup_{t\in[0,T]\cap\mathbb{Q}} \left\{X_t \neq x + \int_0^t \mu\big(X_s\big)\, ds + AW_t\right\}\right) \\ &\geq 1 - \left[\sum_{t\in[0,T]\cap\mathbb{Q}} \mathbb{P}\left(X_t \neq x + \int_0^t \mu\big(X_s\big)\, ds + AW_t\right)\right] = 1. \end{aligned} \tag{161}$$

Next note that the triangle inequality and the hypothesis that for all $y \in \mathbb{R}^d$ it holds that $\|\mu(y)\| \leq C + c\|y\|$ ensure that for all $\omega \in B$, $t \in [0,T]$ it holds that

$$\begin{aligned}\|X_t(\omega)\| &\leq \|x\| + \|AW_t(\omega)\| + \int_0^t \|\mu(X_s(\omega))\| \, ds \\ &\leq \|x\| + \|AW_t(\omega)\| + Ct + c\int_0^t \|X_s(\omega)\| \, ds \\ &\leq \|x\| + \left[\sup_{s\in[0,T]} \|AW_s(\omega)\|\right] + CT + c\int_0^t \|X_s(\omega)\| \, ds.\end{aligned} \tag{162}$$

Moreover, note that the assumption that $X\colon [0,T] \times \Omega \to \mathbb{R}^d$ is a stochastic process with continuous sample paths assures that for all $\omega \in \Omega$ it holds that

$$\int_0^T \|X_s(\omega)\| \, ds < \infty. \tag{163}$$

Grohs et al. [28, Lemma 2.11] (with $\alpha = \|x\| + \sup_{t\in[0,T]} \|AW_t(\omega)\| + CT$, $\beta = c$, $f = ([0,T] \ni t \mapsto \|X_t(\omega)\| \in [0,\infty))$ for $\omega \in B$ in the notation of Grohs et al. [28, Lemma 2.11]) and (162) hence prove that for all $\omega \in B$, $t \in [0,T]$ it holds that

$$\|X_t(\omega)\| \leq \left(\|x\| + \left[\sup_{s\in[0,T]} \|AW_s(\omega)\|\right] + CT\right) e^{ct}. \tag{164}$$

Next note that

$$\left(\mathbb{E}\left[\sup_{s\in[0,T]} \|AW_s\|^p\right]\right)^{1/p} \leq \|A\|_{L(\mathbb{R}^m,\mathbb{R}^d)} \left(\mathbb{E}\left[\sup_{s\in[0,T]} \|W_s\|^p_{\mathbb{R}^m}\right]\right)^{1/p}. \tag{165}$$

Moreover, if $p \in [1,2]$, then it holds that

$$\begin{aligned}\sup_{s\in[0,T]} \|W_s\|^p_{\mathbb{R}^m} &= \sup_{s\in[0,T]} \left(\sum_{i=1}^m |\beta_s^{(i)}|^2\right)^{p/2} \leq \sup_{s\in[0,T]} \left(1 + \sum_{i=1}^m |\beta_s^{(i)}|^2\right)^{p/2} \\ &\leq 1 + \sum_{i=1}^m \left[\sup_{s\in[0,T]} |\beta_s^{(i)}|^2\right].\end{aligned} \tag{166}$$

Hence, if $p \in [1,2]$, then the Burkholder–Davis–Gundy inequality (see, e.g., Karatzas and Shreve [44, Theorem 3.28]) implies that

$$\mathbb{E}\left[\sup_{s\in[0,T]} \|W_s\|^p_{\mathbb{R}^m}\right] \leq 1 + m\,\mathbb{E}\left[\sup_{s\in[0,T]} |\beta_s^{(1)}|^2\right] < \infty\,. \tag{167}$$

Next note that if $p \in (2,\infty)$, then Hölder's inequality implies that

$$\begin{aligned}\sup_{s\in[0,T]} \|W_s\|^p_{\mathbb{R}^m} &= \sup_{s\in[0,T]} \left(\sum_{i=1}^m |\beta_s^{(i)}|^2\right)^{p/2} \leq m^{p/2-1} \sup_{s\in[0,T]} \left(\sum_{i=1}^m |\beta_s^{(i)}|^p\right) \\ &\leq m^{p/2-1} \sum_{i=1}^m \left[\sup_{s\in[0,T]} |\beta_s^{(i)}|^p\right].\end{aligned} \tag{168}$$

Hence, if $p \in (2,\infty)$, then the Burkholder–Davis–Gundy inequality (see, e.g., Karatzas and Shreve [44, Theorem 3.28]) implies that

$$\mathbb{E}\left[\sup_{s\in[0,T]} \|W_s\|_{\mathbb{R}^m}^p\right] \leq m^{p/2-1}\, m\mathbb{E}\left[\sup_{s\in[0,T]} |\beta_s^{(1)}|^p\right] < \infty\,. \tag{169}$$

This, (165), and (167) hence imply that

$$\left(\mathbb{E}\left[\sup_{s\in[0,T]} \|AW_s\|^p\right]\right)^{1/p} < \infty. \tag{170}$$

Combining this, (164), and (161) demonstrates that

$$\begin{aligned}
&\int_0^T \big(\mathbb{E}\big[\|X_t\|^p\big]\big)^{1/p}\, dt \\
&\leq T\left[\sup_{t\in[0,T]} \big(\mathbb{E}\big[\|X_t\|^p\big]\big)^{1/p}\right] \\
&\leq T\left|\mathbb{E}\Big[\big|\|x\| + \sup\nolimits_{s\in[0,T]} \|AW_s\| + CT\big|^p e^{pcT}\Big]\right|^{1/p} \\
&\leq T\left[\|x\| + \left|\mathbb{E}\left[\sup_{s\in[0,T]} \|AW_s\|^p\right]\right|^{1/p} + CT\right] e^{cT} < \infty.
\end{aligned} \tag{171}$$

Next observe that the hypothesis that $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ is a stochastic process with continuous sample paths ensures that $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ is a $(\mathcal{B}([0,T])\otimes \mathcal{F})/\mathcal{B}(\mathbb{R}^d)$-measurable function. Hence, we obtain that for all $t \in [0,T]$ it holds that

$$\left(\mathbb{E}\left[\left|\int_0^t \|X_s\|\, ds\right|^p\right]\right)^{1/p} \leq \int_0^t (\mathbb{E}[\|X_s\|^p])^{1/p}\, ds \tag{172}$$

(cf., for example, Garling [24, Corollary 5.4.2] or Jentzen & Kloeden [42, Proposition 8 in Appendix A]). The triangle inequality, the fact that for all $t \in [0,T]$ it holds that W_t has the same distribution as $\frac{\sqrt{t}}{\sqrt{T}}W_T$, (161), and (162) therefore show that for all $t \in [0,T]$ it holds that

$$\begin{aligned}
&\big(\mathbb{E}\big[\|X_t\|^p\big]\big)^{1/p} \\
&\leq \|x\| + \big(\mathbb{E}\big[\|AW_t\|^p\big]\big)^{1/p} + Ct + c\int_0^t \big(\mathbb{E}\big[\|X_s\|^p\big]\big)^{1/p}\, ds \\
&= \|x\| + \tfrac{\sqrt{t}}{\sqrt{T}}\big(\mathbb{E}\big[\|AW_T\|^p\big]\big)^{1/p} + Ct + c\int_0^t \big(\mathbb{E}\big[\|X_s\|^p\big]\big)^{1/p}\, ds \\
&\leq \|x\| + \big(\mathbb{E}\big[\|AW_T\|^p\big]\big)^{1/p} + CT + c\int_0^t \big(\mathbb{E}\big[\|X_s\|^p\big]\big)^{1/p}\, ds.
\end{aligned} \tag{173}$$

Combining Grohs et al. [28, Lemma 2.11] (with $\alpha = \|x\| + (\mathbb{E}[\|AW_T\|^p])^{1/p} + CT$, $\beta = c$, $f = ([0,T] \ni t \mapsto (\mathbb{E}[\|X_t\|^p])^{1/p} \in [0,\infty))$ in the notation of Grohs et al. [28, Lemma 2.11]) and (171) hence establishes that for all $t \in [0,T]$ it holds that

$$\big(\mathbb{E}\big[\|X_t\|^p\big]\big)^{1/p} \leq \Big(\|x\| + \big(\mathbb{E}\big[\|AW_T\|^p\big]\big)^{1/p} + CT\Big)\, e^{ct}. \tag{174}$$

The proof of Lemma 3.4 is thus completed. □

Corollary 3.5. *Let $d, m \in \mathbb{N}$, $x \in \mathbb{R}^d$, $p \in [1,\infty)$, $c, C, T \in [0,\infty)$, $A \in \mathbb{R}^{d\times m}$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be a standard Brownian motion, let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ be a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable function which satisfies for all $y \in \mathbb{R}^d$ that $\|\mu(y)\| \leq C + c\|y\|$, and let $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a stochastic process with continuous sample paths which satisfies for all $t \in [0,T]$ that*

$$\mathbb{P}\bigg(X_t = x + \int_0^t \mu(X_s)\, ds + AW_t\bigg) = 1 \tag{175}$$

(cf. Lemma 3.3). Then

$$\big(\mathbb{E}\big[\|X_T\|^p\big]\big)^{1/p} \leq \Big(\|x\| + CT + \sqrt{\max\{1, p-1\}\operatorname{Trace}(A^*A)\, T}\Big)\, e^{cT}. \tag{176}$$

Proof of Corollary 3.5. Observe that Lemma 3.1 and Lemma 3.4 establish (176). The proof of Corollary 3.5 is thus completed. □

3.3 A priori estimates for differences of solutions

In this subsection we provide in Lemma 3.6 below well-known a priori estimates for differences of solutions of stochastic differential equations with Lipschitz continuous drift coefficient functions and constant diffusion coefficient functions. Lemma 3.6 is one of the main ingredients in our proof of Lemma 4.1 in Subsection 4.1 below. Our proof of Lemma 3.6 is a slight adaption of the proof of Lemma 2.6 in Beck et al. [5].

Lemma 3.6. *Let $d, m \in \mathbb{N}$, $p \in [1,\infty)$, $l \in [0,\infty)$, $T \in [0,\infty)$, $A \in \mathbb{R}^{d\times m}$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be a standard Brownian motion, let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ be a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable function which satisfies for all $x, y \in \mathbb{R}^d$ that $\|\mu(x) - \mu(y)\| \leq l\|x - y\|$, and let $X^x\colon [0,T]\times\Omega \to \mathbb{R}^d$, $x \in \mathbb{R}^d$, be stochastic processes with continuous sample paths which satisfy for all $x \in \mathbb{R}^d$, $t \in [0,T]$ that*

$$\mathbb{P}\bigg(X_t^x = x + \int_0^t \mu(X_s^x)\, ds + AW_t\bigg) = 1. \tag{177}$$

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

$$(\mathbb{E}[\|X_t^x - X_t^y\|^p])^{1/p} \leq e^{lt}\|x - y\|. \tag{178}$$

Proof of Lemma 3.6. Throughout this proof let $B_x \subseteq \Omega$, $x \in \mathbb{R}^d$, be the sets which satisfy for all $x \in \mathbb{R}^d$ that

$$\begin{aligned} B_x &= \bigcap_{t\in[0,T]} \left\{ X_t^x = x + \int_0^t \mu\big(X_s^x\big)\, ds + AW_t \right\} \\ &= \left\{ \omega \in \Omega \colon \left(\forall\, t \in [0,T] \colon X_t^x(\omega) = x + \int_0^t \mu\big(X_s^x(\omega)\big)\, ds + AW_t(\omega) \right) \right\}. \end{aligned} \tag{179}$$

Observe that the fact that for all $x \in \mathbb{R}^d$ it holds that $X^x \colon [0,T] \times \Omega \to \mathbb{R}^d$ and $W \colon [0,T] \times \Omega \to \mathbb{R}^m$ are stochastic processes with continuous sample paths demonstrates that for all $x \in \mathbb{R}^d$ it holds that

$$B_x = \left(\bigcap_{t\in[0,T]\cap\mathbb{Q}} \left\{ X_t^x = x + \int_0^t \mu\big(X_s^x\big)\, ds + AW_t \right\} \right) \in \mathcal{F}. \tag{180}$$

Combining this and (177) proves that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \mathbb{P}(B_x) &= \mathbb{P}\left(\bigcap_{t\in[0,T]\cap\mathbb{Q}} \left\{ X_t^x = x + \int_0^t \mu\big(X_s^x\big)\, ds + AW_t \right\} \right) \\ &= 1 - \mathbb{P}\left(\Omega \setminus \left[\bigcap_{t\in[0,T]\cap\mathbb{Q}} \left\{ X_t^x = x + \int_0^t \mu\big(X_s^x\big)\, ds + AW_t \right\} \right] \right) \\ &= 1 - \mathbb{P}\left(\bigcup_{t\in[0,T]\cap\mathbb{Q}} \left\{ X_t^x \neq x + \int_0^t \mu\big(X_s^x\big)\, ds + AW_t \right\} \right) \\ &\geq 1 - \left[\sum_{t\in[0,T]\cap\mathbb{Q}} \mathbb{P}\left(X_t^x \neq x + \int_0^t \mu\big(X_s^x\big)\, ds + AW_t \right) \right] = 1. \end{aligned} \tag{181}$$

Hence, we obtain that for all $x, y \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\mathbb{P}(B_x \cap B_y) = 1 - \mathbb{P}(\Omega \backslash [B_x \cap B_y]) \\ &= 1 - \mathbb{P}\big(B_x^c \cup B_y^c\big) \geq 1 - \big[\mathbb{P}(B_x^c) + P\big(B_y^c\big)\big] = 1. \end{aligned} \tag{182}$$

Next note that the triangle inequality and the assumption that for all $x, y \in \mathbb{R}^d$ it holds that $\|\mu(x) - \mu(y)\| \leq l\|x - y\|$ ensure that for all $x, y \in \mathbb{R}^d$, $\omega \in B_x \cap B_y$, $t \in [0,T]$ it holds that

$$\begin{aligned} \|X_t^x(\omega) - X_t^y(\omega)\| &\leq \|x - y\| + \int_0^t \|\mu(X_s^x(\omega)) - \mu(X_s^y(\omega))\|\, ds \\ &\leq \|x - y\| + l \int_0^t \|X_s^x(\omega) - X_s^y(\omega)\|\, ds. \end{aligned} \tag{183}$$

Moreover, note that the assumption that for all $x \in \mathbb{R}^d$ it holds that $X^x \colon [0,T] \times \Omega \to \mathbb{R}^d$ is a stochastic process with continuous sample paths assures that for all $x, y \in \mathbb{R}^d$, $\omega \in \Omega$ it holds that

$$\int_0^T \|X_s^x(\omega) - X_s^y(\omega)\| \, ds < \infty. \tag{184}$$

Grohs et al. [28, Lemma 2.11] (with $\alpha = \|x - y\|$, $\beta = l$, $f = ([0,T] \ni t \mapsto \|X_t^x(\omega) - X_t^y(\omega)\| \in [0,\infty))$ for $x, y \in \mathbb{R}^d$, $\omega \in B_x \cap B_y$ in the notation of Grohs et al. [28, Lemma 2.11]) and (183) hence prove that for all $x, y \in \mathbb{R}^d$, $\omega \in B_x \cap B_y$, $t \in [0,T]$ it holds that

$$\|X_t^x(\omega) - X_t^y(\omega)\| \leq \|x - y\| e^{lt}. \tag{185}$$

This and (182) prove that for all $x, y \in \mathbb{R}^d$, $t \in [0,T]$ it holds that

$$\left(\mathbb{E}[\|X_t^x - X_t^y\|^p]\right)^{1/p} \leq e^{lt} \|x - y\|. \tag{186}$$

The proof of Lemma 3.6 is thus completed. □

4 Error estimates

4.1 Quantitative error estimates

In this subsection we establish in Corollary 4.3 below a quantitative approximation result for viscosity solutions (cf., for example, Hairer et al. [31]) of Kolmogorov PDEs with constant coefficient functions. Our proof of Corollary 4.3 employs the quantitative approximation results in Lemma 4.1 and Corollary 4.2 below. Corollary 4.3 is one of the key ingredients which we use in our proof of Proposition 4.4 and Proposition 4.6 below, respectively, in order to construct ANN approximations for viscosity solutions of Kolmogorov PDEs with constant coefficient functions.

