

Solving stochastic differential equations and  
Kolmogorov equations  
by means of deep learning

Ch. Beck and S. Becker and Ph. Grohs and A. Jentzen

Research Report No. 2018-21  
June 2018

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

---

# Solving stochastic differential equations and Kolmogorov equations by means of deep learning

Christian Beck<sup>1</sup>, Sebastian Becker<sup>2</sup>, Philipp Grohs<sup>3</sup>,  
Nor Jaafari<sup>4</sup>, and Arnulf Jentzen<sup>5</sup>

<sup>1</sup> Department of Mathematics, ETH Zurich,  
Zurich, Switzerland, e-mail: christian.beck@math.ethz.ch

<sup>2</sup> ZENAI AG, Zurich, Switzerland, e-mail: sebastian.becker@zenai.ch

<sup>3</sup> Faculty of Mathematics and Research Platform Data Science,  
University of Vienna, Vienna, Austria, e-mail: philipp.grohs@univie.ac.at

<sup>4</sup> ZENAI AG, Zurich, Switzerland, e-mail: nor.jaafari@zenai.ch

<sup>5</sup> Department of Mathematics, ETH Zurich, Zurich,  
Switzerland, e-mail: arnulf.jentzen@sam.math.ethz.ch

June 11, 2018

## Abstract

Stochastic differential equations (SDEs) and the Kolmogorov partial differential equations (PDEs) associated to them have been widely used in models from engineering, finance, and the natural sciences. In particular, SDEs and Kolmogorov PDEs, respectively, are highly employed in models for the approximative pricing of financial derivatives. Kolmogorov PDEs and SDEs, respectively, can typically not be solved explicitly and it has been and still is an active topic of research to design and analyze numerical methods which are able to approximately solve Kolmogorov PDEs and SDEs, respectively. Nearly all approximation methods for Kolmogorov PDEs in the literature suffer under the curse of dimensionality or only provide approximations of the solution of the PDE at a single fixed space-time point. In this paper we derive and propose a numerical approximation method which aims to overcome both of the above mentioned drawbacks and intends to deliver a numerical approximation of the Kolmogorov PDE on an entire region  $[a, b]^d$  without suffering from the curse of dimensionality. Numerical results on examples including the heat equation, the Black-Scholes model, the stochastic Lorenz equation, and the Heston model suggest that the proposed approximation algorithm is quite effective in high dimensions in terms of both accuracy and speed.

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Derivation and description of the proposed approximation algorithm</b>	<b>3</b>
2.1	Kolmogorov partial differential equations (PDEs)	3
2.2	On stochastic differential equations and Kolmogorov PDEs	3
2.3	Formulation as minimization problem	4
2.4	Discretization of the stochastic differential equation	21
2.5	Deep artificial neural network approximations	22
2.6	Stochastic gradient descent-type minimization	23
2.7	Description of the algorithm in a special case	24
2.8	Description of the algorithm in the general case	25
<b>3</b>	<b>Examples</b>	<b>26</b>
3.1	Setting	27
3.2	Heat equation	27
3.3	Geometric Brownian motions	35
3.4	Black-Scholes model with correlated noise	37
3.5	Stochastic Lorenz equations	39
3.6	Heston model	40
<b>4</b>	<b>PYTHON source codes</b>	<b>42</b>
4.1	PYTHON source code for the algorithm	42
4.2	A PYTHON source code associated to Subsection 3.2	45
4.3	A PYTHON source code associated to Subsection 3.3	46
4.4	A PYTHON source code associated to Subsection 3.4	47
4.5	A PYTHON source code associated to Subsection 3.5	48
4.6	A PYTHON source code associated to Subsection 3.6	50

## 1 Introduction

Stochastic differential equations (SDEs) and the Kolmogorov partial differential equations (PDEs) associated to them have been widely used in models from engineering, finance, and the natural sciences. In particular, SDEs and Kolmogorov PDEs, respectively, are highly employed in models for the approximative pricing of financial derivatives. Kolmogorov PDEs and SDEs, respectively, can typically not be solved explicitly and it has been and still is an active topic of research to design and analyze numerical methods which are able to approximately solve Kolmogorov PDEs and SDEs, respectively (see, e.g., [17], [20], [21], [27], [28], [29], [32], [40], [41], [42], [44], [45], [46], [47], [49], [52]). In particular, there are nowadays several different types of numerical approximation methods for Kolmogorov

PDEs in the literature including deterministic numerical approximation methods such as finite differences based approximation methods (cf., for example, [7], [8], [23], [43], [58], [56]) and finite elements based approximation methods (cf., for example, [9], [10], [59]) as well as random numerical approximation methods based on Monte Carlo methods (cf., for example, [17], [20]) and discretizations of the underlying SDEs (cf., for example, [21], [27], [28], [29], [32], [40], [41], [42], [44], [45], [46], [47], [49], [52]). The above mentioned deterministic approximation methods for PDEs work quite efficiently in one or two space dimensions but cannot be used in the case of high-dimensional PDEs as they suffer from the so-called curse of dimensionality (cf. Bellman [5]) in the sense that the computational effort of the considered approximation algorithm grows exponentially in the PDE dimension. The above mentioned random numerical approximation methods involving Monte Carlo approximations typically overcome this curse of dimensionality but only provide approximations of the Kolmogorov PDE at a single fixed space-time point.