Lemma 4.1. *Let $d, n \in \mathbb{N}$, $\varphi \in C(\mathbb{R}^d, \mathbb{R})$, $c, l, a \in \mathbb{R}$, $b \in (a, \infty)$, $\zeta, \varepsilon, T \in (0,\infty)$, $v, \mathbf{v}, w, \mathbf{w}, z, \mathbf{z} \in [0,\infty)$, $p \in [1,\infty)$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $\langle \cdot, \cdot \rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, let $\mathfrak{K} \in (0,\infty)$ be the $(\max\{2,p\}, 2)$-Kahane–Khintchine constant (cf. Definition 2.1 and Lemma 2.2), assume for all $\Phi \in C^1((0,1)^d, \mathbb{R})$ that*

$$\sup_{x \in (0,1)^d} |\Phi(x)| \leq \zeta \left[\int_{(0,1)^d} \left(|\Phi(x)|^{\max\{2,p\}} + \|(\nabla \Phi)(x)\|^{\max\{2,p\}} \right) dx \right]^{1/\max\{2,p\}}, \tag{187}$$

let $\phi \in C^1(\mathbb{R}^d, \mathbb{R})$ satisfy for all $x \in \mathbb{R}^d$ that

$$|\phi(x)| \leq c d^z (1 + \|x\|^{\mathbf{z}}), \qquad \|(\nabla \phi)(x)\| \leq c d^w (1 + \|x\|^{\mathbf{w}}), \tag{188}$$

$$\textit{and} \qquad |\varphi(x) - \phi(x)| \leq \varepsilon d^v (1 + \|x\|^{\mathbf{v}}), \tag{189}$$

let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ *satisfy for all* $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ *that* $\mu(\lambda x + y) + \lambda\mu(0) = \lambda\mu(x) + \mu(y)$ *and*

$$\|\mu(x) - \mu(y)\| \leq l\|x - y\|, \tag{190}$$

let $A = (A_{i,j})_{(i,j)\in\{1,\ldots,d\}^2} \in \mathbb{R}^{d\times d}$ *be a symmetric and positive semi-definite matrix, let* $u \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$*, assume for all* $x \in \mathbb{R}^d$ *that* $u(0,x) = \varphi(x)$*, assume that* $\inf_{\gamma\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \big(\frac{|u(t,x)|}{1+\|x\|^\gamma}\big) < \infty$*, and assume that* $u|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u)(t,x) = \langle \mu(x), (\nabla_x u)(t,x)\rangle + \tfrac{1}{2}\textstyle\sum_{i,j=1}^d A_{i,j}\,(\tfrac{\partial^2}{\partial x_i \partial x_j}u)(t,x) \tag{191}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$*. Then there exist* $W_1, \ldots, W_n \in \mathbb{R}^{d\times d}$, $B_1, \ldots, B_n \in \mathbb{R}^d$ *such that*

$$\begin{aligned}
&\sup_{x\in[a,b]^d}\left|u(T,x) - \left[\frac{1}{n}\sum_{k=1}^n \phi(W_k x + B_k)\right]\right| \\
&\leq \varepsilon d^v\Big(1 + e^{\mathbf{v}lT}\big[T\|\mu(0)\| + \sqrt{\max\{1, \mathbf{v}-1\}\,T\,\mathrm{Trace}(A)} \\
&+ \sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{v}}\Big) \\
&+ \frac{4\mathfrak{K}\zeta}{\sqrt{n}}\bigg(cd^z\Big(1 + e^{\mathbf{z}lT}\big[T\|\mu(0)\| + \sqrt{\max\{1, \mathbf{z}\max\{2,p\} - 1\}\,T\,\mathrm{Trace}(A)} \\
&+ \sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{z}}\Big) + (b-a)cd^{w+1/2}e^{lT}\Big(1 + e^{\mathbf{w}lT}\big[T\|\mu(0)\| \\
&+ \sqrt{\max\{1, \mathbf{w}\max\{2,p\} + \mathbf{w} - 1\}\,T\,\mathrm{Trace}(A)} + \sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{w}}\Big)\bigg).
\end{aligned} \tag{192}$$

Proof of Lemma 4.1. Throughout this proof let $e_1, \ldots, e_d \in \mathbb{R}^d$ satisfy that $e_1 = (1, 0, \ldots, 0), \ldots, e_d = (0, \ldots, 0, 1)$, let $m_r\colon (0,\infty) \to [r,\infty)$, $r \in \mathbb{R}$, satisfy for all $r \in \mathbb{R}$, $x \in (0,\infty)$ that $m_r(x) = \max\{r, x\}$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a normal filtration $(\mathbb{F}_t)_{t\in[0,T]}$, 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, and 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

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

$$X_t^{i,x} = x + \int_0^t \mu(X_s^{i,x})\,ds + \sqrt{A}W_t^i \tag{193}$$

and

(b) for all $i \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$, $t \in [0,T]$, $\omega \in \Omega$ it holds that $X_t^{i,\lambda x+y}(\omega) + \lambda X_t^{i,0}(\omega) = \lambda X_t^{i,x}(\omega) + X_t^{i,y}(\omega)$

(cf. Grohs et al. [28, Proposition 2.20]). Note that item (b) and Grohs et al. [28, Corollary 2.8] (with $d = d$, $m = d$, $\varphi = (\mathbb{R}^d \ni x \mapsto X_T^{i,x}(\omega) \in \mathbb{R}^d)$ for $i \in \mathbb{N}$, $\omega \in \Omega$ in the notation of Grohs et al. [28, Corollary 2.8]) ensure that for all $i \in \mathbb{N}$, $\omega \in \Omega$ there exist $\mathcal{W}_{i,\omega} \in \mathbb{R}^{d\times d}$ and $\mathcal{B}_{i,\omega} \in \mathbb{R}^d$ which satisfy that for all $x \in \mathbb{R}^d$ it holds that

$$X_T^{i,x}(\omega) = \mathcal{W}_{i,\omega} x + \mathcal{B}_{i,\omega}. \tag{194}$$

Combining this with the assumption that $\phi \in C^1(\mathbb{R}^d, \mathbb{R})$ proves that for all $\omega \in \Omega$, $i \in \mathbb{N}$ it holds that

$$\left(\mathbb{R}^d \ni x \mapsto \phi\left(X_T^{i,x}(\omega)\right) \in \mathbb{R}\right) \in C^1(\mathbb{R}^d, \mathbb{R}). \tag{195}$$

Next note that (188) and (189) ensure that $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ is an at most polynomially growing function. This, (190), and the Feynman-Kac formula (cf., for example, Grohs et al. [28, Proposition 2.22] or Hairer et al. [31, Corollary 4.17]) imply that for all $x \in \mathbb{R}^d$ it holds that

$$u(T, x) = \mathbb{E}\big[\varphi(X_T^{1,x})\big]. \tag{196}$$

This shows that

$$\begin{aligned}
&\sup_{x\in[a,b]^d} \left| u(T,x) - \mathbb{E}\big[\phi(X_T^{1,x})\big] \right| \\
&= \sup_{x\in[a,b]^d} \left| \mathbb{E}\big[\varphi(X_T^{1,x}) - \phi(X_T^{1,x})\big] \right| \\
&\leq \sup_{x\in[a,b]^d} \mathbb{E}\left[\left|\varphi(X_T^{1,x}) - \phi(X_T^{1,x})\right| \frac{d^v\big(1 + \|X_T^{1,x}\|^{\mathbf{v}}\big)}{d^v\big(1 + \|X_T^{1,x}\|^{\mathbf{v}}\big)} \right] \\
&\leq d^v \left[\sup_{y\in\mathbb{R}^d} \left(\frac{|\varphi(y) - \phi(y)|}{d^v(1 + \|y\|^{\mathbf{v}})} \right) \right] \left[\sup_{x\in[a,b]^d} \mathbb{E}\Big[1 + \|X_T^{1,x}\|^{\mathbf{v}}\Big] \right].
\end{aligned} \tag{197}$$

Jensen's inequality and (189) therefore show that

$$\begin{aligned}
&\sup_{x\in[a,b]^d} \left| u(T,x) - \mathbb{E}\big[\phi(X_T^{1,x})\big] \right| \\
&\leq d^v \left[\sup_{y\in\mathbb{R}^d} \left(\frac{|\varphi(y) - \phi(y)|}{d^v(1 + \|y\|^{\mathbf{v}})} \right) \right] \left[1 + \sup_{x\in[a,b]^d} \mathbb{E}\Big[\|X_T^{1,x}\|^{\mathbf{v}}\Big] \right] \\
&\leq \varepsilon d^v \left[1 + \sup_{x\in[a,b]^d} \left| \mathbb{E}\big[\|X_T^{1,x}\|^{m_1(\mathbf{v})}\big] \right|^{\mathbf{v}/m_1(\mathbf{v})} \right].
\end{aligned} \tag{198}$$

Item (a), Corollary 3.5 (with $m = d$, $C = \|\mu(0)\|$, $c = l$, $W = W^1$, $A = \sqrt{A}$, $X = X^{1,x}$, $p = m_1(\mathbf{v})$ for $x \in [a,b]^d$ in the notation of Corollary 3.5), and the fact that $m_1(m_1(\mathbf{v})-1) = m_1(\mathbf{v}-1)$ hence demonstrate that

$$\begin{aligned}
&\sup_{x\in[a,b]^d} \left| u(T,x) - \mathbb{E}\big[\phi(X_T^{1,x})\big] \right| \\
&\leq \varepsilon d^v \Big(1 + e^{\mathbf{v}lT} \big[\|\mu(0)\| T + \sqrt{m_1(\mathbf{v}-1)\, T \operatorname{Trace}(A)} \\
&+ \sup_{x\in[a,b]^d} \|x\| \big]^{\mathbf{v}} \Big).
\end{aligned} \tag{199}$$

Next note that

$$\begin{aligned}
&\sup_{x\in[a,b]^d} \left| \mathbb{E}\Big[|\phi(X_T^{1,x})|^{m_2(p)} \Big] \right|^{1/m_2(p)} \\
&= \sup_{x\in[a,b]^d} \left| \mathbb{E}\left[\left| \phi(X_T^{1,x}) \frac{d^z\big(1+\|X_T^{1,x}\|^{\mathbf{z}}\big)}{d^z\big(1+\|X_T^{1,x}\|^{\mathbf{z}}\big)} \right|^{m_2(p)} \right] \right|^{1/m_2(p)} \\
&\leq d^z \left[\sup_{y\in\mathbb{R}^d} \frac{|\phi(y)|}{d^z(1+\|y\|^{\mathbf{z}})} \right] \left[\sup_{x\in[a,b]^d} \left| \mathbb{E}\Big[\big(1+\|X_T^{1,x}\|^{\mathbf{z}}\big)^{m_2(p)} \Big] \right|^{1/m_2(p)} \right].
\end{aligned} \tag{200}$$

This, (188), the triangle inequality, and Hölder's inequality show that

$$\begin{aligned}
&\sup_{x\in[a,b]^d} \left| \mathbb{E}\Big[|\phi(X_T^{1,x})|^{m_2(p)} \Big] \right|^{1/m_2(p)} \\
&\leq c d^z \left[1 + \sup_{x\in[a,b]^d} \left| \mathbb{E}\Big[\|X_T^{1,x}\|^{\mathbf{z}m_2(p)} \Big] \right|^{1/m_2(p)} \right] \\
&\leq c d^z \left[1 + \sup_{x\in[a,b]^d} \left| \mathbb{E}\Big[\|X_T^{1,x}\|^{m_1(\mathbf{z}m_2(p))} \Big] \right|^{\mathbf{z}/m_1(\mathbf{z}m_2(p))} \right].
\end{aligned} \tag{201}$$

Item (a), Corollary 3.5 (with $m = d$, $C = \|\mu(0)\|$, $c = l$, $A = \sqrt{A}$, $W = W^1$, $X = X^{1,x}$, $p = m_1(\mathbf{z}m_2(p))$ for $x \in [a,b]^d$ in the notation of Corollary 3.5), and the fact that $m_1(m_1(\mathbf{z}m_2(p)) - 1) = m_1(\mathbf{z}m_2(p) - 1)$ hence demonstrate that

$$\begin{aligned}
&\sup_{x\in[a,b]^d} \left| \mathbb{E}\Big[|\phi(X_T^{1,x})|^{m_2(p)} \Big] \right|^{1/m_2(p)} \\
&\leq c d^z \Big(1 + e^{\mathbf{z}lT} \big[T\|\mu(0)\| + \sqrt{m_1(\mathbf{z}m_2(p)-1)\, T \operatorname{Trace}(A)} \\
&+ \sup_{x\in[a,b]^d} \|x\| \big]^{\mathbf{z}} \Big).
\end{aligned} \tag{202}$$

Next note that

$$
\begin{aligned}
&\sup_{x\in[a,b]^d}\left|\mathbb{E}\!\left[\limsup_{\mathbb{R}^d\setminus\{0\}\ni h\to 0}\left(\frac{|\phi(X_T^{1,x+h})-\phi(X_T^{1,x})|^{m_2(p)}}{\|h\|^{m_2(p)}}\right)\right]\right|^{1/m_2(p)}\\
&=\sup_{x\in[a,b]^d}\left|\mathbb{E}\!\left[\left\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)^*(\nabla\phi)(X_T^{1,x})\right\|^{m_2(p)}\right]\right|^{1/m_2(p)}\\
&\leq \sup_{x\in[a,b]^d}\left|\mathbb{E}\!\left[\left\|(\nabla\phi)(X_T^{1,x})\right\|^{m_2(p)}\left\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)\right\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^{m_2(p)}\right]\right|^{1/m_2(p)}.
\end{aligned}
\tag{203}
$$

Moreover, observe that for all $x\in[a,b]^d$ it holds that

$$
\begin{aligned}
&\left|\mathbb{E}\!\left[\left\|(\nabla\phi)(X_T^{1,x})\right\|^{m_2(p)}\left\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)\right\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^{m_2(p)}\right]\right|^{1/m_2(p)}\\
&=\left|\mathbb{E}\!\left[\left\|(\nabla\phi)(X_T^{1,x})\right\|^{m_2(p)}\frac{\big[d^w\big(1+\|X_T^{1,x}\|^{\mathbf{w}}\big)\big]^{m_2(p)}}{\big[d^w\big(1+\|X_T^{1,x}\|^{\mathbf{w}}\big)\big]^{m_2(p)}}\left\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)\right\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^{m_2(p)}\right]\right|^{1/m_2(p)}\\
&\leq d^w\left[\sup_{y\in\mathbb{R}^d}\frac{\|(\nabla\phi)(y)\|}{d^w(1+\|y\|^{\mathbf{w}})}\right]\left|\mathbb{E}\!\left[\big(1+\|X_T^{1,x}\|^{\mathbf{w}}\big)^{m_2(p)}\left\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)\right\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^{m_2(p)}\right]\right|^{1/m_2(p)}.
\end{aligned}
\tag{204}
$$

This, (203), and (188) prove that

$$
\begin{aligned}
&\sup_{x\in[a,b]^d}\left|\mathbb{E}\!\left[\limsup_{\mathbb{R}^d\setminus\{0\}\ni h\to 0}\left(\frac{|\phi(X_T^{1,x+h})-\phi(X_T^{1,x})|^{m_2(p)}}{\|h\|^{m_2(p)}}\right)\right]\right|^{1/m_2(p)}\\
&\leq c d^w\sup_{x\in[a,b]^d}\left|\mathbb{E}\!\left[\big(1+\|X_T^{1,x}\|^{\mathbf{w}}\big)^{m_2(p)}\left\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)\right\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^{m_2(p)}\right]\right|^{1/m_2(p)}.
\end{aligned}
\tag{205}
$$

This, Hölder's inequality, and the triangle inequality show that

$$
\begin{aligned}
&\sup_{x\in[a,b]^d}\left|\mathbb{E}\!\left[\limsup_{\mathbb{R}^d\setminus\{0\}\ni h\to 0}\left(\frac{|\phi(X_T^{1,x+h})-\phi(X_T^{1,x})|^{m_2(p)}}{\|h\|^{m_2(p)}}\right)\right]\right|^{1/m_2(p)}\\
&\leq c d^w\left[\sup_{x\in[a,b]^d}\left(\left|\mathbb{E}\!\left[\big\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)\big\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^{m_2(p)(m_2(p)+1)}\right]\right|^{1/(m_2(p)(m_2(p)+1))}\right.\right.\\
&\left.\left.\cdot\left|\mathbb{E}\!\left[\big(1+\|X_T^{1,x}\|^{\mathbf{w}}\big)^{m_2(p)+1}\right]\right|^{1/(m_2(p)+1)}\right)\right]\\
&\leq c d^w\left[\sup_{x\in[a,b]^d}\left(\left|\mathbb{E}\!\left[\big\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)\big\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^{m_2(p)(m_2(p)+1)}\right]\right|^{1/(m_2(p)(m_2(p)+1))}\right.\right.\\
&\left.\left.\cdot\left(1+\left|\mathbb{E}\!\left[\big\|X_T^{1,x}\big\|^{\mathbf{w}(m_2(p)+1)}\right]\right|^{1/(m_2(p)+1)}\right)\right)\right].
\end{aligned}
\tag{206}
$$

Jensen's inequality therefore demonstrates that

$$\begin{aligned}
&\sup_{x\in[a,b]^d}\left|\mathbb{E}\left[\limsup_{\mathbb{R}^d\setminus\{0\}\ni h\to 0}\left(\frac{|\phi(X_T^{1,x+h})-\phi(X_T^{1,x})|^{m_2(p)}}{\|h\|^{m_2(p)}}\right)\right]\right|^{1/m_2(p)}\\
&\le cd^w\left[\sup_{x\in[a,b]^d}\left(\left|\mathbb{E}\Big[\big\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)\big\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^{m_2(p)(m_2(p)+1)}\Big]\right|^{1/(m_2(p)(m_2(p)+1))}\right.\right.\\
&\quad\left.\left.\cdot\left(1+\left|\mathbb{E}\Big[\big\|X_T^{1,x}\big\|^{m_1(\mathbf{w}(m_2(p)+1))}\Big]\right|^{\mathbf{w}/m_1(\mathbf{w}(m_2(p)+1))}\right)\right)\right].
\end{aligned}\tag{207}$$