The key contribution of this paper is to derive and propose a numerical approximation method which aims to overcome both of the above mentioned drawbacks and intends to deliver a numerical approximation of the Kolmogorov PDE on an entire region  $[a, b]^d$  without suffering from the curse of dimensionality. The numerical scheme, which we propose in this work, is inspired by recently developed deep learning based approximation algorithms for PDEs in the literature (cf., for example, [3], [4], [14], [15], [16], [24], [26], [51], [57]). To derive the proposed approximation scheme we first reformulate the considered Kolmogorov PDE as a suitable infinite dimensional stochastic optimization problem (see items (ii)–(iii) in Proposition 2.7 below for details). This infinite dimensional stochastic optimization problem is then temporally discretized by means of suitable discretizations of the underlying SDE and it is spatially discretized by means of fully connected deep artificial neural network approximations (see (107) in Subsection 2.6 as well as Subsections 2.4–2.5 below). The resulting finite dimensional stochastic optimization problem is then solved by means of stochastic gradient descent type optimization algorithms (see (109) in Subsection 2.6, Framework 2.9 in Subsection 2.7, Framework 2.10 in Subsection 2.8, as well as (124)–(125) in Subsection 3.1). We test the proposed approximation method numerically in the case of several examples of SDEs and PDEs, respectively (see Subsections 3.2–3.6 below for details). The obtained numerical results indicate that the proposed approximation algorithm is quite effective in high dimensions in terms of both accuracy and speed.

The remainder of this article is organized as follows. In Section 2 we derive the proposed approximation algorithm (see Subsections 2.1–2.6 below) and we present a detailed description of the proposed approximation algorithm in a special case (see Subsection 2.7 below) as well as in the general case (see Subsection 2.8 below). In Section 3 we test the proposed algorithm numerically in the case of several examples of SDEs and PDEs, respectively. The employed source codes for the numerical simulations in Section 3 are postponed to Section 4.

## 2 Derivation and description of the proposed approximation algorithm

In this section we describe the approximation problem which we intend to solve (see Subsection 2.1 below) and we derive (see Subsections 2.2–2.6 below) and specify (see Subsections 2.7–2.8 below) the numerical scheme which we suggest to use to solve this approximation problem (cf., for example, E et al. [14], Han et al. [24], Sirignano & Spiliopoulos [57], Beck et al. [3], Fujii, Takahashi, A., & Takahashi, M. [16], and Henry-Labordere [26] for related derivations and related approximation schemes).

### 2.1 Kolmogorov partial differential equations (PDEs)

Let  $T \in (0, \infty)$ ,  $d \in \mathbb{N}$ , let  $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  be Lipschitz continuous functions, let  $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}$  be a function, and let  $u = (u(t, x))_{(t,x) \in [0,T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  be a function with at most polynomially growing partial derivatives which satisfies for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  that  $u(0, x) = \varphi(x)$  and

$$\frac{\partial u}{\partial t}(t, x) = \frac{1}{2} \text{Trace}_{\mathbb{R}^d}(\sigma(x)[\sigma(x)]^*(\text{Hess}_x u)(t, x)) + \langle \mu(x), (\nabla_x u)(t, x) \rangle_{\mathbb{R}^d}. \quad (1)$$

Our goal is to approximately calculate the function  $\mathbb{R}^d \ni x \mapsto u(T, x) \in \mathbb{R}$  on some subset of  $\mathbb{R}^d$ . To fix ideas we consider real numbers  $a, b \in \mathbb{R}$  with  $a < b$  and we suppose that our goal is to approximately calculate the function  $[a, b]^d \ni x \mapsto u(T, x) \in \mathbb{R}$ .

### 2.2 On stochastic differential equations and Kolmogorov PDEs

In this subsection we provide a probabilistic representation for the solutions of the PDE (1), that is, we recall the classical Feynman-Kac formula for the PDE (1) (cf., for example, Øksendal [50, Chapter 8]).

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space with a normal filtration  $(\mathbb{F}_t)_{t \in [0, T]}$ , let  $W: [0, T] \times \Omega \rightarrow \mathbb{R}^d$  be a standard  $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t \in [0, T]})$ -Brownian motion, and for every  $x \in \mathbb{R}^d$  let  $X^x = (X_t^x)_{t \in [0, T]}: [0, T] \times \Omega \rightarrow \mathbb{R}^d$  be an  $(\mathbb{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous sample paths which satisfies that for every  $t \in [0, T]$  it holds  $\mathbb{P}$ -a.s. that

$$X_t^x = x + \int_0^t \mu(X_s^x) ds + \int_0^t \sigma(X_s^x) dW_s. \quad (2)$$

The Feynman-Kac formula (cf., for example, Hairer et al. [22, Corollary 4.17 and Remark 4.1]) and (1) hence yield that for every  $x \in \mathbb{R}^d$  it holds that

$$u(T, x) = \mathbb{E}[u(0, X_T^x)] = \mathbb{E}[\varphi(X_T^x)]. \quad (3)$$

## 2.3 Formulation as minimization problem

In the next step we exploit (3) to formulate a minimization problem which is uniquely solved by the function  $[a, b]^d \ni x \mapsto u(T, x) \in \mathbb{R}$  (cf. (1) above). For this we first recall the  $L^2$ -minimization property of the expectation of a real-valued random variable (see Lemma 2.1 below). Then we extend this minimization result to certain random fields (see Proposition 2.2 below). Thereafter, we apply Proposition 2.2 to random fields in the context of the Feynman-Kac representation (3) to obtain Proposition 2.7 below. Proposition 2.7 provides a minimization problem (see, for instance, (93) below) which has the function  $[a, b]^d \ni x \mapsto u(T, x) \in \mathbb{R}$  as the unique global minimizer.