Next note that (194) implies that for all $\omega\in\Omega$, $x,y\in\mathbb{R}^d$ it holds that

$$\big(\tfrac{\partial}{\partial x}X_T^{1,x}(\omega)\big)y = X_T^{1,y}(\omega)-X_T^{1,0}(\omega).\tag{208}$$

This and Hölder's inequality show that for all $x\in\mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left|\mathbb{E}\Big[\big\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)\big\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^{m_2(p)(m_2(p)+1)}\Big]\right|^{1/(m_2(p)(m_2(p)+1))}\\
&\le\left|\mathbb{E}\left[\left(\sum_{i=1}^d\big\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)e_i\big\|^2\right)^{(m_2(p)(m_2(p)+1))/2}\right]\right|^{1/(m_2(p)(m_2(p)+1))}\\
&\le\left|\mathbb{E}\left[d^{\frac{m_2(p)(m_2(p)+1)-2}{2}}\left(\sum_{i=1}^d\big\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)e_i\big\|^{m_2(p)(m_2(p)+1)}\right)\right]\right|^{1/(m_2(p)(m_2(p)+1))}\\
&=d^{\frac{m_2(p)(m_2(p)+1)-2}{2m_2(p)(m_2(p)+1)}}\left|\sum_{i=1}^d\mathbb{E}\left[\big\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)e_i\big\|^{m_2(p)(m_2(p)+1)}\right]\right|^{1/(m_2(p)(m_2(p)+1))}\\
&\le d^{1/2}\max_{i\in\{1,\dots,d\}}\left(\left|\mathbb{E}\left[\big\|\big(\tfrac{\partial}{\partial x}X_T^{1,x}\big)e_i\big\|^{m_2(p)(m_2(p)+1)}\right]\right|^{1/(m_2(p)(m_2(p)+1))}\right)\\
&=d^{1/2}\max_{i\in\{1,\dots,d\}}\left(\left|\mathbb{E}\left[\big\|X_T^{1,e_i}-X_T^{1,0}\big\|^{m_2(p)(m_2(p)+1)}\right]\right|^{1/(m_2(p)(m_2(p)+1))}\right).
\end{aligned}\tag{209}$$

Combining this with (207) shows that

$$\begin{aligned}
&\sup_{x\in[a,b]^d}\left|\mathbb{E}\left[\limsup_{\mathbb{R}^d\setminus\{0\}\ni h\to 0}\left(\frac{|\phi(X_T^{1,x+h})-\phi(X_T^{1,x})|^{m_2(p)}}{\|h\|^{m_2(p)}}\right)\right]\right|^{1/m_2(p)}\\
&\le cd^{w+1/2}\max_{i\in\{1,\dots,d\}}\left(\left|\mathbb{E}\left[\big\|X_T^{1,e_i}-X_T^{1,0}\big\|^{m_2(p)(m_2(p)+1)}\right]\right|^{1/(m_2(p)(m_2(p)+1))}\right)\\
&\quad\cdot\left[1+\sup_{x\in[a,b]^d}\left|\mathbb{E}\Big[\big\|X_T^{1,x}\big\|^{m_1(\mathbf{w}(m_2(p)+1))}\Big]\right|^{\mathbf{w}/m_1(\mathbf{w}(m_2(p)+1))}\right].
\end{aligned}\tag{210}$$

Item (a), Corollary 3.5 (with $m = d$, $C = \|\mu(0)\|$, $c = l$, $A = \sqrt{A}$, $W = W^1$, $X = X^{1,x}$, $p = m_1(\mathbf{w}(m_2(p)+1))$ for $x \in [a,b]^d$ in the notation of Corollary 3.5), Lemma 3.6 (with $m = d$, $l = l$, $A = \sqrt{A}$, $W = W^1$, $X = X^1$, $x = e_i$, $y = 0$, $p = m_2(p)(m_2(p)+1)$ for $i \in \{1,\ldots,d\}$ in the notation of Lemma 3.6), and the fact that $m_1(m_1(\mathbf{w}(m_2(p)+1))-1) = m_1(\mathbf{w}(m_2(p)+1)-1)$ hence demonstrate that

$$\begin{aligned}
&\sup_{x\in[a,b]^d}\left|\mathbb{E}\left[\limsup_{\mathbb{R}^d\backslash\{0\}\ni h\to 0}\left(\frac{|\phi(X_T^{1,x+h})-\phi(X_T^{1,x})|^{m_2(p)}}{\|h\|^{m_2(p)}}\right)\right]\right|^{1/m_2(p)} \\
&\leq cd^{w+1/2}\max_{i\in\{1,\ldots,d\}}\left(e^{lT}\|e_i\|\right)\Big(1+e^{\mathbf{w}lT}\big[T\|\mu(0)\| \\
&+\sqrt{m_1(\mathbf{w}(m_2(p)+1)-1)\,T\,\mathrm{Trace}(A)}+\sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{w}}\Big) \\
&= cd^{w+1/2}e^{lT}\Big(1+e^{\mathbf{w}lT}\big[T\|\mu(0)\| \\
&+\sqrt{m_1(\mathbf{w}m_2(p)+\mathbf{w}-1)\,T\,\mathrm{Trace}(A)}+\sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{w}}\Big).
\end{aligned} \tag{211}$$

Combining this, (202), and (195) with item (ii) in Lemma 2.16 (with $n = n$, $\xi_i = ((\mathbb{R}^d\times\Omega)\ni(x,\omega)\mapsto\phi(X_T^{i,x}(\omega))\in\mathbb{R})$ for $i\in\{1,\ldots,n\}$ in the notation of Lemma 2.16) implies that

$$\begin{aligned}
&\left|\mathbb{E}\left[\sup_{x\in[a,b]^d}\left|\mathbb{E}\big[\phi(X_T^{1,x})\big]-\left[\frac{1}{n}\sum_{k=1}^n\phi(X_T^{k,x})\right]\right|^p\right]\right|^{1/p} \\
&\leq\frac{4\mathfrak{K}\zeta}{\sqrt{n}}\bigg(cd^z\Big(1+e^{\mathbf{z}lT}\big[T\|\mu(0)\|+\sqrt{m_1(\mathbf{z}m_2(p)-1)\,T\,\mathrm{Trace}(A)} \\
&+\sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{z}}\Big)+(b-a)cd^{w+1/2}e^{lT}\Big(1+e^{\mathbf{w}lT}\big[T\|\mu(0)\| \\
&+\sqrt{m_1(\mathbf{w}m_2(p)+\mathbf{w}-1)\,T\,\mathrm{Trace}(A)}+\sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{w}}\Big)\bigg).
\end{aligned} \tag{212}$$

Grohs et al. [28, Proposition 3.3] hence shows that there exists $\omega_n \in \Omega$ which satisfies that

$$\begin{aligned}
&\sup_{x\in[a,b]^d}\left|\mathbb{E}\big[\phi(X_T^{1,x})\big]-\left[\frac{1}{n}\sum_{k=1}^n\phi(X_T^{k,x}(\omega_n))\right]\right| \\
&\leq\frac{4\mathfrak{K}\zeta}{\sqrt{n}}\bigg(cd^z\Big(1+e^{\mathbf{z}lT}\big[T\|\mu(0)\|+\sqrt{m_1(\mathbf{z}m_2(p)-1)\,T\,\mathrm{Trace}(A)} \\
&+\sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{z}}\Big)+(b-a)cd^{w+1/2}e^{lT}\Big(1+e^{\mathbf{w}lT}\big[T\|\mu(0)\| \\
&+\sqrt{m_1(\mathbf{w}m_2(p)+\mathbf{w}-1)\,T\,\mathrm{Trace}(A)}+\sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{w}}\Big)\bigg).
\end{aligned} \tag{213}$$

Combining this, (194), (196), and (199) with the triangle inequality demonstrates that

$$
\begin{aligned}
&\sup_{x\in[a,b]^d}\left|u(T,x)-\left[\frac{1}{n}\sum_{k=1}^{n}\phi(\mathcal{W}_{k,\omega_n}x+\mathcal{B}_{k,\omega_n})\right]\right|\\
&=\sup_{x\in[a,b]^d}\left|\mathbb{E}[\varphi(X_T^{1,x})]-\left[\frac{1}{n}\sum_{k=1}^{n}\phi(X_T^{k,x}(\omega_n))\right]\right|\\
&\leq\sup_{x\in[a,b]^d}\left|\mathbb{E}[\varphi(X_T^{1,x})]-\mathbb{E}[\phi(X_T^{1,x})]\right|\\
&+\sup_{x\in[a,b]^d}\left|\mathbb{E}[\phi(X_T^{1,x})]-\left[\frac{1}{n}\sum_{k=1}^{n}\phi(X_T^{k,x}(\omega_n))\right]\right|\\
&\leq\varepsilon d^v\Big(1+e^{\mathbf{v}lT}\big[T\|\mu(0)\|+\sqrt{m_1(\mathbf{v}-1)\,T\,\mathrm{Trace}(A)}\\
&+\sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{v}}\Big)\\
&+\frac{4\mathfrak{K}\zeta}{\sqrt{n}}\left(cd^z\Big(1+e^{\mathbf{z}lT}\big[T\|\mu(0)\|+\sqrt{m_1(\mathbf{z}m_2(p)-1)\,T\,\mathrm{Trace}(A)}\right.\\
&+\sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{z}}\Big)+(b-a)cd^{w+1/2}e^{lT}\Big(1+e^{\mathbf{w}lT}\big[T\|\mu(0)\|\\
&\left.+\sqrt{m_1(\mathbf{w}m_2(p)+\mathbf{w}-1)\,T\,\mathrm{Trace}(A)}+\sup_{x\in[a,b]^d}\|x\|\big]^{\mathbf{w}}\Big)\right).
\end{aligned}
\tag{214}
$$

The proof of Lemma 4.1 is thus completed. □

Corollary 4.2. *Let $d,n\in\mathbb{N}$, $\varphi\in C(\mathbb{R}^d,\mathbb{R})$, $\alpha,a\in\mathbb{R}$, $\beta\in[0,\infty)$, $b\in(a,\infty)$, $\varepsilon,T,c\in(0,\infty)$, $v,\mathbf{v},w,\mathbf{w},z,\mathbf{z}\in[0,\infty)$, $\mu\in\mathbb{R}^d$, let $\|\cdot\|:\mathbb{R}^d\to[0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $\phi\in C^1(\mathbb{R}^d,\mathbb{R})$, let $A=(A_{i,j})_{(i,j)\in\{1,\ldots,d\}^2}\in\mathbb{R}^{d\times d}$ be a symmetric and positive semi-definite matrix, assume for all $x\in\mathbb{R}^d$ that*

$$|\phi(x)|\leq cd^z(1+\|x\|^{\mathbf{z}}),\qquad \|(\nabla\phi)(x)\|\leq cd^w(1+\|x\|^{\mathbf{w}}),\tag{215}$$

$$|\varphi(x)-\phi(x)|\leq\varepsilon d^v(1+\|x\|^{\mathbf{v}}),\qquad \sqrt{\mathrm{Trace}(A)}\leq cd^\beta,\tag{216}$$

and $\|\mu\|\leq cd^\alpha$, let $u\in C([0,T]\times\mathbb{R}^d,\mathbb{R})$, assume for all $x\in\mathbb{R}^d$ that $u(0,x)=\varphi(x)$, assume that $\inf_{\gamma\in(0,\infty)}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\big(\frac{|u(t,x)|}{1+\|x\|^\gamma}\big)<\infty$, and assume that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u)(t,x)=\textstyle\sum_{i,j=1}^d A_{i,j}\,(\frac{\partial^2}{\partial x_i\partial x_j}u)(t,x)+\sum_{i=1}^d\mu_i\,(\frac{\partial}{\partial x_i}u)(t,x)\tag{217}$$

for $(t,x)\in(0,T)\times\mathbb{R}^d$. Then there exist $W_1,\ldots,W_n\in\mathbb{R}^{d\times d}$, $B_1,\ldots,B_n\in\mathbb{R}^d$

such that

$$\begin{aligned}
&\sup_{x\in[a,b]^d}\left|u(T,x)-\left[\frac{1}{n}\sum_{k=1}^{n}\phi(W_kx+B_k)\right]\right| \\
&\leq \varepsilon d^{v+\mathbf{v}\max\{\alpha,\beta,1/2\}} \\
&\cdot\Big(1+\big[\sqrt{2}\max\{1,T\}\max\{1,\sqrt{\mathbf{v}}\}(2c+\max\{|a|,|b|\})\big]^{\mathbf{v}}\Big) \\
&+d^{1+\max\{z+\mathbf{z}\max\{\alpha,\beta+1\},w+1/2+\mathbf{w}\max\{\alpha,\beta+1\}\}} \\
&\cdot\frac{32\sqrt{e}c}{\sqrt{n}}\Big[(1+(b-a)) \\
&\cdot\Big(1+\big[\sqrt{6}\max\{1,T\}\max\{1,\sqrt{\mathbf{z}},\sqrt{\mathbf{w}}\}(2c+\max\{|a|,|b|\})\big]^{\max\{\mathbf{z},\mathbf{w}\}}\Big)\Big].
\end{aligned}\tag{218}$$

Proof of Corollary 4.2. Throughout this proof let $\mathfrak{K}\in(0,\infty)$ be the $(\max\{2,d^2\},2)$-Kahane–Khintchine constant (cf. Definition 2.1 and Lemma 2.2). Observe that for all $x\in[a,b]^d$ it holds that

$$\|x\|=\left[\sum_{i=1}^{d}|x_i|^2\right]^{1/2}\leq\left[\sum_{i=1}^{d}[\max\{|a|,|b|\}]^2\right]^{1/2}=d^{1/2}\max\{|a|,|b|\}.\tag{219}$$

This proves that

$$\sup_{x\in[a,b]^d}\|x\|\leq d^{1/2}\max\{|a|,|b|\}.\tag{220}$$

Next note that Lemma 2.2 (with $p=\max\{2,d^2\}$ in the notation of Lemma 2.2) ensures that

$$\mathfrak{K}\leq\sqrt{\max\{1,\max\{2,d^2\}-1\}}\leq d.\tag{221}$$

Combining this, (220), Corollary 2.15, (216), and the hypothesis that $\|\mu\|\leq cd^\alpha$ with Lemma 4.1 (with $l=0$, $n=n$, $\zeta=8\sqrt{e}$, $p=d^2$, $\mu=(\mathbb{R}^d\ni x\mapsto\mu\in\mathbb{R}^d)$, $A=2A$ in the notation of Lemma 4.1) demonstrates that there exist $W_1,\dots,W_n\in\mathbb{R}^{d\times d}$, $B_1,\dots,B_n\in\mathbb{R}^d$ which satisfy that

$$\begin{aligned}
&\sup_{x\in[a,b]^d}\left|u(T,x)-\left[\frac{1}{n}\sum_{k=1}^{n}\phi(W_kx+B_k)\right]\right| \\
&\leq\varepsilon d^v\Big(1+\big[Tcd^\alpha+\sqrt{2\max\{1,\mathbf{v}-1\}\,T}cd^\beta \\
&+d^{1/2}\max\{|a|,|b|\}\big]^{\mathbf{v}}\Big) \\
&+\frac{32\sqrt{e}d}{\sqrt{n}}\bigg(cd^z\Big(1+\big[Tcd^\alpha+\sqrt{2\max\{1,\mathbf{z}\max\{2,d^2\}-1\}\,T}cd^\beta \\
&+d^{1/2}\max\{|a|,|b|\}\big]^{\mathbf{z}}\Big)+(b-a)cd^{w+1/2}\Big(1+\big[Tcd^\alpha \\
&+\sqrt{2\max\{1,\mathbf{w}\max\{2,d^2\}+\mathbf{w}-1\}\,T}cd^\beta+d^{1/2}\max\{|a|,|b|\}\big]^{\mathbf{w}}\Big)\bigg).
\end{aligned}\tag{222}$$

Next note that

$$\begin{aligned}\max\{1,\mathbf{z}\max\{2,d^2\}-1\} &\leq \max\{1,\mathbf{z}\}\max\{2,d^2\} \leq \max\{1,\mathbf{z}\}(d^2+1)\\ &\leq 2\max\{1,\mathbf{z}\}d^2.\end{aligned} \tag{223}$$

Moreover, note that

$$\begin{aligned}&\max\{1,\mathbf{w}\max\{2,d^2\}+\mathbf{w}-1\}\\ &\leq \max\{1,\mathbf{w}\}\max\{2,d^2\}+\max\{1,\mathbf{w}\} \leq \max\{1,\mathbf{w}\}(d^2+1)+\max\{1,\mathbf{w}\}\\ &\leq 2\max\{1,\mathbf{w}\}d^2+\max\{1,\mathbf{w}\} \leq 3\max\{1,\mathbf{w}\}d^2.\end{aligned} \tag{224}$$