Our proof of Proposition 2.7 is based on the elementary auxiliary results in Lemmas 2.3–2.6. For completeness we also present the proofs of Lemmas 2.3–2.6 here. The statement and the proof of Lemma 2.3 are based on the proof of Da Prato & Zabczyk [13, Lemma 1.1].

**Lemma 2.1.** *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and let  $X: \Omega \rightarrow \mathbb{R}$  be an  $\mathcal{F}/\mathcal{B}(\mathbb{R})$ -measurable random variable which satisfies  $\mathbb{E}[|X|^2] < \infty$ . Then*

(i) *it holds for every  $y \in \mathbb{R}$  that*

$$\mathbb{E}[|X - y|^2] = \mathbb{E}[|X - \mathbb{E}[X]|^2] + |\mathbb{E}[X] - y|^2, \quad (4)$$

(ii) *it holds that there exists a unique real number  $z \in \mathbb{R}$  such that*

$$\mathbb{E}[|X - z|^2] = \inf_{y \in \mathbb{R}} \mathbb{E}[|X - y|^2], \quad (5)$$

and

(iii) *it holds that*

$$\mathbb{E}[|X - \mathbb{E}[X]|^2] = \inf_{y \in \mathbb{R}} \mathbb{E}[|X - y|^2]. \quad (6)$$

*Proof of Lemma 2.1.* Observe that the fact that  $\mathbb{E}[|X|^2] < \infty$  ensures that for every  $y \in \mathbb{R}$  it holds that

$$\begin{aligned} \mathbb{E}[|X - y|^2] &= \mathbb{E}[|X - \mathbb{E}[X] + \mathbb{E}[X] - y|^2] \\ &= \mathbb{E}[|X - \mathbb{E}[X]|^2 + 2(X - \mathbb{E}[X])(\mathbb{E}[X] - y) + |\mathbb{E}[X] - y|^2] \\ &= \mathbb{E}[|X - \mathbb{E}[X]|^2] + 2(\mathbb{E}[X] - y)\mathbb{E}[X - \mathbb{E}[X]] + |\mathbb{E}[X] - y|^2 \\ &= \mathbb{E}[|X - \mathbb{E}[X]|^2] + |\mathbb{E}[X] - y|^2. \end{aligned} \quad (7)$$

This establishes item (i). Item (ii) and item (iii) are immediate consequences of item (i). The proof of Lemma 2.1 is thus completed.  $\square$

**Proposition 2.2.** *Let  $a \in \mathbb{R}$ ,  $b \in (a, \infty)$ , let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $X = (X_x)_{x \in [a, b]^d} : [a, b]^d \times \Omega \rightarrow \mathbb{R}$  be a  $(\mathcal{B}([a, b]^d) \otimes \mathcal{F})/\mathcal{B}(\mathbb{R})$ -measurable function, assume for every  $x \in [a, b]^d$  that  $\mathbb{E}[|X_x|^2] < \infty$ , and assume that the function  $[a, b]^d \ni x \mapsto \mathbb{E}[X_x] \in \mathbb{R}$  is continuous. Then*

(i) *it holds that there exists a unique continuous function  $u : [a, b]^d \rightarrow \mathbb{R}$  such that*

$$\int_{[a, b]^d} \mathbb{E}[|X_x - u(x)|^2] dx = \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \int_{[a, b]^d} \mathbb{E}[|X_x - v(x)|^2] dx \right) \quad (8)$$

and

(ii) *it holds for every  $x \in [a, b]^d$  that  $u(x) = \mathbb{E}[X_x]$ .*

*Proof of Proposition 2.2.* Observe that item (i) in Lemma 2.1 and the hypothesis that  $\forall x \in [a, b]^d : \mathbb{E}[|X_x|^2] < \infty$  ensure that for every function  $u : [a, b]^d \rightarrow \mathbb{R}$  and every  $x \in [a, b]^d$  it holds that

$$\mathbb{E}[|X_x - u(x)|^2] = \mathbb{E}[|X_x - \mathbb{E}[X_x]|^2] + |\mathbb{E}[X_x] - u(x)|^2. \quad (9)$$

Fubini's theorem (see, e.g., Klenke [39, Theorem 14.16]) hence proves that for every continuous function  $u : [a, b]^d \rightarrow \mathbb{R}$  it holds that

$$\int_{[a, b]^d} \mathbb{E}[|X_x - u(x)|^2] dx = \int_{[a, b]^d} \mathbb{E}[|X_x - \mathbb{E}[X_x]|^2] dx + \int_{[a, b]^d} |\mathbb{E}[X_x] - u(x)|^2 dx. \quad (10)$$

The hypothesis that the function  $[a, b]^d \ni x \mapsto \mathbb{E}[X_x] \in \mathbb{R}$  is continuous therefore demonstrates that

$$\begin{aligned} & \int_{[a, b]^d} \mathbb{E}[|X_x - \mathbb{E}[X_x]|^2] dx \\ & \geq \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \int_{[a, b]^d} \mathbb{E}[|X_x - v(x)|^2] dx \right) \\ & = \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \int_{[a, b]^d} \mathbb{E}[|X_x - \mathbb{E}[X_x]|^2] dx + \int_{[a, b]^d} |\mathbb{E}[X_x] - v(x)|^2 dx \right) \\ & \geq \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \int_{[a, b]^d} \mathbb{E}[|X_x - \mathbb{E}[X_x]|^2] dx \right) \\ & = \int_{[a, b]^d} \mathbb{E}[|X_x - \mathbb{E}[X_x]|^2] dx. \end{aligned} \quad (11)$$

Hence, we obtain that

$$\int_{[a, b]^d} \mathbb{E}[|X_x - \mathbb{E}[X_x]|^2] dx = \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \int_{[a, b]^d} \mathbb{E}[|X_x - v(x)|^2] dx \right). \quad (12)$$

Again the fact that the function  $[a, b]^d \ni x \mapsto \mathbb{E}[X_x] \in \mathbb{R}$  is continuous therefore proves that there exists a continuous function  $u: [a, b]^d \rightarrow \mathbb{R}$  such that

$$\int_{[a, b]^d} \mathbb{E}[|X_x - u(x)|^2] dx = \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \int_{[a, b]^d} \mathbb{E}[|X_x - v(x)|^2] dx \right). \quad (13)$$

Next observe that (10) and (12) yield that for every continuous function  $u: [a, b]^d \rightarrow \mathbb{R}$  with

$$\int_{[a, b]^d} \mathbb{E}[|X_x - u(x)|^2] dx = \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \int_{[a, b]^d} \mathbb{E}[|X_x - v(x)|^2] dx \right) \quad (14)$$

it holds that

$$\begin{aligned} & \int_{[a, b]^d} \mathbb{E}[|X_x - \mathbb{E}[X_x]|^2] dx \\ &= \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \int_{[a, b]^d} \mathbb{E}[|X_x - v(x)|^2] dx \right) = \int_{[a, b]^d} \mathbb{E}[|X_x - u(x)|^2] dx \\ &= \int_{[a, b]^d} \mathbb{E}[|X_x - \mathbb{E}[X_x]|^2] dx + \int_{[a, b]^d} |\mathbb{E}[X_x] - u(x)|^2 dx. \end{aligned} \quad (15)$$

Hence, we obtain that for every continuous function  $u: [a, b]^d \rightarrow \mathbb{R}$  with

$$\int_{[a, b]^d} \mathbb{E}[|X_x - u(x)|^2] dx = \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \int_{[a, b]^d} \mathbb{E}[|X_x - v(x)|^2] dx \right) \quad (16)$$

it holds that

$$\int_{[a, b]^d} |\mathbb{E}[X_x] - u(x)|^2 dx = 0. \quad (17)$$

This and again the hypothesis that the function  $[a, b]^d \ni x \mapsto \mathbb{E}[X_x] \in \mathbb{R}$  is continuous yield that for every continuous function  $u: [a, b]^d \rightarrow \mathbb{R}$  with

$$\int_{[a, b]^d} \mathbb{E}[|X_x - u(x)|^2] dx = \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \int_{[a, b]^d} \mathbb{E}[|X_x - v(x)|^2] dx \right) \quad (18)$$

and every  $x \in [a, b]^d$  it holds that  $u(x) = \mathbb{E}[X_x]$ . Combining this with (13) completes the proof of Proposition 2.2.  $\square$

**Lemma 2.3** (Projections in metric spaces). *Let  $(E, d)$  be a metric space, let  $n \in \mathbb{N}$ ,  $e_1, e_2, \dots, e_n \in E$ , and let  $P: E \rightarrow E$  be the function which satisfies for every  $x \in E$  that*

$$P(x) = e_{\min\{k \in \{1, 2, \dots, n\} : d(x, e_k) = \min\{d(x, e_1), d(x, e_2), \dots, d(x, e_n)\}\}}. \quad (19)$$

*Then*



(i) it holds for every  $x \in E$  that

$$d(x, P(x)) = \min_{k \in \{1, 2, \dots, n\}} d(x, e_k) \quad (20)$$

and

(ii) it holds for every  $A \subseteq E$  that  $P^{-1}(A) \in \mathcal{B}(E)$ .

*Proof of Lemma 2.3.* Throughout this proof let  $D = (D_1, \dots, D_n): E \rightarrow \mathbb{R}^n$  be the function which satisfies for every  $x \in E$  that

$$D(x) = (D_1(x), D_2(x), \dots, D_n(x)) = (d(x, e_1), d(x, e_2), \dots, d(x, e_n)). \quad (21)$$

Note that (19) ensures that for every  $x \in E$  it holds that

$$d(x, P(x)) = d(x, e_{\min\{k \in \{1, 2, \dots, n\}: d(x, e_k) = \min\{d(x, e_1), d(x, e_2), \dots, d(x, e_n)\}\}}) = \min_{k \in \{1, 2, \dots, n\}} d(x, e_k). \quad (22)$$

This establishes item (i). It thus remains to prove item (ii). For this observe that the fact that the function  $d: E \times E \rightarrow [0, \infty)$  is continuous ensures that the function  $D: E \rightarrow \mathbb{R}^n$  is continuous. Hence, we obtain that the function  $D: E \rightarrow \mathbb{R}^n$  is  $\mathcal{B}(E)/\mathcal{B}(\mathbb{R}^n)$ -measurable. Next note that item (i) demonstrates that for every  $k \in \{1, 2, \dots, n\}$ ,  $x \in P^{-1}(\{e_k\})$  it holds that

$$d(x, e_k) = d(x, P(x)) = \min_{l \in \{1, 2, \dots, n\}} d(x, e_l). \quad (23)$$

Hence, we obtain that for every  $k \in \{1, 2, \dots, n\}$ ,  $x \in P^{-1}(\{e_k\})$  it holds that

$$k \geq \min\{l \in \{1, 2, \dots, n\}: d(x, e_l) = \min\{d(x, e_1), d(x, e_2), \dots, d(x, e_n)\}\}. \quad (24)$$