Combining this, the fact that $\max\{1,\mathbf{v}-1\} \leq \max\{1,\mathbf{v}\}$, (222), and (223) demonstrates that

$$\begin{aligned}&\sup_{x\in[a,b]^d}\left|u(T,x)-\left[\frac{1}{n}\sum_{k=1}^{n}\phi(W_kx+B_k)\right]\right|\\ &\leq \varepsilon d^{v+\mathbf{v}\max\{\alpha,\beta,1/2\}}\Big(1+\big[Tc+\sqrt{2\max\{1,\mathbf{v}\}Tc}+\max\{|a|,|b|\}\big]^{\mathbf{v}}\Big)\\ &+d^{1+\max\{z+\mathbf{z}\max\{\alpha,\beta+1\},w+1/2+\mathbf{w}\max\{\alpha,\beta+1\}\}}\\ &\cdot\frac{32\sqrt{e}c}{\sqrt{n}}\Bigg(\Big(1+\big[Tc+\sqrt{4\max\{1,\mathbf{z}\}Tc}+\max\{|a|,|b|\}\big]^{\mathbf{z}}\Big)\\ &+(b-a)\Big(1+\big[Tc+\sqrt{6\max\{1,\mathbf{w}\}Tc}+\max\{|a|,|b|\}\big]^{\mathbf{w}}\Big)\Bigg).\end{aligned} \tag{225}$$

Hence, we obtain that

$$\begin{aligned}&\sup_{x\in[a,b]^d}\left|u(T,x)-\left[\frac{1}{n}\sum_{k=1}^{n}\phi(W_kx+B_k)\right]\right|\\ &\leq \varepsilon d^{v+\mathbf{v}\max\{\alpha,\beta,1/2\}}\\ &\cdot\Big(1+\big[\sqrt{2}\max\{1,T\}\max\{1,\sqrt{\mathbf{v}}\}(2c+\max\{|a|,|b|\})\big]^{\mathbf{v}}\Big)\\ &+d^{1+\max\{z+\mathbf{z}\max\{\alpha,\beta+1\},w+1/2+\mathbf{w}\max\{\alpha,\beta+1\}\}}\\ &\cdot\frac{32\sqrt{e}c}{\sqrt{n}}\Big[(1+(b-a))\\ &\cdot\Big(1+\big[\sqrt{6}\max\{1,T\}\max\{1,\sqrt{\mathbf{z}},\sqrt{\mathbf{w}}\}(2c+\max\{|a|,|b|\})\big]^{\max\{\mathbf{z},\mathbf{w}\}}\Big)\Big].\end{aligned} \tag{226}$$

The proof of Corollary 4.2 is thus completed. □

Corollary 4.3. *Let $d,n\in\mathbb{N}$, $\varphi\in C(\mathbb{R}^d,\mathbb{R})$, $\alpha,a\in\mathbb{R}$, $\beta\in[0,\infty)$, $b\in(a,\infty)$, $\varepsilon,T\in(0,\infty)$, $c\in[1/2,\infty)$, $v,\mathbf{v},w,\mathbf{w},z,\mathbf{z}\in[0,\infty)$, $\mu\in\mathbb{R}^d$, let $\|\cdot\|:\mathbb{R}^d\to[0,\infty)$*

be the standard norm on $\mathbb{R}^d$*, let* $A = (A_{i,j})_{(i,j)\in\{1,\ldots,d\}^2} \in \mathbb{R}^{d\times d}$ *be a symmetric and positive semi-definite matrix, let* $\phi \in C^1(\mathbb{R}^d, \mathbb{R})$*, assume for all* $x \in \mathbb{R}^d$ *that*

$$|\phi(x)| \le cd^z(1 + \|x\|^{\mathbf{z}}), \qquad \|(\nabla\phi)(x)\| \le cd^w(1 + \|x\|^{\mathbf{w}}), \tag{227}$$

$$|\varphi(x) - \phi(x)| \le \varepsilon d^v(1 + \|x\|^{\mathbf{v}}), \qquad \sqrt{\mathrm{Trace}(A)} \le cd^\beta, \tag{228}$$

and $\|\mu\| \le cd^\alpha$*, let* $u \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$*, assume for all* $x \in \mathbb{R}^d$ *that* $u(0,x) = \varphi(x)$*, assume that* $\inf_{\gamma\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \big(\frac{|u(t,x)|}{1+\|x\|^\gamma}\big) < \infty$*, and assume that* $u|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u)(t,x) = \textstyle\sum_{i,j=1}^d A_{i,j}\, (\tfrac{\partial^2}{\partial x_i \partial x_j}u)(t,x) + \sum_{i=1}^d \mu_i\, (\tfrac{\partial}{\partial x_i}u)(t,x) \tag{229}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$*. Then there exist* $W_1, \ldots, W_n \in \mathbb{R}^{d\times d}$*,* $B_1, \ldots, B_n \in \mathbb{R}^d$ *such that*

$$\begin{aligned}
&\sup_{x\in[a,b]^d} \left| u(T,x) - \left[\frac{1}{n}\sum_{k=1}^n \phi(W_k x + B_k)\right]\right| \\
&\le \varepsilon d^{v+\mathbf{v}\max\{\alpha,\beta,1/2\}} \big[5c + T + \sqrt{\mathbf{v}} + |a| + |b|\big]^{4\mathbf{v}+1} \\
&+ d^{1+\max\{z+\mathbf{z}\max\{\alpha,\beta+1\}, w+1/2+\mathbf{w}\max\{\alpha,\beta+1\}\}} \\
&\cdot \frac{1}{\sqrt{n}}\big[7c + T + \sqrt{\mathbf{z}} + \sqrt{\mathbf{w}} + |a| + |b|\big]^{5+4(\mathbf{z}+\mathbf{w})}.
\end{aligned} \tag{230}$$

Proof of Corollary 4.3. Observe that Corollary 4.2 ensures that there exist $W_1, \ldots, W_n \in \mathbb{R}^{d\times d}$, $B_1, \ldots, B_n \in \mathbb{R}^d$ which satisfy that

$$\begin{aligned}
&\sup_{x\in[a,b]^d} \left| u(T,x) - \left[\frac{1}{n}\sum_{k=1}^n \phi(W_k x + B_k)\right]\right| \\
&\le \varepsilon d^{v+\mathbf{v}\max\{\alpha,\beta,1/2\}} \\
&\cdot \Big(1 + \big[\sqrt{2}\max\{1,T\}\max\{1,\sqrt{\mathbf{v}}\}(2c + \max\{|a|,|b|\})\big]^{\mathbf{v}}\Big) \\
&+ d^{1+\max\{z+\mathbf{z}\max\{\alpha,\beta+1\}, w+1/2+\mathbf{w}\max\{\alpha,\beta+1\}\}}\, \frac{32\sqrt{e}c}{\sqrt{n}}\Big[(1 + (b-a)) \\
&\cdot \Big(1 + \big[\sqrt{6}\max\{1,T\}\max\{1,\sqrt{\mathbf{z}},\sqrt{\mathbf{w}}\}(2c + \max\{|a|,|b|\})\big]^{\max\{\mathbf{z},\mathbf{w}\}}\Big)\Big].
\end{aligned} \tag{231}$$

Next note that the assumption that $c \in [1/2, \infty)$ implies that

$$\begin{aligned}
&1 + \big[\sqrt{2}\max\{1,T\}\max\{1,\sqrt{\mathbf{v}}\}(2c + \max\{|a|,|b|\})\big]^{\mathbf{v}} \\
&\le 1 + \max\{\sqrt{2}, T, \sqrt{\mathbf{v}}, 2c + \max\{|a|,|b|\}\}^{4\mathbf{v}} \\
&\le 1 + \big[2\sqrt{2}c + 2c + T + \sqrt{\mathbf{v}} + |a| + |b|\big]^{4\mathbf{v}} \\
&\le 2\big[5c + T + \sqrt{\mathbf{v}} + |a| + |b|\big]^{4\mathbf{v}} \\
&\le \big[5c + T + \sqrt{\mathbf{v}} + |a| + |b|\big]^{4\mathbf{v}+1}.
\end{aligned} \tag{232}$$

Moreover, note that the assumption that $c \in [1/2, \infty)$ shows that

$$
\begin{aligned}
&32\sqrt{e}c\Big[(1+(b-a)) \\
&\cdot\Big(1+\big[\sqrt{6}\max\{1,T\}\max\{1,\sqrt{\mathbf{z}},\sqrt{\mathbf{w}}\}(2c+\max\{|a|,|b|\})\big]^{\max\{\mathbf{z},\mathbf{w}\}}\Big)\Big] \\
&\leq 53c(1+|a|+|b|)\Big[1+\big(2\sqrt{6}c+2c+T+\sqrt{\mathbf{z}}+\sqrt{\mathbf{w}}+|a|+|b|\big)^{4\max\{\mathbf{z},\mathbf{w}\}}\Big] \\
&\leq 53c(1+|a|+|b|)\Big[1+\big(7c+T+\sqrt{\mathbf{z}}+\sqrt{\mathbf{w}}+|a|+|b|\big)^{4\max\{\mathbf{z},\mathbf{w}\}}\Big].
\end{aligned}
\tag{233}
$$

Hence, we obtain that

$$
\begin{aligned}
&32\sqrt{e}c\Big[(1+(b-a)) \\
&\cdot\Big(1+\big[\sqrt{6}\max\{1,T\}\max\{1,\sqrt{\mathbf{z}},\sqrt{\mathbf{w}}\}(2c+\max\{|a|,|b|\})\big]^{\max\{\mathbf{z},\mathbf{w}\}}\Big)\Big] \\
&\leq \big[7c+T+\sqrt{\mathbf{z}}+\sqrt{\mathbf{w}}+|a|+|b|\big]^{5+4(\mathbf{z}+\mathbf{w})}.
\end{aligned}
\tag{234}
$$

Combining this with (231) and (232) establishes (230). The proof of Corollary 4.3 is thus completed. □

4.2 Qualitative error estimates

In this subsection we provide in Proposition 4.4 below a qualitative approximation result for viscosity solutions (cf., for example, Hairer et al. [31]) of Kolomogorov PDEs with constant coefficient functions. Informally speaking, we can think of the approximations in Proposition 4.4 as linear combinations of realizations of ANNs with a suitable continuously differentiable activation function. Proposition 4.4 will be employed in our proof of Proposition 4.6 in Subsection 4.3 below.

Proposition 4.4. *Let $d \in \mathbb{N}$, $\varphi \in C(\mathbb{R}^d, \mathbb{R})$, $\alpha, a \in \mathbb{R}$, $\beta \in [0, \infty)$, $b \in (a, \infty)$, $r, T \in (0, \infty)$, $c \in [1/2, \infty)$, $v, \mathbf{v}, w, \mathbf{w}, z, \mathbf{z} \in [0, \infty)$, $C = \frac{1}{2}[5c + T + \sqrt{\mathbf{v}} + |a| + |b|]^{-4\mathbf{v}-1}$, $\mathbf{C} = 4[7c+T+\sqrt{\mathbf{z}}+\sqrt{\mathbf{w}}+|a|+|b|]^{10+8(\mathbf{z}+\mathbf{w})}$, $p = v+\mathbf{v}\max\{\alpha, \beta, 1/2\}$, $\mathbf{p} = 2 + \max\{2z + 2\mathbf{z}\max\{\alpha, \beta + 1\}, 2w + 1 + 2\mathbf{w}\max\{\alpha, \beta + 1\}\}$, $\mu \in \mathbb{R}^d$, let $\|\cdot\| \colon \mathbb{R}^d \to [0, \infty)$ be the standard norm on $\mathbb{R}^d$, let $\phi_\varepsilon \in C^1(\mathbb{R}^d, \mathbb{R})$, $\varepsilon \in (0, r]$, let $A = (A_{i,j})_{(i,j)\in\{1,\ldots,d\}^2} \in \mathbb{R}^{d\times d}$ be a symmetric and positive semi-definite matrix, and assume for all $\varepsilon \in (0, r]$, $x \in \mathbb{R}^d$ that*

$$|\phi_\varepsilon(x)| \leq cd^z(1 + \|x\|^{\mathbf{z}}), \qquad \|(\nabla\phi_\varepsilon)(x)\| \leq cd^w(1 + \|x\|^{\mathbf{w}}), \tag{235}$$

$$|\varphi(x) - \phi_\varepsilon(x)| \leq \varepsilon cd^v(1 + \|x\|^{\mathbf{v}}), \qquad \sqrt{\operatorname{Trace}(A)} \leq cd^\beta, \tag{236}$$

and $\|\mu\| \leq cd^\alpha$. Then

(i) there exists a unique $u \in C([0, T]\times\mathbb{R}^d, \mathbb{R})$ which satisfies for all $x \in \mathbb{R}^d$ that $u(0, x) = \varphi(x)$, which satisfies that $\inf_{\gamma\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \big(\frac{|u(t,x)|}{1+\|x\|^\gamma}\big) <$

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

$$(\tfrac{\partial}{\partial t}u)(t,x) = \textstyle\sum_{i,j=1}^{d} A_{i,j}\,(\tfrac{\partial^2}{\partial x_i \partial x_j}u)(t,x) + \sum_{i=1}^{d}\mu_i\,(\tfrac{\partial}{\partial x_i}u)(t,x) \tag{237}$$

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

(ii) it holds for all $\varepsilon\in(0,r]$, $n\in\mathbb{N}\cap[\varepsilon^{-2}d^{\mathbf{p}}\mathbf{C},\infty)$ *that there exist* $W_1,\ldots,W_n\in\mathbb{R}^{d\times d}$, $B_1,\ldots,B_n\in\mathbb{R}^d$ *such that*

$$\sup_{x\in[a,b]^d}\left|u(T,x)-\left[\frac{1}{n}\sum_{k=1}^{n}\phi_{\varepsilon c^{-1}d^{-p}C}(W_k x+B_k)\right]\right|\le\varepsilon. \tag{238}$$

Proof of Proposition 4.4. First, note that (235) and (236) ensure that $\varphi\colon\mathbb{R}^d\to\mathbb{R}$ is an at most polynomially growing function. Grohs et al. [28, Corollary 2.23] (see also Hairer et al. [31, Corollary 4.17]) hence implies 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 $\inf_{\gamma\in(0,\infty)}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\big(\frac{|u(t,x)|}{1+\|x\|^\gamma}\big)<\infty$, and which satisfies that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u)(t,x) = \textstyle\sum_{i,j=1}^{d} A_{i,j}\,(\tfrac{\partial^2}{\partial x_i \partial x_j}u)(t,x) + \sum_{i=1}^{d}\mu_i\,(\tfrac{\partial}{\partial x_i}u)(t,x) \tag{239}$$

for $(t,x)\in(0,T)\times\mathbb{R}^d$. This establishes item (i). Next note that the hypothesis that $c\in[1/2,\infty)$ shows that

$$\frac{C}{c}=\frac{1}{2c}\frac{1}{[5c+T+\sqrt{\mathbf{v}}+|a|+|b|]^{4\mathbf{v}+1}}\le 1. \tag{240}$$

Hence, we obtain that for all $\varepsilon\in(0,r]$ it holds that $\varepsilon c^{-1}d^{-p}C\le\varepsilon$. This and (236) prove that for all $\varepsilon\in(0,r]$, $x\in\mathbb{R}^d$ it holds that

$$|\varphi(x)-\phi_{\varepsilon c^{-1}d^{-p}C}(x)|\le\varepsilon c^{-1}d^{-p}Ccd^{v}(1+\|x\|^{\mathbf{v}})=\varepsilon d^{-p}Cd^{v}(1+\|x\|^{\mathbf{v}}). \tag{241}$$

Corollary 4.3 (with $\varepsilon=\varepsilon d^{-p}C$, $\phi=\phi_{\varepsilon c^{-1}d^{-p}C}$ for $\varepsilon\in(0,r]$ in the notation of Corollary 4.3) hence demonstrates that for all $\varepsilon\in(0,r]$, $n\in\mathbb{N}\cap[\varepsilon^{-2}d^{\mathbf{p}}\mathbf{C},\infty)$ there exist $W_1,\ldots,W_n\in\mathbb{R}^{d\times d}$, $B_1,\ldots,B_n\in\mathbb{R}^d$ such that

$$\begin{aligned}&\sup_{x\in[a,b]^d}\left|u(T,x)-\left[\frac{1}{n}\sum_{k=1}^{n}\phi_{\varepsilon c^{-1}d^{-p}C}(W_k x+B_k)\right]\right|\\&\le\varepsilon d^{-p}C\frac{d^p}{2C}+\frac{\sqrt{\mathbf{C}d^{\mathbf{p}}}}{2\sqrt{n}}\le\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.\end{aligned} \tag{242}$$

This establishes item (ii). The proof of Proposition 4.4 is thus completed. □

4.3 Qualitative error estimates for artificial neural networks (ANNs)

In this subsection we prove in Corollary 4.7 below that ANNs with continuously differentiable activation functions can overcome the curse of dimensionality in the uniform approximation of viscosity solutions of Kolmogorov PDEs with constant coefficient functions. Corollary 4.7 is an immediate consequence of Proposition 4.6 below. Proposition 4.6, in turn, follows from Proposition 4.4 above and the well-known fact that linear combinations of realizations of ANNs are again realizations of ANNs. To formulate Corollary 4.7 we introduce in Setting 4.5 below a common framework from the scientific literature (cf., e.g., Grohs et al. [29, Section 2.1] and Petersen & Voigtlaender [52, Section 2]) to mathematically describe ANNs.

Setting 4.5. *Let* $\mathbf{N}$ *be the set given by*

$$\mathbf{N} = \cup_{L\in\mathbb{N}\cap[2,\infty)} \cup_{(l_0,\dots,l_L)\in((\mathbb{N}^L)\times\{1\})} \big(\times_{k=1}^{L} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\big), \tag{243}$$

let $\mathbf{a} \in C^1(\mathbb{R},\mathbb{R})$, *let* $\mathbf{A}_n\colon \mathbb{R}^n \to \mathbb{R}^n$, $n \in \mathbb{N}$, *satisfy for all* $n \in \mathbb{N}$, $x = (x_1,\dots,x_n) \in \mathbb{R}^n$ *that* $\mathbf{A}_n(x) = (\mathbf{a}(x_1),\dots,\mathbf{a}(x_n))$, *and let* $\mathcal{N},\mathcal{L},\mathcal{P},\mathfrak{P}\colon \mathbf{N} \to \mathbb{N}$ *and* $\mathcal{R}\colon \mathbf{N} \to \cup_{d\in\mathbb{N}} C(\mathbb{R}^d,\mathbb{R})$ *satisfy for all* $L \in \mathbb{N}\cap[2,\infty)$, $(l_0,\dots,l_L) \in ((\mathbb{N}^L)\times\{1\})$, $\Phi = ((W_1,B_1),\dots,(W_L,B_L)) = (((W_1^{(i,j)})_{i\in\{1,\dots,l_1\},j\in\{1,\dots,l_0\}}, (B_1^i)_{i\in\{1,\dots,l_1\}}), \dots, ((W_L^{(i,j)})_{i\in\{1,\dots,l_L\},j\in\{1,\dots,l_{L-1}\}}, (B_k^i)_{i\in\{1,\dots,l_L\}})) \in (\times_{k=1}^{L} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $x_0 \in \mathbb{R}^{l_0}$, $\dots$, $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ *with* $\forall\, k \in \mathbb{N}\cap(0,L)\colon x_k = \mathbf{A}_{l_k}(W_k x_{k-1}+B_k)$ *that* $\mathcal{N}(\Phi) = \sum_{k=0}^{L} l_k$, $\mathcal{L}(\Phi) = L+1$, $\mathcal{P}(\Phi) = \sum_{k=1}^{L} l_k(l_{k-1}+1)$, $(\mathcal{R}\Phi) \in C(\mathbb{R}^{l_0},\mathbb{R})$, $(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L$, *and*

$$\mathfrak{P}(\Phi) = \sum_{k=1}^{L}\sum_{i=1}^{l_k}\left[\mathbb{1}_{\mathbb{R}\setminus\{0\}}(B_k^i) + \sum_{j=1}^{l_{k-1}} \mathbb{1}_{\mathbb{R}\setminus\{0\}}(W_k^{(i,j)})\right]. \tag{244}$$

Proposition 4.6. *Assume Setting 4.5, let* $d \in \mathbb{N}$, $\mu \in \mathbb{R}^d$, $\varphi \in C(\mathbb{R}^d,\mathbb{R})$, $\alpha, a \in \mathbb{R}$, $\beta \in [0,\infty)$, $b \in (a,\infty)$, $r,T \in (0,\infty)$, $c \in \big[1/2,\infty\big)$, $v,\mathbf{v},w,\mathbf{w},z,\mathbf{z} \in [0,\infty)$, *let*

$$\begin{aligned}
C &= (4[7c + T + \sqrt{\mathbf{z}} + \sqrt{\mathbf{w}} + |a| + |b|]^{10+8(\mathbf{z}+\mathbf{w})} + 1)(1+r^2)\,,\\
p &= 2 + \max\{2z + 2\mathbf{z}\max\{\alpha,\beta+1\}, 2w + 1 + 2\mathbf{w}\max\{\alpha,\beta+1\}\}\,,\\
\mathcal{C} &= \tfrac{1}{2}[5c + T + \sqrt{\mathbf{v}} + |a| + |b|]^{-4\mathbf{v}-1}\,,\\
\mathbf{p} &= v + \mathbf{v}\max\{\alpha,\beta,1/2\}\,,
\end{aligned} \tag{245}$$

let $\|\cdot\|: \mathbb{R}^d \to [0,\infty)$ *be the standard norm on* $\mathbb{R}^d$, *let* $(\phi_\varepsilon)_{\varepsilon\in(0,r]} \subseteq \mathbf{N}$, *let* $A = (A_{i,j})_{(i,j)\in\{1,\dots,d\}^2} \in \mathbb{R}^{d\times d}$ *be a symmetric and positive semi-definite matrix, and assume for all* $\varepsilon \in (0,r]$, $x \in \mathbb{R}^d$ *that* $(\mathcal{R}\phi_\varepsilon) \in C(\mathbb{R}^d,\mathbb{R})$,

$$|(\mathcal{R}\phi_\varepsilon)(x)| \le cd^z(1+\|x\|^{\mathbf{z}}), \quad \|(\nabla(\mathcal{R}\phi_\varepsilon))(x)\| \le cd^w(1+\|x\|^{\mathbf{w}}), \tag{246}$$

$$|\varphi(x) - (\mathcal{R}\phi_\varepsilon)(x)| \le \varepsilon c d^v(1+\|x\|^{\mathbf{v}}), \qquad \sqrt{\operatorname{Trace}(A)} \le cd^\beta, \tag{247}$$

and $\|\mu\| \le cd^\alpha$. *Then*

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

$$(\tfrac{\partial}{\partial t}u)(t,x) = \textstyle\sum_{i,j=1}^d A_{i,j}\,(\tfrac{\partial^2}{\partial x_i \partial x_j}u)(t,x) + \sum_{i=1}^d \mu_i\,(\tfrac{\partial}{\partial x_i}u)(t,x) \tag{248}$$

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

(ii) there exists $(\psi_\varepsilon)_{\varepsilon\in(0,r]} \subseteq \mathbf{N}$ *such that for all* $\varepsilon \in (0,r]$ *it holds that* $\mathcal{N}(\psi_\varepsilon) \le C d^p \varepsilon^{-2} \mathcal{N}(\phi_{\varepsilon c^{-1} d^{-\mathbf{p}}\mathcal{C}})$, $\mathcal{L}(\psi_\varepsilon) = \mathcal{L}(\phi_{\varepsilon c^{-1} d^{-\mathbf{p}}\mathcal{C}})$, $\mathcal{P}(\psi_\varepsilon) \le C^2 d^{2p} \varepsilon^{-4} \mathcal{P}(\phi_{\varepsilon c^{-1} d^{-\mathbf{p}}\mathcal{C}})$, $\mathfrak{P}(\psi_\varepsilon) \le C d^p \varepsilon^{-2} \mathcal{P}(\phi_{\varepsilon c^{-1} d^{-\mathbf{p}}\mathcal{C}})$, $(\mathcal{R}\psi_\varepsilon) \in C(\mathbb{R}^d,\mathbb{R})$*, and*

$$\sup_{x\in[a,b]^d} |u(T,x) - (\mathcal{R}\psi_\varepsilon)(x)| \le \varepsilon. \tag{249}$$

Proof of Proposition 4.6. Throughout this proof let $\varepsilon \in (0,r]$, let

$$\mathbf{C} = 4[7c + T + \sqrt{\mathbf{z}} + \sqrt{\mathbf{w}} + |a| + |b|]^{10+8(\mathbf{z}+\mathbf{w})}\,, \tag{250}$$

let $n = \min(\mathbb{N} \cap [\mathbf{C} d^p \varepsilon^{-2}, \infty))$, let $\gamma_1,\dots,\gamma_n \in \mathbb{R}^{d\times d}$, $\delta_1,\dots,\delta_n \in \mathbb{R}^d$ satisfy

$$\sup_{x\in[a,b]^d} \left| u(T,x) - \left[\frac{1}{n} \sum_{k=1}^n (\mathcal{R}\phi_{\varepsilon c^{-1} d^{-\mathbf{p}}\mathcal{C}})(\gamma_k x + \delta_k) \right] \right| \le \varepsilon \tag{251}$$

(cf. item (ii) in Proposition 4.4), let $L \in \mathbb{N}\cap[2,\infty)$, $(l_0,\dots,l_L) \in (\{d\}\times(\mathbb{N}^{L-1})\times\{1\})$, $((W_1,B_1),\dots,(W_L,B_L)) \in (\times_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$ satisfy

$$\phi_{\varepsilon c^{-1} d^{-\mathbf{p}}\mathcal{C}} = ((W_1,B_1),\dots,(W_L,B_L)), \tag{252}$$

and let

$$\begin{aligned} \psi &= ((\mathcal{W}_1\mathcal{W}_0, \mathcal{W}_1\mathcal{B}_0 + \mathcal{B}_1), (\mathcal{W}_2,\mathcal{B}_2), \dots, (\mathcal{W}_{L-1},\mathcal{B}_{L-1}), (\mathcal{W}_L, B_L)) \\ &\in (\mathbb{R}^{nl_1\times l_0} \times \mathbb{R}^{nl_1}) \times (\times_{k=2}^{L-1} (\mathbb{R}^{nl_k\times nl_{k-1}} \times \mathbb{R}^{nl_k})) \times (\mathbb{R}^{l_L\times nl_{L-1}} \times \mathbb{R}^{l_L}) \end{aligned} \tag{253}$$

satisfy for all $k \in \{1,\dots,L-1\}$ that

$$\mathcal{W}_0 = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix}, \quad \mathcal{W}_k = \operatorname{diag}(W_k,\dots,W_k), \quad \mathcal{W}_L = \frac{1}{n}\begin{pmatrix} W_L & \dots & W_L \end{pmatrix}, \tag{254}$$

$$\mathcal{B}_0 = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_n \end{pmatrix}, \qquad \text{and} \qquad \mathcal{B}_k = \begin{pmatrix} B_k \\ \vdots \\ B_k \end{pmatrix}. \tag{255}$$

Observe that item (i) in Proposition 4.4 (with $r = r$, $\phi_\varepsilon = (\mathcal{R}\phi_\varepsilon)$ for $\varepsilon \in (0, r]$ in the notation of Proposition 4.4) establishes item (i). Next note that the fact that $n \in [\mathbf{C}d^p\varepsilon^{-2}, \mathbf{C}d^p\varepsilon^{-2} + 1]$ and the fact that $r^2\varepsilon^{-2} \in [1, \infty)$ prove that

$$\begin{aligned} n \le \mathbf{C}d^p\varepsilon^{-2} + 1 &\le (\mathbf{C}+1)d^p \max\{1, \varepsilon^{-2}\} \\ &\le (\mathbf{C}+1)\max\{1, r^{-2}\}r^2 d^p \varepsilon^{-2} \\ &\le C d^p \varepsilon^{-2}. \end{aligned} \tag{256}$$

This and (253) show that

$$\begin{aligned} \mathcal{N}(\psi) = l_0 + \sum_{k=1}^{L-1} n l_k + l_L &\le n \sum_{k=0}^{L} l_k \\ &= n\mathcal{N}(\phi_{\varepsilon c^{-1} d^{-\mathbf{p}} \mathcal{C}}) \le C d^p \varepsilon^{-2} \mathcal{N}(\phi_{\varepsilon c^{-1} d^{-\mathbf{p}} \mathcal{C}}). \end{aligned} \tag{257}$$

Next note that (253) and (256) demonstrate that

$$\begin{aligned} \mathcal{P}(\psi) &= n l_1 l_0 + n l_1 + \sum_{k=2}^{L-1} n l_k (n l_{k-1} + 1) + n l_L l_{L-1} + l_L \\ &\le n^2 \left[l_1(l_0+1) + \sum_{k=2}^{L-1} l_k(l_{k-1}+1) + l_L(l_{L-1}+1) \right] \\ &= n^2 \mathcal{P}(\phi_{\varepsilon c^{-1} d^{-\mathbf{p}} \mathcal{C}}) \le C^2 d^{2p} \varepsilon^{-4} \mathcal{P}(\phi_{\varepsilon c^{-1} d^{-\mathbf{p}} \mathcal{C}}). \end{aligned} \tag{258}$$

Moreover, note that (253) and (256) ensure that

$$\begin{aligned} \mathfrak{P}(\psi) &\le n l_1 (l_0 + 1) + n \sum_{k=2}^{L} l_k (l_{k-1} + 1) \\ &= n\mathcal{P}(\phi_{\varepsilon c^{-1} d^{-\mathbf{p}} \mathcal{C}}) \le C d^p \varepsilon^{-2} \mathcal{P}(\phi_{\varepsilon c^{-1} d^{-\mathbf{p}} \mathcal{C}}). \end{aligned} \tag{259}$$

Furthermore, (253), (254), and (255) imply that for all $x \in \mathbb{R}^d$ it holds that

$$(\mathcal{R}\psi)(x) = \frac{1}{n} \sum_{k=1}^{n} (\mathcal{R}\phi_{\varepsilon c^{-1} d^{-\mathbf{p}} \mathcal{C}})(\gamma_k x + \delta_k). \tag{260}$$

Combining this and (251) demonstrates that

$$\sup_{x \in [a,b]^d} |u(T,x) - (\mathcal{R}\psi)(x)| \le \varepsilon. \tag{261}$$

Next observe that (252) and (253) show that $\mathcal{L}(\psi) = L + 1 = \mathcal{L}(\phi_{\varepsilon c^{-1} d^{-\mathbf{p}} \mathcal{C}})$. Combining this with (257), (258), (259), and (261) establishes item (ii). The proof of Proposition 4.6 is thus completed. □

Corollary 4.7. *Assume Setting 4.5, let $\alpha, c, a \in \mathbb{R}$, $\beta \in [0, \infty)$, $b \in (a, \infty)$, $r, T \in (0, \infty)$, p, q, v, $\mathbf{v}$, w, $\mathbf{w}$, z, $\mathbf{z} \in [0, \infty)$, $\mathbf{p} = 2 + \max\{2z + 2\mathbf{z}\max\{\alpha, \beta +$*

$1\}, 2w+1+2\mathbf{w}\max\{\alpha,\beta+1\}\}$, $\mathfrak{p} = v + \mathbf{v}\max\{\alpha,\beta,1/2\}$, *for every* $d \in \mathbb{N}$ *let* $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0,\infty)$ *be the standard norm on* $\mathbb{R}^d$, *let* $(\phi_{\varepsilon,d})_{(\varepsilon,d)\in(0,r]\times\mathbb{N}} \subseteq \mathbf{N}$, *let* $\mu_d \in \mathbb{R}^d$, $d \in \mathbb{N}$, *let* $A_d = (A_d^{(i,j)})_{(i,j)\in\{1,\dots,d\}^2} \in \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, *be symmetric and positive semi-definite matrices, let* $\varphi_d \in C(\mathbb{R}^d,\mathbb{R})$, $d \in \mathbb{N}$, *and assume for all* $\varepsilon \in (0,r]$, $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ *that* $(\mathcal{R}\phi_{\varepsilon,d}) \in C(\mathbb{R}^d,\mathbb{R})$,

$$|(\mathcal{R}\phi_{\varepsilon,d})(x)| \le cd^z(1+\|x\|_{\mathbb{R}^d}^{\mathbf{z}}), \qquad \|(\nabla(\mathcal{R}\phi_{\varepsilon,d}))(x)\|_{\mathbb{R}^d} \le cd^w(1+\|x\|_{\mathbb{R}^d}^{\mathbf{w}}), \tag{262}$$

$$|\varphi_d(x) - (\mathcal{R}\phi_{\varepsilon,d})(x)| \le \varepsilon c d^v(1+\|x\|_{\mathbb{R}^d}^{\mathbf{v}}), \qquad \sqrt{\mathrm{Trace}(A_d)} \le cd^\beta, \tag{263}$$

$$\|\mu_d\|_{\mathbb{R}^d} \le cd^\alpha, \qquad \text{and} \qquad \mathcal{P}(\phi_{\varepsilon,d}) \le cd^p\varepsilon^{-q}. \tag{264}$$

Then

(i) there exist unique $u_d \in C([0,T]\times\mathbb{R}^d,\mathbb{R})$, $d \in \mathbb{N}$, *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* $\inf_{\gamma\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\big(\frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^\gamma}\big) < \infty$, *and which satisfy that for all* $d \in \mathbb{N}$ *it holds that* $u_d|_{(0,T)\times\mathbb{R}^d}$ *is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u_d)(t,x) = (\tfrac{\partial}{\partial x}u_d)(t,x)\,\mu_d + \textstyle\sum_{i,j=1}^d A_d^{(i,j)}\,(\tfrac{\partial^2}{\partial x_i\partial x_j}u_d)(t,x) \tag{265}$$

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

(ii) there exist $(\psi_{\varepsilon,d})_{(\varepsilon,d)\in(0,r]\times\mathbb{N}} \subseteq \mathbf{N}$, $C \in \mathbb{R}$ *such that for all* $\varepsilon \in (0,r]$, $d \in \mathbb{N}$ *it holds that* $\mathcal{N}(\psi_{\varepsilon,d}) \le Cd^{p+\mathbf{p}+\mathfrak{p}q}\varepsilon^{-(q+2)}$, $\mathcal{P}(\psi_{\varepsilon,d}) \le Cd^{p+2\mathbf{p}+\mathfrak{p}q}\varepsilon^{-(q+4)}$, $\mathfrak{P}(\psi_{\varepsilon,d}) \le Cd^{p+\mathbf{p}+\mathfrak{p}q}\varepsilon^{-(q+2)}$, $(\mathcal{R}\psi_{\varepsilon,d}) \in C(\mathbb{R}^d,\mathbb{R})$, *and*

$$\sup_{x\in[a,b]^d} |u_d(T,x) - (\mathcal{R}\psi_{\varepsilon,d})(x)| \le \varepsilon. \tag{266}$$