Moreover, note that (19) ensures that for every  $k \in \{1, 2, \dots, n\}$ ,  $x \in P^{-1}(\{e_k\})$  it holds that

$$\begin{aligned} & \min\left\{l \in \{1, 2, \dots, n\}: d(x, e_l) = \min_{u \in \{1, 2, \dots, n\}} d(x, e_u)\right\} \\ & \in \{l \in \{1, 2, \dots, n\}: e_l = e_k\} \subseteq \{k, k+1, \dots, n\}. \end{aligned} \quad (25)$$

Therefore, we obtain that for every  $k \in \{1, 2, \dots, n\}$ ,  $x \in P^{-1}(\{e_k\})$  with  $e_k \notin (\cup_{l \in \mathbb{N} \cap [0, k)} \{e_l\})$  it holds that

$$\min\left\{l \in \{1, 2, \dots, n\}: d(x, e_l) = \min_{u \in \{1, 2, \dots, n\}} d(x, e_u)\right\} \geq k. \quad (26)$$

Combining this with (24) yields that for every  $k \in \{1, 2, \dots, n\}$ ,  $x \in P^{-1}(\{e_k\})$  with  $e_k \notin (\cup_{l \in \mathbb{N} \cap [0, k)} \{e_l\})$  it holds that

$$\min\left\{l \in \{1, 2, \dots, n\}: d(x, e_l) = \min_{u \in \{1, 2, \dots, n\}} d(x, e_u)\right\} = k. \quad (27)$$

Hence, we obtain that for every  $k \in \{1, 2, \dots, n\}$  with  $e_k \notin (\cup_{l \in \mathbb{N} \cap [0, k)} \{e_l\})$  it holds that

$$P^{-1}(\{e_k\}) \subseteq \left\{ x \in E : \min \left\{ l \in \{1, 2, \dots, n\} : d(x, e_l) = \min_{u \in \{1, 2, \dots, n\}} d(x, e_u) \right\} = k \right\}. \quad (28)$$

This and (19) show that for every  $k \in \{1, 2, \dots, n\}$  with  $e_k \notin (\cup_{l \in \mathbb{N} \cap [0, k)} \{e_l\})$  it holds that

$$P^{-1}(\{e_k\}) = \left\{ x \in E : \min \left\{ l \in \{1, 2, \dots, n\} : d(x, e_l) = \min_{u \in \{1, 2, \dots, n\}} d(x, e_u) \right\} = k \right\}. \quad (29)$$

Combining (21) with the fact that the function  $D: E \rightarrow \mathbb{R}^n$  is  $\mathcal{B}(E)/\mathcal{B}(\mathbb{R}^n)$ -measurable therefore demonstrates that for every  $k \in \{1, 2, \dots, n\}$  with  $e_k \notin (\cup_{l \in \mathbb{N} \cap [0, k)} \{e_l\})$  it holds that

$$\begin{aligned} & P^{-1}(\{e_k\}) \\ &= \left\{ x \in E : \min \left\{ l \in \{1, 2, \dots, n\} : d(x, e_l) = \min_{u \in \{1, 2, \dots, n\}} d(x, e_u) \right\} = k \right\} \\ &= \left\{ x \in E : \min \left\{ l \in \{1, 2, \dots, n\} : D_l(x) = \min_{u \in \{1, 2, \dots, n\}} D_u(x) \right\} = k \right\} \\ &= \left\{ x \in E : \left( \begin{array}{l} \forall l \in \mathbb{N} \cap [0, k): D_k(x) < D_l(x) \text{ and} \\ \forall l \in \{1, 2, \dots, n\}: D_k(x) \leq D_l(x) \end{array} \right) \right\} \\ &= \left[ \bigcap_{l=1}^{k-1} \underbrace{\{x \in E : D_k(x) < D_l(x)\}}_{\in \mathcal{B}(E)} \right] \cap \left[ \bigcap_{l=1}^n \underbrace{\{x \in E : D_k(x) \leq D_l(x)\}}_{\in \mathcal{B}(E)} \right] \in \mathcal{B}(E). \end{aligned} \quad (30)$$

Hence, we obtain that for every  $f \in \{e_1, e_2, \dots, e_n\}$  it holds that

$$P^{-1}(\{f\}) \in \mathcal{B}(E). \quad (31)$$

Therefore, we obtain that for every  $A \subseteq E$  it holds that

$$\begin{aligned} P^{-1}(A) &= P^{-1}(A \cap \{e_1, e_2, \dots, e_n\}) \\ &= \cup_{f \in A \cap \{e_1, e_2, \dots, e_n\}} \underbrace{P^{-1}(\{f\})}_{\in \mathcal{B}(E)} \in \mathcal{B}(E). \end{aligned} \quad (32)$$

This establishes item (ii). The proof of Lemma 2.3 is thus completed.  $\square$

**Lemma 2.4.** *Let  $(E, d)$  be a separable metric space, let  $(\mathcal{E}, \delta)$  be a metric space, let  $(\Omega, \mathcal{F})$  be a measurable space, let  $X: E \times \Omega \rightarrow \mathcal{E}$  be a function, assume for every  $e \in E$  that the function  $\Omega \ni \omega \mapsto X(e, \omega) \in \mathcal{E}$  is  $\mathcal{F}/\mathcal{B}(\mathcal{E})$ -measurable, and assume for every  $\omega \in \Omega$  that the function  $E \ni e \mapsto X(e, \omega) \in \mathcal{E}$  is continuous. Then it holds that the function  $X: E \times \Omega \rightarrow \mathcal{E}$  is  $(\mathcal{B}(E) \otimes \mathcal{F})/\mathcal{B}(\mathcal{E})$ -measurable.*