Proof of Corollary 4.7. Throughout this proof let $\mathbf{C} = (4[7\max\{1/2,c\} + T + \sqrt{\mathbf{z}} + \sqrt{\mathbf{w}} + |a| + |b|]^{10+8(\mathbf{z}+\mathbf{w})} + 1)(1+r^2)$, $\mathbf{c} = \max\{1/2,c\}$, $\mathcal{C} = \frac{1}{2}[5\mathbf{c} + T + \sqrt{\mathbf{v}} + |a| + |b|]^{-4\mathbf{v}-1}$. Note that item (i) in Proposition 4.6 (with $c = \max\{1/2,c\}$, $\mu = \mu_d$, $\varphi = \varphi_d$, $(\phi_\varepsilon)_{\varepsilon\in(0,r]} = (\phi_{\varepsilon,d})_{\varepsilon\in(0,r]}$, $A = A_d$, for $d \in \mathbb{N}$ in the notation of Proposition 4.6) implies that for every $d \in \mathbb{N}$ there exists a unique continuous function $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ which satisfies for all $x \in \mathbb{R}^d$ that $u_d(0,x) = \varphi_d(x)$, which satisfies that $\inf_{\gamma\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\big(\frac{|u_d(t,x)|}{1+\|x\|_{\mathbb{R}^d}^\gamma}\big) < \infty$, and which satisfies that $u_d|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u_d)(t,x) = (\tfrac{\partial}{\partial x}u_d)(t,x)\,\mu_d + \textstyle\sum_{i,j=1}^d A_d^{(i,j)}\,(\tfrac{\partial^2}{\partial x_i\partial x_j}u_d)(t,x) \tag{267}$$

for $(t,x) \in (0,T)\times\mathbb{R}^d$. This establishes item (i). Next note that for all $L \in \mathbb{N}\cap[2,\infty)$, $(l_0,\dots,l_L) \in \mathbb{N}^L\times\{1\}$, $\Phi \in (\times_{k=1}^L(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k}))$ it holds that

$$\mathcal{N}(\Phi) = \sum_{k=0}^L l_k \le l_1 l_0 + \sum_{k=1}^L l_k + \sum_{k=2}^L l_k l_{k-1} = \mathcal{P}(\Phi). \tag{268}$$

Moreover, note that Proposition 4.6 (with $c = \max\{1/2, c\}$, $\mu = \mu_d$, $\varphi = \varphi_d$, $(\phi_\varepsilon)_{\varepsilon\in(0,r]} = (\phi_{\varepsilon,d})_{\varepsilon\in(0,r]}$, $A = A_d$, for $d \in \mathbb{N}$ in the notation of Proposition 4.6) assures that there exist $\psi_{\varepsilon,d} \in \mathbf{N}$, $d \in \mathbb{N}$, $\varepsilon \in (0, r]$, which satisfy that for every $d \in \mathbb{N}$, $\varepsilon \in (0, r]$ it holds that $\mathcal{N}(\psi_{\varepsilon,d}) \leq \mathbf{C}d^{\mathbf{p}}\varepsilon^{-2}\mathcal{N}(\phi_{\varepsilon\mathbf{c}^{-1}d^{-\mathfrak{p}}\mathcal{C},d})$, $\mathcal{P}(\psi_{\varepsilon,d}) \leq \mathbf{C}^2 d^{2\mathbf{p}}\varepsilon^{-4}\mathcal{P}(\phi_{\varepsilon\mathbf{c}^{-1}d^{-\mathfrak{p}}\mathcal{C},d})$, $\mathfrak{P}(\psi_{\varepsilon,d}) \leq \mathbf{C}d^{\mathbf{p}}\varepsilon^{-2}\mathcal{P}(\phi_{\varepsilon\mathbf{c}^{-1}d^{-\mathfrak{p}}\mathcal{C},d})$, $(\mathcal{R}\psi_{\varepsilon,d}) \in C(\mathbb{R}^d, \mathbb{R})$, and

$$\sup_{x\in[a,b]^d} |u_d(T, x) - (\mathcal{R}\psi_{\varepsilon,d})(x)| \leq \varepsilon. \tag{269}$$

Next note that (264) implies that for all $\varepsilon \in (0, r], d \in \mathbb{N}$ it holds that

$$\mathcal{P}(\phi_{\varepsilon\mathbf{c}^{-1}d^{-\mathfrak{p}}\mathcal{C},d}) \leq cd^p(\varepsilon\mathbf{c}^{-1}d^{-\mathfrak{p}}\mathcal{C})^{-q} = c\mathbf{c}^q d^{p+\mathfrak{p}q}\mathcal{C}^{-q}\varepsilon^{-q}\,. \tag{270}$$

Combining this, (268), and (269) hence demonstrates that for all $d \in \mathbb{N}$, $\varepsilon \in (0, r]$ it holds that $\mathcal{N}(\psi_{\varepsilon,d}) \leq \mathbf{C}c\mathbf{c}^q d^{\mathbf{p}+p+\mathfrak{p}q}\mathcal{C}^{-q}\varepsilon^{-(q+2)}$, $\mathcal{P}(\psi_{\varepsilon,d}) \leq \mathbf{C}^2 c\mathbf{c}^q d^{2\mathbf{p}+p+\mathfrak{p}q}\mathcal{C}^{-q}\varepsilon^{-(q+4)}$, and $\mathfrak{P}(\psi_{\varepsilon,d}) \leq \mathbf{C}c\mathbf{c}^q d^{\mathbf{p}+p+\mathfrak{p}q}\mathcal{C}^{-q}\varepsilon^{-(q+2)}$. This and (269) establish item (ii). The proof of Corollary 4.7 is thus completed. □

5 Artifical neural network approximations for heat equations

5.1 Viscosity solutions for heat equations

In this subsection we establish in Lemma 5.3 below a well-known connection between viscosity solutions and classical solutions of heat equations with at most polynomially growing initial conditions. Lemma 5.3 will be employed in our proof of Theorem 5.4 below, the main result of this article. Lemma 5.3 is a simple consequence of Lemma 5.2 and the Feynman-Kac formula for viscosity solutions of Kolmogorov PDEs (cf., for example, Hairer et al. [31]). Lemma 5.2, in turn, is an elementary and well-known existence result for solutions of heat equations (cf., for example, Evans [21, Theorem 1 in Subsection 2.3.1]). For completeness we also provide in this subsection a detailed proof for Lemma 5.2. Our proof of Lemma 5.2 employs the elementary and well-known result in Lemma 5.1 below.

Lemma 5.1. *Let $p \in [0, \infty)$, $d \in \mathbb{N}$, and let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the standard norm on $\mathbb{R}^d$. Then it holds for all $t \in (0, \infty)$, $x \in \mathbb{R}^d$ that*

$$\int_{\mathbb{R}^d} \|y\|^p e^{-\frac{\|x-y\|^2}{4t}}\, dy < \infty. \tag{271}$$

Proof of Lemma 5.1. Throughout this proof let S be the set given by

$$S = \begin{cases} (-1, 1) & : d = 1 \\ (0, 2\pi) & : d = 2 \\ (0, 2\pi) \times (0, \pi)^{d-2} & : d \in \{3, 4, \dots\} \end{cases} \tag{272}$$

and for every $n \in \{2, 3, \ldots\}$ let $T_n \colon (0,\infty) \times (0, 2\pi) \times (0,\pi)^{n-2} \to \mathbb{R}$ be the function which satisfies for all $n \in \{2, 3, \ldots\}$, $r \in (0,\infty)$, $\varphi \in (0, 2\pi)$, $\vartheta_1, \ldots, \vartheta_{n-2} \in (0,\pi)$ that if $n = 2$ then $T_2(r, \varphi) = r$ and if $n \geq 3$ then

$$T_n(r, \varphi, \vartheta_1, \ldots, \vartheta_{n-2}) = r^{n-1} \left[\textstyle\prod_{i=1}^{n-2} [\sin(\vartheta_i)]^i \right]. \tag{273}$$

Observe that the integral transformation theorem with the diffeomorphism $(0,\infty) \ni r \mapsto \sqrt{r} \in (0,\infty)$ implies that

$$\int_0^\infty r^{p+d-1} e^{-r^2}\, dr = \int_0^\infty r^{(p+d-1)/2} e^{-r} \frac{1}{2r^{1/2}}\, dr = \frac{1}{2} \int_0^\infty r^{(p+d)/2-1} e^{-r}\, dr. \tag{274}$$

Item (iv) in Lemma 2.4 (with $x = \frac{p+d}{2}$ in the notation of Lemma 2.4) hence shows that

$$\int_0^\infty r^{p+d-1} e^{-r^2}\, dr \leq \frac{1}{2} \sqrt{\frac{4\pi}{p+d}} \left[\frac{p+d}{2e} \right]^{\frac{p+d}{2}} e^{\frac{1}{6(p+d)}} < \infty. \tag{275}$$

Next note that the integral transformation theorem with the diffeomorphism $\mathbb{R}^d \ni y \mapsto 2\sqrt{t}y \in \mathbb{R}^d$ for $t \in (0,\infty)$, the triangle inequality, and the fact that for all $a, b \in [0,\infty)$ it holds that $(a+b)^p \leq \max\{1, 2^{p-1}\}(a^p + b^p)$ ensure that for all $t \in (0,\infty)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
\int_{\mathbb{R}^d} \|y\|^p e^{-\frac{\|x-y\|^2}{4t}}\, dy &= \int_{\mathbb{R}^d} \|x-y\|^p e^{-\frac{\|y\|^2}{4t}}\, dy \\
&= \int_{\mathbb{R}^d} \|x - 2\sqrt{t}y\|^p e^{-\|y\|^2} (2\sqrt{t})^d\, dy \\
&\leq \max\{1, 2^{p-1}\} (2\sqrt{t})^d \|x\|^p \int_{\mathbb{R}^d} e^{-\|y\|^2}\, dy \\
&\quad + \max\{1, 2^{p-1}\} (2\sqrt{t})^{p+d} \int_{\mathbb{R}^d} \|y\|^p e^{-\|y\|^2}\, dy.
\end{aligned} \tag{276}$$

To establish (271) we distinguish between the case $d = 1$ and the case $d \in \mathbb{N} \cap [2,\infty)$. First, we consider the case $d = 1$. Note that

$$\int_{\mathbb{R}^d} \|y\|^p e^{-\|y\|^2}\, dy = 2 \int_0^\infty y^p e^{-y^2}\, dy. \tag{277}$$

Combining this with (275) and (276) establishes (271) in the case $d = 1$. Next we consider the case $d \in \{2, 3, \ldots\}$. Note that (272), (273), item (iii) in Lemma 2.6, and Fubini's theorem ensure that

$$\begin{aligned}
\int_{\mathbb{R}^d} \|y\|^p e^{-\|y\|^2}\, dy &= \int_0^\infty \int_S r^{p+d-1} e^{-r^2} T_d(1, \phi)\, d\phi\, dr \\
&= \int_S \int_0^\infty r^{p+d-1} e^{-r^2}\, dr\, T_d(1, \phi)\, d\phi \\
&\leq 2\pi^{d-1} \int_0^\infty r^{p+d-1} e^{-r^2}\, dr.
\end{aligned} \tag{278}$$

Combining this with (275) and (276) establishes (271) in the case $d \in \{2, 3, \ldots\}$. The proof of Lemma 5.1 is thus completed. □

Lemma 5.2. *Let $d \in \mathbb{N}$, $\varphi \in C(\mathbb{R}^d, \mathbb{R})$, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the standard norm on $\mathbb{R}^d$, assume that $\inf_{\gamma \in (0,\infty)} \sup_{x \in \mathbb{R}^d} \big(\frac{|\varphi(x)|}{1+\|x\|^\gamma}\big) < \infty$, and let $\Phi \colon (0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ be the function which satisfies for all $t \in (0, \infty)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ that*

$$\Phi(t, x) = \int_{\mathbb{R}^d} \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{(x_1-y_1)^2+\ldots+(x_d-y_d)^2}{4t}} \varphi(y)\, dy. \tag{279}$$

Then it holds for all $t \in (0, \infty)$, $x \in \mathbb{R}^d$ that $\Phi \in C^{1,2}((0, \infty) \times \mathbb{R}^d, \mathbb{R})$ and

$$(\tfrac{\partial}{\partial t}\Phi)(t, x) = (\Delta_x \Phi)(t, x). \tag{280}$$

Proof of Lemma 5.2. Throughout this proof let $\rho \colon (0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ be the function which satisfies for all $t \in (0, \infty)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ that

$$\rho(t, x) = \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{x_1^2+\ldots+x_d^2}{4t}}. \tag{281}$$

Observe that for all $t \in (0, \infty)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$(\tfrac{\partial}{\partial t}\rho)(t, x) = \left[\frac{x_1^2 + \ldots + x_d^2}{4t^2} - \frac{d}{2t}\right] \rho(t, x). \tag{282}$$

Next note that for all $i \in \{1, \ldots, d\}$, $t \in (0, \infty)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$(\tfrac{\partial}{\partial x_i}\rho)(t, x) = -\frac{x_i}{2t}\rho(t, x). \tag{283}$$

This implies that for all $i, j \in \{1, \ldots, d\}$, $t \in (0, \infty)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$(\tfrac{\partial^2}{\partial x_i \partial x_j}\rho)(t, x) = \begin{cases} \Big[\frac{x_i^2}{4t^2} - \frac{1}{2t}\Big]\rho(t, x) & : i = j \\ \frac{x_i x_j}{4t^2}\rho(t, x) & : i \neq j. \end{cases} \tag{284}$$

Hence, we obtain that for all $t \in (0, \infty)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$(\Delta_x \rho)(t, x) = \sum_{i=1}^{d} (\tfrac{\partial^2}{\partial x_i^2}\rho)(t, x) = \left[\frac{x_1^2 + \ldots + x_d^2}{4t^2} - \frac{d}{2t}\right] \rho(t, x). \tag{285}$$

Combining this with (282) demonstrates that for all $t \in (0, \infty)$, $x \in \mathbb{R}^d$ it holds that

$$(\tfrac{\partial}{\partial t}\rho)(t, x) - (\Delta_x \rho)(t, x) = 0. \tag{286}$$

Next note that the hypothesis that $\inf_{\gamma \in (0,\infty)} \sup_{x \in \mathbb{R}^d} \big(\frac{|\varphi(x)|}{1+\|x\|^\gamma}\big) < \infty$ ensures that there exist $\gamma \in (0, \infty)$, $C \in \mathbb{R}$ which satisfy that for all $x \in \mathbb{R}^d$ it holds that

$$|\varphi(x)| \leq C(1 + \|x\|^\gamma). \tag{287}$$

This and Lemma 5.1 show that for all $t \in (0, \infty)$, $x \in \mathbb{R}^d$ it holds that

$$|\Phi(t,x)| \leq \int_{\mathbb{R}^d} |\rho(t, x-y)\varphi(y)| \, dy \leq C \int_{\mathbb{R}^d} \rho(t, x-y)(1 + \|y\|^\gamma) \, dy < \infty. \quad (288)$$

Next note that (282), (287), the triangle inequality, and Lemma 5.1 demonstrate that for all $t \in (0, \infty)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\int_{\mathbb{R}^d} \big|(\tfrac{\partial}{\partial t}\rho)(t, x-y)\varphi(y)\big| \, dy \\ &\leq C \int_{\mathbb{R}^d} \left[\frac{\|x-y\|^2}{4t^2} + \frac{d}{2t} \right] \rho(t, x-y)(1 + \|y\|^\gamma) \, dy < \infty. \end{aligned} \quad (289)$$

Combining this, (288), and (282) with Amann & Escher [2, Ch. X, Theorem 3.18] shows that for all $t \in (0, \infty)$, $x \in \mathbb{R}^d$ it holds that $\Phi \in C^{1,0}((0, \infty) \times \mathbb{R}^d, \mathbb{R})$ and

$$(\tfrac{\partial}{\partial t}\Phi)(t,x) = \int_{\mathbb{R}^d} (\tfrac{\partial}{\partial t}\rho)(t, x-y)\varphi(y) \, dy. \quad (290)$$

Next observe that (283), (287), and Lemma 5.1 ensure that for all $i \in \{1, \ldots, d\}$, $t \in (0, \infty)$, $x \in \mathbb{R}^d$ it holds that

$$\int_{\mathbb{R}^d} \big|(\tfrac{\partial}{\partial x_i}\rho)(t, x-y)\varphi(y)\big| \, dy \leq \frac{C}{2t} \int_{\mathbb{R}^d} \|x-y\| \rho(t, x-y)(1 + \|y\|^\gamma) \, dy < \infty. \quad (291)$$