*Proof of Lemma 2.4.* Throughout this proof let  $(e_m)_{m \in \mathbb{N}} \subseteq E$  be a sequence which satisfies that  $\overline{\{e_m : m \in \mathbb{N}\}} = E$ , let  $P_n : E \rightarrow E$ ,  $n \in \mathbb{N}$ , be the functions which satisfy for every  $n \in \mathbb{N}$ ,  $x \in E$  that

$$P_n(x) = e_{\min\{k \in \{1, 2, \dots, n\} : d(x, e_k) = \min\{d(x, e_1), d(x, e_2), \dots, d(x, e_n)\}\}}, \quad (33)$$

and let  $\mathcal{X}_n : E \times \Omega \rightarrow \mathcal{E}$ ,  $n \in \mathbb{N}$ , be the functions which satisfy for every  $n \in \mathbb{N}$ ,  $x \in E$ ,  $\omega \in \Omega$  that

$$\mathcal{X}_n(x, \omega) = X(P_n(x), \omega). \quad (34)$$

Note that (34) shows that for all  $n \in \mathbb{N}$ ,  $B \in \mathcal{B}(\mathcal{E})$  it holds that

$$\begin{aligned} (\mathcal{X}_n)^{-1}(B) &= \{(x, \omega) \in E \times \Omega : \mathcal{X}_n(x, \omega) \in B\} \\ &= \bigcup_{y \in \text{Im}(P_n)} \left( [(\mathcal{X}_n)^{-1}(B)] \cap [(P_n)^{-1}(\{y\}) \times \Omega] \right) \\ &= \bigcup_{y \in \text{Im}(P_n)} \left\{ (x, \omega) \in E \times \Omega : \left[ \mathcal{X}_n(x, \omega) \in B \text{ and } x \in (P_n)^{-1}(\{y\}) \right] \right\} \\ &= \bigcup_{y \in \text{Im}(P_n)} \left\{ (x, \omega) \in E \times \Omega : \left[ X(P_n(x), \omega) \in B \text{ and } x \in (P_n)^{-1}(\{y\}) \right] \right\}. \end{aligned} \quad (35)$$

Item (ii) in Lemma 2.3 hence implies that for all  $n \in \mathbb{N}$ ,  $B \in \mathcal{B}(\mathcal{E})$  it holds that

$$\begin{aligned} (\mathcal{X}_n)^{-1}(B) &= \bigcup_{y \in \text{Im}(P_n)} \left\{ (x, \omega) \in E \times \Omega : \left[ X(y, \omega) \in B \text{ and } x \in (P_n)^{-1}(\{y\}) \right] \right\} \\ &= \bigcup_{y \in \text{Im}(P_n)} \left( \{(x, \omega) \in E \times \Omega : X(y, \omega) \in B\} \cap [(P_n)^{-1}(\{y\}) \times \Omega] \right) \\ &= \bigcup_{y \in \text{Im}(P_n)} \left( \underbrace{[E \times ((X(y, \cdot))^{-1}(B))]}_{\in (\mathcal{B}(E) \otimes \mathcal{F})} \cap \underbrace{[(P_n)^{-1}(\{y\}) \times \Omega]}_{\in (\mathcal{B}(E) \otimes \mathcal{F})} \right) \in (\mathcal{B}(E) \otimes \mathcal{F}). \end{aligned} \quad (36)$$

This proves that for every  $n \in \mathbb{N}$  it holds that the function  $\mathcal{X}_n$  is  $(\mathcal{B}(E) \otimes \mathcal{F})/\mathcal{B}(\mathcal{E})$ -measurable. In addition, note that item (i) in Lemma 2.3 and the hypothesis that for every  $\omega \in \Omega$  it holds that the function  $E \ni x \mapsto X(x, \omega) \in \mathcal{E}$  is continuous imply that for every  $x \in E$ ,  $\omega \in \Omega$  it holds that

$$\lim_{n \rightarrow \infty} \mathcal{X}_n(x, \omega) = \lim_{n \rightarrow \infty} X(P_n(x), \omega) = X(x, \omega). \quad (37)$$

Combining this with the fact that for every  $n \in \mathbb{N}$  it holds that the function  $\mathcal{X}_n : E \times \Omega \rightarrow \mathcal{E}$  is  $(\mathcal{B}(E) \otimes \mathcal{F})/\mathcal{B}(\mathcal{E})$ -measurable shows that the function  $X : E \times \Omega \rightarrow \mathcal{E}$  is  $(\mathcal{B}(E) \otimes \mathcal{F})/\mathcal{B}(\mathcal{E})$ -measurable. The proof of Lemma 2.4 is thus completed.  $\square$

**Lemma 2.5.** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $(E, d)$  and  $(\mathcal{E}, \delta)$  be separable metric spaces, let  $X_n: \Omega \rightarrow E$ ,  $n \in \mathbb{N}_0$ , be random variables which satisfy for every  $\varepsilon \in (0, \infty)$  that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(d(X_n, X_0) \geq \varepsilon) = 0, \quad (38)$$

and let  $\Phi: E \rightarrow \mathcal{E}$  be a continuous function. Then it holds for every  $\varepsilon \in (0, \infty)$  that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\delta(\Phi(X_n), \Phi(X_0)) \geq \varepsilon) = 0. \quad (39)$$

*Proof of Lemma 2.5.* Note that (38), e.g., Cox et al. [12, Lemma 2.4], and, e.g., Hutzenthaler et al. [35, Lemma 4.2] establish (39). The proof of Lemma 2.5 is thus completed.  $\square$

**Lemma 2.6.** Let  $d, m \in \mathbb{N}$ ,  $T \in (0, \infty)$ ,  $L, a \in \mathbb{R}$ ,  $b \in (a, \infty)$ , let  $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$  be functions which satisfy for every  $x, y \in \mathbb{R}^d$  that  $\max\{\|\mu(x) - \mu(y)\|_{\mathbb{R}^d}, \|\sigma(x) - \sigma(y)\|_{HS(\mathbb{R}^m, \mathbb{R}^d)}\} \leq L\|x - y\|_{\mathbb{R}^d}$ , let  $\Phi: C([0, T], \mathbb{R}^d) \rightarrow \mathbb{R}$  be an at most polynomially growing continuous function, let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space with a normal filtration  $(\mathbb{F}_t)_{t \in [0, T]}$ , let  $\xi: \Omega \rightarrow [a, b]^d$  be a continuous uniformly distributed  $\mathbb{F}_0/\mathcal{B}([a, b]^d)$ -measurable random variable, let  $W: [0, T] \times \Omega \rightarrow \mathbb{R}^m$  be a standard  $(\mathbb{F}_t)_{t \in [0, T]}$ -Brownian motion, for every  $x \in [a, b]^d$  let  $X^x = (X_t^x)_{t \in [0, T]}: [0, T] \times \Omega \rightarrow \mathbb{R}^d$  be an  $(\mathbb{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous sample paths which satisfies that for every  $t \in [0, T]$  it holds  $\mathbb{P}$ -a.s. that

$$X_t^x = x + \int_0^t \mu(X_s^x) ds + \int_0^t \sigma(X_s^x) dW_s, \quad (40)$$

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

$$\mathbb{X}_t = \xi + \int_0^t \mu(\mathbb{X}_s) ds + \int_0^t \sigma(\mathbb{X}_s) dW_s. \quad (41)$$

Then

(i) it holds for every  $x \in [a, b]^d$  that the functions  $\Omega \ni \omega \mapsto \Phi((X_t^x(\omega))_{t \in [0, T]}) \in \mathbb{R}$  and  $\Omega \ni \omega \mapsto \Phi((\mathbb{X}_t(\omega))_{t \in [0, T]}) \in \mathbb{R}$  are  $\mathcal{F}/\mathcal{B}(\mathbb{R})$ -measurable,

(ii) it holds for every  $p \in [2, \infty)$ ,  $x, y \in [a, b]^d$  that

$$\left( \mathbb{E} \left[ \sup_{t \in [0, T]} \|X_t^x - X_t^y\|_{\mathbb{R}^d}^p \right] \right)^{1/p} \leq \sqrt{2} \exp\left(L^2 T [p + \sqrt{T}]^2\right) \|x - y\|_{\mathbb{R}^d}, \quad (42)$$

(iii) it holds for every  $x \in [a, b]^d$  that  $\mathbb{E}[|\Phi((X_t^x)_{t \in [0, T]})| + |\Phi((\mathbb{X}_t)_{t \in [0, T]})|] < \infty$ ,

(iv) it holds that the function  $[a, b]^d \ni x \mapsto \mathbb{E}[\Phi((X_t^x)_{t \in [0, T]})] \in \mathbb{R}$  is continuous, and

(v) it holds that

$$\mathbb{E}[\Phi((\mathbb{X}_t)_{t \in [0, T]})] = \frac{1}{(b-a)^d} \left( \int_{[a, b]^d} \mathbb{E}[\Phi((X_t^x)_{t \in [0, T]})] dx \right). \quad (43)$$

*Proof of Lemma 2.6.* Throughout this proof let  $c \in [1, \infty)$  be a real number which satisfies for every  $w \in C([0, T], \mathbb{R}^m)$  that

$$|\Phi(w)| \leq c \left[ 1 + \sup_{t \in [0, T]} \|w_t\|_{\mathbb{R}^d} \right]^c, \quad (44)$$

let  $p_t: C([0, T], \mathbb{R}^m) \rightarrow \mathbb{R}^m$ ,  $t \in [0, T]$ , be the functions which satisfy for every  $t \in [0, T]$ ,  $w = (w_s)_{s \in [0, T]} \in C([0, T], \mathbb{R}^m)$  that  $p_t(w) = w_t$ , and let  $\Psi_{x, w}^N: [0, T] \rightarrow \mathbb{R}^d$ ,  $N \in \mathbb{N}$ ,  $x \in \mathbb{R}^d$ ,  $w \in C([0, T], \mathbb{R}^m)$ , be the functions which satisfy for every  $w \in C([0, T], \mathbb{R}^m)$ ,  $x \in \mathbb{R}^d$ ,  $N \in \mathbb{N}$ ,  $n \in \{0, 1, \dots, N-1\}$ ,  $t \in [\frac{nT}{N}, \frac{(n+1)T}{N}]$  that  $\Psi_{x, w}^N(0) = x$  and

$$\Psi_{x, w}^N(t) = \Psi_{x, w}^N\left(\frac{nT}{N}\right) + \left(\frac{nt}{T} - n\right) \left[ \mu\left(\Psi_{x, w}^N\left(\frac{nT}{N}\right)\right) \frac{T}{N} + \sigma\left(\Psi_{x, w}^N\left(\frac{nT}{N}\right)\right) \left(w_{\frac{(n+1)T}{N}} - w_{\frac{nT}{N}}\right) \right]. \quad (45)$$