Combining this, (288), and (283) with (290) and Amann & Escher [2, Ch. X, Theorem 3.18] shows that for all $i \in \{1, \ldots, d\}$, $t \in (0, \infty)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that $\Phi \in C^{1,1}((0, \infty) \times \mathbb{R}^d, \mathbb{R})$ and

$$(\tfrac{\partial}{\partial x_i}\Phi)(t,x) = \int_{\mathbb{R}^d} (\tfrac{\partial}{\partial x_i}\rho)(t, x-y)\varphi(y) \, dy. \quad (292)$$

Next note that (284), (287), the fact that for all $a, b \in \mathbb{R}$ it holds that $ab \leq a^2 + b^2$, and Lemma 5.1 ensure that for all $i, j \in \{1, \ldots, d\}$, $t \in (0, \infty)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\int_{\mathbb{R}^d} \big|(\tfrac{\partial^2}{\partial x_i \partial x_j}\rho)(t, x-y)\varphi(y)\big| \, dy \\ &\leq C \int_{\mathbb{R}^d} \left[\frac{\|x-y\|^2}{4t^2} + \frac{1}{2t} \right] \rho(t, x-y)(1 + \|y\|^\gamma) \, dy < \infty. \end{aligned} \quad (293)$$

Combining this, (291), and (284) with (292) and Amann & Escher [2, Ch. X, Theorem 3.18] shows that for all $i, j \in \{1, \ldots, d\}$, $t \in (0, \infty)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that $\Phi \in C^{1,2}((0, \infty) \times \mathbb{R}^d, \mathbb{R})$ and

$$(\tfrac{\partial^2}{\partial x_i \partial x_j}\Phi)(t,x) = \int_{\mathbb{R}^d} (\tfrac{\partial^2}{\partial x_i \partial x_j}\rho)(t, x-y)\varphi(y) \, dy. \quad (294)$$

Hence, we obtain that for all $t \in (0,\infty)$, $x \in \mathbb{R}^d$ it holds that

$$(\Delta_x \Phi)(t,x) = \sum_{i=1}^{d} (\tfrac{\partial^2}{\partial x_i^2} \Phi)(t,x) = \int_{\mathbb{R}^d} \left[\sum_{i=1}^{d} (\tfrac{\partial^2}{\partial x_i^2} \rho)(t, x-y)\varphi(y) \right] dy = \int_{\mathbb{R}^d} (\Delta_x \rho)(t,x-y)\varphi(y)\, dy. \tag{295}$$

Combining this and (290) with (286) establishes (280). The proof of Lemma 5.2 is thus completed. □

Lemma 5.3. *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, $\varphi \in C(\mathbb{R}^d,\mathbb{R})$, $u \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be a norm, assume for all $x \in \mathbb{R}^d$ that $u(0,x) = \varphi(x)$, assume that $\inf_{\gamma\in(0,\infty)} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \big(\frac{|u(t,x)|}{1+\|x\|^\gamma}\big) < \infty$, and assume that $u|_{(0,T)\times\mathbb{R}^d}$ is a viscosity solution of*

$$(\tfrac{\partial}{\partial t} u)(t,x) = (\Delta_x u)(t,x) \tag{296}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$. Then it holds for all $t \in (0,T]$, $x \in \mathbb{R}^d$ that $u|_{(0,T]\times\mathbb{R}^d} \in C^{1,2}((0,T] \times \mathbb{R}^d, \mathbb{R})$ and

$$(\tfrac{\partial}{\partial t} u)(t,x) = (\Delta_x u)(t,x). \tag{297}$$

Proof of Lemma 5.3. Throughout this proof let $\|\cdot\|_2 : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a normal filtration $(\mathbb{F}_t)_{t\in[0,T]}$, and let $W : [0,T] \times \Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion. Observe that there exist $C \in \mathbb{R}$, $c \in (0,\infty)$ such that for all $x \in \mathbb{R}^d$ it holds that

$$c\|x\|_2 \le \|x\| \le C\|x\|_2. \tag{298}$$

This and the fact that $\inf_{\gamma\in(0,\infty)} \sup_{x\in\mathbb{R}^d} \big(\frac{|\varphi(x)|}{1+\|x\|^\gamma}\big) < \infty$ show that

$$\inf_{\gamma\in(0,\infty)} \sup_{x\in\mathbb{R}^d} \left(\frac{|\varphi(x)|}{1+\|x\|_2^\gamma} \right) < \infty. \tag{299}$$

Hence, we obtain that $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ is an at most polynomially growing function. The Feynman-Kac formula (cf., for example, Grohs et al. [28, Proposition 2.22(iii)] and Hairer et al. [31, Corollary 4.17]) hence ensures that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$u(t,x) = \mathbb{E}[\varphi(x + \sqrt{2} W_t)]. \tag{300}$$

Next note that the fact that for all $t \in (0,T]$ it holds that W_t is a $\mathcal{N}_{0,tI_{\mathbb{R}^d}}$ distributed random variable implies that for all $t \in (0,T]$, $x \in \mathbb{R}^d$ it holds that

$x + \sqrt{2}W_t$ is a $\mathcal{N}_{x,2tI_{\mathbb{R}^d}}$ distributed random variable. Combining this with (300) demonstrates that for all $t \in (0,T]$, $x \in \mathbb{R}^d$ it holds that

$$u(t,x) = \int_{\mathbb{R}^d} \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{(x_1-y_1)^2+\ldots+(x_d-y_d)^2}{4t}} \varphi(y)\, dy. \tag{301}$$

Lemma 5.2 hence proves that for all $t \in (0,T]$, $x \in \mathbb{R}^d$ it holds that $u|_{(0,T]\times\mathbb{R}^d} \in C^{1,2}((0,T] \times \mathbb{R}^d, \mathbb{R})$ and

$$(\tfrac{\partial}{\partial t} u)(t,x) = (\Delta_x u)(t,x). \tag{302}$$

The proof of Lemma 5.3 is thus completed. □

5.2 Qualitative error estimates for heat equations

In this subsection we establish in Theorem 5.4 below the main result of this article. Theorem 5.4 proves that ANNs do not suffer from the curse of dimensionality in the uniform numerical approximation of heat equations. Corollary 5.5 below specializes Theorem 5.4 to the case in which the constants $c \in \mathbb{R}$, $p, q, v, \mathbf{v}, w, \mathbf{w}, z, \mathbf{z} \in [0,\infty)$, which are used to formulate the hypotheses in (303)–(304) below, all coincide.

Theorem 5.4. *Assume Setting 4.5, let $c, a \in \mathbb{R}$, $b \in (a,\infty)$, $r, T \in (0,\infty)$, $p, q, v, \mathbf{v}, w, \mathbf{w}, z, \mathbf{z} \in [0,\infty)$, $\mathbf{p} = 2 + \max\{2z + 3\mathbf{z}, 2w + 3\mathbf{w} + 1\}$, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $\varphi_d \in C(\mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, let $(\phi_{\varepsilon,d})_{(\varepsilon,d)\in(0,r]\times\mathbb{N}} \subseteq \mathbf{N}$, and assume for all $\varepsilon \in (0,r]$, $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $(\mathcal{R}\phi_{\varepsilon,d}) \in C(\mathbb{R}^d, \mathbb{R})$,*

$$|(\mathcal{R}\phi_{\varepsilon,d})(x)| \leq cd^z(1 + \|x\|_{\mathbb{R}^d}^{\mathbf{z}}), \qquad \|(\nabla(\mathcal{R}\phi_{\varepsilon,d}))(x)\|_{\mathbb{R}^d} \leq cd^w(1 + \|x\|_{\mathbb{R}^d}^{\mathbf{w}}), \tag{303}$$

$$|\varphi_d(x) - (\mathcal{R}\phi_{\varepsilon,d})(x)| \leq \varepsilon c d^v(1 + \|x\|_{\mathbb{R}^d}^{\mathbf{v}}), \quad \text{and} \quad \mathcal{P}(\phi_{\varepsilon,d}) \leq cd^p \varepsilon^{-q}. \tag{304}$$

Then

(i) there exist unique $u_d \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, which satisfy for all $d \in \mathbb{N}$, $t \in (0,T]$, $x \in \mathbb{R}^d$ that $u_d|_{(0,T]\times\mathbb{R}^d} \in C^{1,2}((0,T] \times \mathbb{R}^d, \mathbb{R})$, $u_d(0,x) = \varphi_d(x)$, $\inf_{\gamma\in(0,\infty)} \sup_{(s,y)\in[0,T]\times\mathbb{R}^d} \big(\frac{|u_d(s,y)|}{1+\|y\|_{\mathbb{R}^d}^{\gamma}}\big) < \infty$, and

$$(\tfrac{\partial}{\partial t} u_d)(t,x) = (\Delta_x u_d)(t,x) \tag{305}$$

and

(ii) there exist $(\psi_{\varepsilon,d})_{(\varepsilon,d)\in(0,r]\times\mathbb{N}} \subseteq \mathbf{N}$, $C \in \mathbb{R}$ such that for all $\varepsilon \in (0,r]$, $d \in \mathbb{N}$ it holds that $(\mathcal{R}\psi_{\varepsilon,d}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{N}(\psi_{\varepsilon,d}) \leq Cd^{p+\mathbf{p}+q(v+\frac{1}{2}\mathbf{v})}\varepsilon^{-(q+2)}$, $\mathcal{P}(\psi_{\varepsilon,d}) \leq Cd^{p+2\mathbf{p}+q(v+\frac{1}{2}\mathbf{v})}\varepsilon^{-(q+4)}$, $\mathfrak{P}(\psi_{\varepsilon,d}) \leq Cd^{p+\mathbf{p}+q(v+\frac{1}{2}\mathbf{v})}\varepsilon^{-(q+2)}$, and

$$\sup_{x\in[a,b]^d} |u_d(T,x) - (\mathcal{R}\psi_{\varepsilon,d})(x)| \leq \varepsilon. \tag{306}$$

Proof of Theorem 5.4. First, observe that for all $d \in \mathbb{N}$ it holds that

$$\sqrt{\operatorname{Trace}(I_{\mathbb{R}^d})} = \left[\sum_{i=1}^{d} 1\right]^{1/2} = d^{1/2} \leq \max\{1, c\} d^{1/2}. \tag{307}$$

Corollary 4.7 (with $\alpha = 0$, $\beta = \frac{1}{2}$, $c = \max\{1, c\}$, $\mu_d = 0$, $A_d = I_{\mathbb{R}^d}$ for $d \in \mathbb{N}$ in the notation of Corollary 4.7) hence implies that there exist unique $u_d \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, 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 $\inf_{\gamma \in (0,\infty)} \sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \big(\frac{|u_d(t,x)|}{1 + \|x\|_{\mathbb{R}^d}^{\gamma}}\big) < \infty$, and which satisfy that for all $d \in \mathbb{N}$ it holds that $u_d|_{(0,T) \times \mathbb{R}^d}$ is a viscosity solution of

$$(\tfrac{\partial}{\partial t} u_d)(t,x) = (\Delta_x u_d)(t,x) \tag{308}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$ and there exist $(\psi_{\varepsilon,d})_{(\varepsilon,d) \in (0,r] \times \mathbb{N}} \subseteq \mathbf{N}$, $C \in \mathbb{R}$ such that for all $\varepsilon \in (0,r]$, $d \in \mathbb{N}$ it holds that $\mathcal{N}(\psi_{\varepsilon,d}) \leq C d^{p + \mathbf{p} + q(v + \frac{1}{2}\mathbf{v})} \varepsilon^{-(q+2)}$, $\mathcal{P}(\psi_{\varepsilon,d}) \leq C d^{p + 2\mathbf{p} + q(v + \frac{1}{2}\mathbf{v})} \varepsilon^{-(q+4)}$, $\mathfrak{P}(\psi_{\varepsilon,d}) \leq C d^{p + \mathbf{p} + q(v + \frac{1}{2}\mathbf{v})} \varepsilon^{-(q+2)}$, $(\mathcal{R}\psi_{\varepsilon,d}) \in C(\mathbb{R}^d, \mathbb{R})$, and

$$\sup_{x \in [a,b]^d} |u_d(T,x) - (\mathcal{R}\psi_{\varepsilon,d})(x)| \leq \varepsilon. \tag{309}$$

This proves item (ii). Next note that (303) and (304) ensure that for all $d \in \mathbb{N}$ it holds that $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$ is an at most polynomially growing function. Hence, we obtain that for all $d \in \mathbb{N}$ it holds that

$$\inf_{\gamma \in (0,\infty)} \sup_{x \in \mathbb{R}^d} \left(\frac{|\varphi_d(x)|}{1 + \|x\|_{\mathbb{R}^d}^{\gamma}}\right) < \infty. \tag{310}$$

Lemma 5.3 (with $T = T$, $\varphi = \varphi_d$, $u = u_d$ for $d \in \mathbb{N}$ in the notation of Lemma 5.3) hence shows that for all $d \in \mathbb{N}$, $t \in (0,T]$, $x \in \mathbb{R}^d$ it holds that $u_d|_{(0,T] \times \mathbb{R}^d} \in C^{1,2}((0,T] \times \mathbb{R}^d, \mathbb{R})$ and

$$(\tfrac{\partial}{\partial t} u_d)(t,x) = (\Delta_x u_d)(t,x). \tag{311}$$

This, (308), and Hairer et al. [31, Remark 4.1]) prove item (i). The proof of Theorem 5.4 is thus completed. □

Corollary 5.5. *Assume Setting 4.5, let $a \in \mathbb{R}$, $b \in (a,\infty)$, $c, T \in (0,\infty)$, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0,\infty)$ be the standard norm on $\mathbb{R}^d$, let $\varphi_d \in C(\mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, let $(\phi_{\varepsilon,d})_{(\varepsilon,d) \in (0,1] \times \mathbb{N}} \subseteq \mathbf{N}$, and assume for all $\varepsilon \in (0,1]$, $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $(\mathcal{R}\phi_{\varepsilon,d}) \in C(\mathbb{R}^d, \mathbb{R})$,*

$$\mathcal{P}(\phi_{\varepsilon,d}) \leq c d^c \varepsilon^{-c}, \quad |\varphi_d(x) - (\mathcal{R}\phi_{\varepsilon,d})(x)| \leq \varepsilon c d^c (1 + \|x\|_{\mathbb{R}^d}^c), \tag{312}$$

and

$$|(\mathcal{R}\phi_{\varepsilon,d})(x)| + \|(\nabla(\mathcal{R}\phi_{\varepsilon,d}))(x)\|_{\mathbb{R}^d} \leq c d^c (1 + \|x\|_{\mathbb{R}^d}^c). \tag{313}$$

Then

(i) *there exist unique* $u_d \in C([0,T]\times\mathbb{R}^d,\mathbb{R})$, $d \in \mathbb{N}$, *which satisfy for all* $d \in \mathbb{N}$, $t \in (0,T]$, $x \in \mathbb{R}^d$ *that* $u_d|_{(0,T]\times\mathbb{R}^d} \in C^{1,2}((0,T]\times\mathbb{R}^d,\mathbb{R})$, $u_d(0,x) = \varphi_d(x)$, $\inf_{\gamma\in(0,\infty)} \sup_{(s,y)\in[0,T]\times\mathbb{R}^d} \big(\frac{|u_d(s,y)|}{1+\|y\|^{\gamma}_{\mathbb{R}^d}}\big) < \infty$, *and*

$$(\tfrac{\partial}{\partial t} u_d)(t,x) = (\Delta_x u_d)(t,x) \tag{314}$$

and

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

$$\sup_{x\in[a,b]^d} |u_d(T,x) - (\mathcal{R}\psi_{\varepsilon,d})(x)| \leq \varepsilon. \tag{315}$$

Proof of Corollary 5.5. First, observe that (312), (313), and item (i) in Theorem 5.4 (with $r = 1$, $c = c$, $p = c$, $q = c$, $v = c$, $\mathbf{v} = c$, $w = c$, $\mathbf{w} = c$, $z = c$, $\mathbf{z} = c$ in the notation of Theorem 5.4) establish item (i). Next note that (312), (313), and item (ii) in Theorem 5.4 (with $r = 1$, $c = c$, $p = c$, $q = c$, $v = c$, $\mathbf{v} = c$, $w = c$, $\mathbf{w} = c$, $z = c$, $\mathbf{z} = c$ in the notation of Theorem 5.4) ensure that there exist $(\psi_{\varepsilon,d})_{(\varepsilon,d)\in(0,1]\times\mathbb{N}} \subseteq \mathbf{N}$, $C \in \mathbb{R}$ which satisfy that for all $\varepsilon \in (0,1]$, $d \in \mathbb{N}$ it holds that $\mathcal{P}(\psi_{\varepsilon,d}) \leq C d^{\frac{3}{2}c^2+11c+6}\varepsilon^{-(c+4)}$, $(\mathcal{R}\psi_{\varepsilon,d}) \in C(\mathbb{R}^d,\mathbb{R})$, and

$$\sup_{x\in[a,b]^d} |u_d(T,x) - (\mathcal{R}\psi_{\varepsilon,d})(x)| \leq \varepsilon. \tag{316}$$

Hence, we obtain that for all $\varepsilon \in (0,1]$, $d \in \mathbb{N}$ it holds that

$$\mathcal{P}(\psi_{\varepsilon,d}) \leq \max\{C, \tfrac{3}{2}c^2 + 11c + 6\} d^{\max\{C,\frac{3}{2}c^2+11c+6\}} \varepsilon^{-\max\{C,\frac{3}{2}c^2+11c+6\}}. \tag{317}$$

Combining this with (316) establishes item (ii). The proof of Corollary 5.5 is thus completed. □

5.3 ANN approximations for geometric Brownian motions

In this subsection we specialize Theorem 5.4 above in Corollary 5.6 below to an example in which the activation function is the softplus function ($\mathbb{R} \ni x \mapsto \ln(1 + e^x) \in (0,\infty)$).

Corollary 5.6. *Let* $c, a \in \mathbb{R}$, $b \in (a,\infty)$, $p \in [0,\infty)$, $T \in (0,\infty)$, *for every* $d \in \mathbb{N}$ *let* $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0,\infty)$ *be the standard norm on* $\mathbb{R}^d$, *let* $\mathbf{N}$ *be the set given by*

$$\mathbf{N} = \cup_{L\in\mathbb{N}\cap[2,\infty)} \cup_{l_0,\ldots,l_L\in\mathbb{N}} \big(\times_{k=1}^{L} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\big), \tag{318}$$

let $\mathbf{A}_d \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, *satisfy for all* $d \in \mathbb{N}$, $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$ *that* $\mathbf{A}_d(x) = (\ln(1 + e^{x_1}),\ldots,\ln(1 + e^{x_d}))$, *let* $\mathcal{P} \colon \mathbf{N} \to \mathbb{N}$ *and* $\mathcal{R} \colon \mathbf{N} \to \cup_{m,n\in\mathbb{N}} C(\mathbb{R}^m,\mathbb{R}^n)$ *be the functions which satisfy that for all* $L \in \mathbb{N} \cap [2,\infty)$, $l_0,\ldots,l_L \in \mathbb{N}$, $\Phi = ((W_1,B_1),\ldots,(W_L,B_L)) \in (\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $x_0 \in$

$\mathbb{R}^{l_0}$, $\ldots, x_{L-1} \in \mathbb{R}^{l_{L-1}}$ with $\forall\, k \in \mathbb{N} \cap (0,L)\colon x_k = \mathbf{A}_{l_k}(W_k x_{k-1} + B_k)$ it holds that $\mathcal{P}(\Phi) = \sum_{k=1}^{L} l_k(l_{k-1}+1)$, $(\mathcal{R}\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$, and

$$(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L, \tag{319}$$

and let $(K_d)_{d\in\mathbb{N}} \subseteq \mathbb{R}$ satisfy for all $d \in \mathbb{N}$ that $|K_d| \leq cd^p$. Then

(i) there exist unique $u_d \in C([0,T]\times\mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, which satisfy for all $d \in \mathbb{N}$, $t \in (0,T]$, $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$ that $u_d|_{(0,T]\times\mathbb{R}^d} \in C^{1,2}((0,T]\times\mathbb{R}^d, \mathbb{R})$, $u_d(0,x) = \ln(1+e^{x_1+\ldots+x_d-K_d})+K_d$, $\inf_{\gamma\in(0,\infty)} \sup_{(s,y)\in[0,T]\times\mathbb{R}^d} \big(\frac{|u_d(s,y)|}{1+\|y\|_{\mathbb{R}^d}^{\gamma}}\big) < \infty$, and

$$(\tfrac{\partial}{\partial t} u_d)(t,x) = (\Delta_x u_d)(t,x) \tag{320}$$

and

(ii) there exist $(\psi_{\varepsilon,d})_{(\varepsilon,d)\in(0,1]\times\mathbb{N}} \subseteq \mathbf{N}$, $\kappa \in \mathbb{R}$ such that for all $\varepsilon \in (0,1]$, $d \in \mathbb{N}$ it holds that $\mathcal{P}(\psi_{\varepsilon,d}) \leq \kappa d^{11+4\max\{p,1/2\}} \varepsilon^{-4}$, $(\mathcal{R}\psi_{\varepsilon,d}) \in C(\mathbb{R}^d,\mathbb{R})$, and

$$\sup_{x\in[a,b]^d} |u_d(T,x) - (\mathcal{R}\psi_{\varepsilon,d})(x)| \leq \varepsilon. \tag{321}$$

Proof of Corollary 5.6. Throughout this proof let $\varphi_d\colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, be the functions which satisfy for all $d \in \mathbb{N}$, $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$ that

$$\varphi_d(x) = \ln(1 + e^{x_1+\ldots+x_d-K_d}) + K_d \tag{322}$$

and let $(\phi_d)_{d\in\mathbb{N}} \subseteq \mathbf{N}$ satisfy for all $d \in \mathbb{N}$ that

$$\phi_d = (((1,\ldots,1), -K_d), (1, K_d)) \in (\mathbb{R}^{1\times d} \times \mathbb{R}) \times (\mathbb{R}\times\mathbb{R}). \tag{323}$$

Observe that (323) assures that for all $d \in \mathbb{N}$ it holds that

$$\mathcal{P}(\phi_d) = 1(d+1) + 1(1+1) = d+3 \leq 4d \leq \max\{4,c\}d. \tag{324}$$

Next note that the fact that $(\mathbb{R} \ni x \mapsto \ln(1+e^x) \in \mathbb{R}) \in C^1(\mathbb{R},\mathbb{R})$, (319), and (323) imply that for all $d \in \mathbb{N}$, $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$ it holds that $(\mathcal{R}\phi_d) \in C(\mathbb{R}^d,\mathbb{R})$ and

$$(\mathcal{R}\phi_d)(x) = \ln(1 + e^{x_1+\ldots+x_d-K_d}) + K_d = \varphi_d(x). \tag{325}$$

Next note that for all $d \in \mathbb{N}$ it holds that $\ln(1+e^{-K_d}) \leq \ln 2 + |K_d|$ and for any $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $\|(\nabla\varphi_d)(x)\|_{\mathbb{R}^d} \leq d^{1/2}$. This and the hypothesis that for all $d \in \mathbb{N}$ it holds that $|K_d| \leq cd^p$ hence demonstrate that for all $d \in \mathbb{N}$, $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
|(\mathcal{R}\phi_d)(x)| = |\varphi_d(x)| &\leq \sup_{y\in\mathbb{R}^d} [\|(\nabla\varphi_d)(y)\|_{\mathbb{R}^d}\|x\|_{\mathbb{R}^d} + |\varphi_d(0)|] \\
&\leq d^{1/2}\|x\|_{\mathbb{R}^d} + \ln 2 + 2|K_d| \leq d^{1/2}\|x\|_{\mathbb{R}^d} + 1 + 2cd^p \\
&\leq \max\{1,2c\}(1 + d^p + d^{1/2}\|x\|_{\mathbb{R}^d}) \\
&\leq 2\max\{1,2c\}d^{\max\{p,1/2\}}(1 + \|x\|_{\mathbb{R}^d}).
\end{aligned} \tag{326}$$

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

$$\left\| \big(\nabla(\mathcal{R}\phi_d)\big)(x) \right\|_{\mathbb{R}^d} = \|(\nabla\varphi_d)(x)\|_{\mathbb{R}^d} \leq d^{1/2} \leq 2\max\{1, 2c\} d^{1/2}(1 + \|x\|_{\mathbb{R}^d}^0). \tag{327}$$

Combining this, (324), (325), (326), and the fact that $(\mathbb{R} \ni x \mapsto \ln(1+e^x) \in \mathbb{R}) \in C^1(\mathbb{R}, \mathbb{R})$ with Theorem 5.4 (with $c = \max\{4, 4c\}$, $r = 1$, $p = 1$, $q = 0$, $v = 0$, $\mathbf{v} = 0$, $w = 1/2$, $\mathbf{w} = 0$, $z = \max\{p, 1/2\}$, $\mathbf{z} = 1$, $\mathbf{a} = \big(\mathbb{R} \ni x \mapsto \ln(1+e^x) \in \mathbb{R}\big)$ in the notation of Theorem 5.4) establishes items (i)–(ii). The proof of Corollary 5.6 is thus completed. □

Acknowledgements

This project is based on the master thesis of DK written from March 2018 to September 2018 at ETH Zurich under the supervision of AJ.

References

[1] Aliprantis, C. D., and Border, K. C. *Infinite dimensional analysis.* Springer, Berlin, 2006.

[2] Amann, H., and Escher, J. *Analysis. III.* Birkhäuser, Basel, 2009.

[3] Artin, E. *The gamma function.* Holt, Rinehart and Winston, New York-Toronto-London, 1964.

[4] Beck, C., Becker, S., Cheridito, P., Jentzen, A., and Neufeld, A. Deep splitting method for parabolic PDEs. *arXiv:1907.03452* (2019), 40 pages.

[5] 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.

[6] Beck, C., E, W., and Jentzen, A. Machine Learning Approximation Algorithms for High-Dimensional Fully Nonlinear Partial Differential Equations and Second-order Backward Stochastic Differential Equations. *J. Nonlinear Sci. 29*, 4 (2019), 1563–1619.

[7] Becker, S., Cheridito, P., and Jentzen, A. Deep optimal stopping. *Journal of Machine Learning Research 20*, 74 (2019), 1–25.

[8] Becker, S., Cheridito, P., Jentzen, A., and Welti, T. Solving high-dimensional optimal stopping problems using deep learning. *arXiv:1908.01602* (2019), 42 pages.

[9] Bellman, R. *Dynamic programming.* Princeton University Press, Princeton, NJ, 1957.

[10] Berg, J., and Nyström, K. A unified deep artificial neural network approach to partial differential equations in complex geometries. *Neurocomputing 317* (2018), 28–41.

[11] Berner, J., Grohs, P., and Jentzen, A. Analysis of the generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *Revision requested from SIAM Journal on Mathematics of Data Science; arXiv:1809.03062* (2018), 35 pages.

[12] Buehler, H., Gonon, L., Teichmann, J., and Wood, B. Deep hedging. *Quantitative Finance 19*, 8 (2019), 1271–1291.

[13] Chan-Wai-Nam, Q., Mikael, J., and Warin, X. Machine learning for semi linear PDEs. *J. Sci. Comput. 79*, 3 (2019), 1667–1712.

[14] Chen, Y., and Wan, J. W. L. Deep neural network framework based on backward stochastic differential equations for pricing and hedging american options in high dimensions. *arXiv:1909.11532* (2019), 35 pages.

[15] Cohn, D. L. *Measure theory.* Birkhäuser, New York, 2013.

[16] 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. *Revision requested from IMA. Journal on Numerical Analysis; arXiv:1605.00856* (2017), 48 pages.

[17] Dockhorn, T. A discussion on solving partial differential equations using neural networks. *arXiv:1904.07200* (2019), 9 pages.

[18] E, W., Han, J., and Jentzen, A. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in Mathematics and Statistics 5*, 4 (2017), 349–380.

[19] 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.

[20] Elbrächter, D., Grohs, P., Jentzen, A., and Schwab, C. DNN Expression Rate Analysis of High-dimensional PDEs: Application to Option Pricing. *arXiv:1809.07669* (2018), 50 pages.

[21] Evans, L. C. *Partial differential equations*, second ed., vol. 19 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2010.

[22] Farahmand, A.-m., Nabi, S., and Nikovski, D. Deep reinforcement learning for partial differential equation control. *2017 American Control Conference (ACC)* (2017), 3120–3127.

[23] Fujii, M., Takahashi, A., and Takahashi, M. Asymptotic expansion as prior knowledge in deep learning method for high dimensional BSDEs. *Asia-Pacific Financial Markets* (Mar 2019).

[24] Garling, D. J. H. *Inequalities: a journey into linear analysis.* Cambridge University Press, Cambridge, 2007.

[25] Gilbarg, D., and Trudinger, N. S. *Elliptic partial differential equations of second order.* Springer, Berlin, 2001. Reprint of the 1998 edition.

[26] Goudenège, L., Molent, A., and Zanette, A. Machine Learning for Pricing American Options in High Dimension. *arXiv:1903.11275* (2019), 11 pages.

[27] Graves, A., Mohamed, A.-r., and Hinton, G. Speech recognition with deep recurrent neural networks. In *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing, ICASSP* (2013), pp. 6645–6649.

[28] Grohs, P., Hornung, F., Jentzen, A., and von Wurstemberger, P. A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *To appear in Memoires of the American Mathematical Society; arXiv:1809.02362* (2018), 124 pages.

[29] Grohs, P., Hornung, F., Jentzen, A., and Zimmermann, P. Space-time error estimates for deep neural network approximations for differential equations. *arXiv:1908.03833* (2019), 86 pages.

[30] Grohs, P., Jentzen, A., and Salimova, D. Deep neural network approximations for Monte Carlo algorithms. *arXiv:1908.10828* (2019), 45 pages.

[31] Hairer, M., Hutzenthaler, M., and Jentzen, A. Loss of regularity for Kolmogorov equations. *Annals of Probabability 43*, 2 (2015), 468–527.

[32] Han, J., Jentzen, A., and E, W. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences 115*, 34 (2018), 8505–8510.

[33] Han, J., and Long, J. Convergence of the Deep BSDE Method for Coupled FBSDEs. *arXiv:1811.01165* (2018), 26 pages.

[34] Henry-Labordère, P. Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM. (November 15, 2017), 16 pages. Available at SSRN: `https://ssrn.com/abstract=3071506`.

[35] Hinton, G., Deng, L., Yu, D., Dahl, G., rahman Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., and Kingsbury, B. Deep neural networks for acoustic modeling in speech recognition. *Signal Processing Magazine* (2012).

[36] Huang, G., Liu, Z., and Weinberger, K. Q. Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 2261–2269.

[37] Huré, C., Pham, H., and Warin, X. Some machine learning schemes for high-dimensional nonlinear PDEs. *arXiv:1902.01599* (2019), 33 pages.

[38] Hutzenthaler, M., Jentzen, A., Kruse, T., and Nguyen, T. A. A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations. *Revision requested from SN Partial Differential Equations and Applications; arXiv:1901.10854* (2019), 29 pages.

[39] Hutzenthaler, M., Jentzen, A., and Wang, X. Exponential integrability properties of numerical approximation processes for nonlinear stochastic differential equations. *Mathematics of Computation 87*, 311 (2018), 1353–1413.

[40] Hytönen, T., van Neerven, J., Veraar, M., and Weis, L. *Analysis in Banach spaces. Vol. II.* Springer, 2017.

[41] Jacquier, A., and Oumgari, M. Deep PPDEs for rough local stochastic volatility. *arXiv:1906.02551* (2019), 21 pages.

[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] Jentzen, A., Salimova, D., and Welti, T. A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. *arXiv:1809.07321* (2018), 48 pages.

[44] Karatzas, I., and Shreve, S. E. *Brownian motion and stochastic calculus*, second ed. Springer, New York, 1991.

[45] Klenke, A. *Probability theory*. Springer, 2014.

[46] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[47] Kutyniok, G., Petersen, P., Raslan, M., and Schneider, R. A theoretical analysis of deep neural networks and parametric PDEs. *arXiv:1904.00377* (2019), 43 pages.

[48] Long, Z., Lu, Y., Ma, X., and Dong, B. PDE-Net: Learning PDEs from Data. In *Proceedings of the 35th International Conference on Machine Learning* (2018), pp. 3208–3216.

[49] Lye, K. O., Mishra, S., and Ray, D. Deep learning observables in computational fluid dynamics. *arXiv:1903.03040* (2019), 57 pages.

[50] Magill, M., Qureshi, F., and de Haan, H. W. Neural networks trained to solve differential equations learn general representations. In *Advances in Neural Information Processing Systems* (2018), pp. 4071–4081.

[51] Mizuguchi, M., Tanaka, K., Sekine, K., and Oishi, S. Estimation of Sobolev embedding constant on a domain dividable into bounded convex domains. *Journal of Inequalities and Applications* (2017).

[52] Petersen, P., and Voigtlaender, F. Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *arXiv:1709.05289* (2017), 54 pages.

[53] Pham, H., and Warin, X. Neural networks-based backward scheme for fully nonlinear PDEs. *arXiv:1908.00412* (2019), 15 pages.

[54] Raissi, M. Deep hidden physics models: Deep learning of nonlinear partial differential equations. *J. Mach. Learn. Res. 19* (2018), 25:1–25:24.

[55] Reisinger, C., and Zhang, Y. Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems. *arXiv:1903.06652* (2019), 34 pages.

[56] Rockafellar, R. T. *Convex Analysis*. Princeton University Press, 1970.

[57] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. Mastering the game of go with deep neural networks and tree search. *Nature 529* (Jan. 2016), 484.

[58] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., and Hassabis, D. Mastering the game of go without human knowledge. *Nature 550* (Oct. 2017), 354.

[59] Simonyan, K., and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556* (2014).

[60] Sirignano, J., and Cont, R. Universal features of price formation in financial markets: perspectives from deep learning. *Quantitative Finance 19*, 9 (2019), 1449–1459.

[61] Sirignano, J., and Spiliopoulos, K. DGM: A deep learning algorithm for solving partial differential equations. *J. Comput. Phys. 375* (2018), 1339–1364.

[62] Wu, C., Karanasou, P., Gales, M. J., and Sim, K. C. Stimulated deep neural network for speech recognition. In *Interspeech 2016* (2016), pp. 400–404.