Observe that the fact that the Borel sigma-algebra  $\mathcal{B}(C([0, T], \mathbb{R}^m))$  is generated by the set  $\cup_{t \in [0, T]} \cup_{A \in \mathcal{B}(\mathbb{R}^m)} \{(p_t)^{-1}(A)\}$  (cf., for example, Klenke [39, Theorem 21.31]), the hypothesis that for every  $x \in [a, b]^d$  it holds that  $X^x: [0, T] \times \Omega \rightarrow \mathbb{R}^d$  is an  $(\mathbb{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous sample paths, and the hypothesis that  $\mathbb{X}: [0, T] \times \Omega \rightarrow \mathbb{R}^d$  is an  $(\mathbb{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous sample paths demonstrate that the functions

$$\Omega \ni \omega \mapsto (X_t^x(\omega))_{t \in [0, T]} \in C([0, T], \mathbb{R}^d) \quad (46)$$

and

$$\Omega \ni \omega \mapsto (\mathbb{X}_t(\omega))_{t \in [0, T]} \in C([0, T], \mathbb{R}^d) \quad (47)$$

are  $\mathcal{F}/\mathcal{B}(C([0, T], \mathbb{R}^d))$ -measurable. Combining this with the fact that the function  $\Phi: C([0, T], \mathbb{R}^d) \rightarrow \mathbb{R}$  is  $\mathcal{B}(C([0, T], \mathbb{R}^d))/\mathcal{B}(\mathbb{R})$ -measurable implies that for every  $x \in [a, b]^d$  it holds that the functions  $\Omega \ni \omega \mapsto \Phi((X_t^x(\omega))_{t \in [0, T]}) \in \mathbb{R}$  and  $\Omega \ni \omega \mapsto \Phi((\mathbb{X}_t(\omega))_{t \in [0, T]}) \in \mathbb{R}$  are  $\mathcal{F}/\mathcal{B}(\mathbb{R})$ -measurable. This proves item (i). Next observe that (45), the hypothesis that  $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$  are globally Lipschitz continuous, and the fact that for every  $p \in (0, \infty)$  it holds that  $\mathbb{E}[\|\xi\|_{\mathbb{R}^d}^p] < \infty$  ensure that for every  $p \in (0, \infty)$  it holds that

$$\begin{aligned} & \sup_{N \in \mathbb{N}} \sup_{x \in [a, b]^d} \left( \mathbb{E} \left[ \sup_{t \in [0, T]} \|\Psi_{x, W}^N(t)\|_{\mathbb{R}^d}^p \right] + \mathbb{E} \left[ \sup_{t \in [0, T]} \|\Psi_{\xi, W}^N(t)\|_{\mathbb{R}^d}^p \right] \right) \\ &= \sup_{N \in \mathbb{N}} \sup_{x \in [a, b]^d} \left( \mathbb{E} \left[ \max_{n \in \{0, 1, \dots, N\}} \|\Psi_{x, W}^N\left(\frac{nT}{N}\right)\|_{\mathbb{R}^d}^p \right] + \mathbb{E} \left[ \max_{n \in \{0, 1, \dots, N\}} \|\Psi_{\xi, W}^N\left(\frac{nT}{N}\right)\|_{\mathbb{R}^d}^p \right] \right) < \infty \end{aligned} \quad (48)$$

(cf., for example, Kloeden & Platen [41, Section 10.6]). Next note that (40), (41), (45), the fact that  $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$  are locally Lipschitz continuous functions, and e.g., Hutzenthaler & Jentzen [32, Theorem 3.3] ensure that for every  $x \in [a, b]^d$ ,  $\varepsilon \in (0, \infty)$  it holds that

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left( \sup_{t \in [0, T]} \|X_t^x - \Psi_{x, W}^N(t)\|_{\mathbb{R}^d} \geq \varepsilon \right) = 0 \quad (49)$$

and

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left( \sup_{t \in [0, T]} \|\mathbb{X}_t - \Psi_{\xi, W}^N(t)\|_{\mathbb{R}^d} \geq \varepsilon \right) = 0. \quad (50)$$

Combining (46), (47), and, e.g., Hutzenthaler & Jentzen [32, Lemma 3.10] hence demonstrates that for every  $x \in [a, b]^d$ ,  $p \in (0, \infty)$  it holds that

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in [0, T]} \|X_t^x\|^p \right] &\leq \liminf_{N \rightarrow \infty} \mathbb{E} \left[ \sup_{t \in [0, T]} \|\Psi_{x, W}^N(t)\|^p \right] \\ &\leq \sup_{N \in \mathbb{N}} \mathbb{E} \left[ \sup_{t \in [0, T]} \|\Psi_{x, W}^N(t)\|^p \right] \end{aligned} \quad (51)$$

and

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in [0, T]} \|\mathbb{X}_t\|^p \right] &\leq \liminf_{N \rightarrow \infty} \mathbb{E} \left[ \sup_{t \in [0, T]} \|\Psi_{\xi, W}^N(t)\|^p \right] \\ &\leq \sup_{N \in \mathbb{N}} \mathbb{E} \left[ \sup_{t \in [0, T]} \|\Psi_{\xi, W}^N(t)\|^p \right]. \end{aligned} \quad (52)$$

This and (48) assure that for every  $p \in (0, \infty)$  it holds that

$$\sup_{x \in [a, b]^d} \left( \mathbb{E} \left[ \sup_{t \in [0, T]} \|X_t^x\|_{\mathbb{R}^d}^p \right] + \mathbb{E} \left[ \sup_{t \in [0, T]} \|\mathbb{X}_t\|_{\mathbb{R}^d}^p \right] \right) < \infty. \quad (53)$$

Combining (44) and the fact that  $\forall r \in (0, \infty)$ ,  $a, b \in \mathbb{R}: |a + b|^r \leq 2^r(|a|^r + |b|^r)$  therefore demonstrates that for every  $p \in (0, \infty)$  it holds that

$$\begin{aligned} &\sup_{x \in [a, b]^d} \left( \mathbb{E} [|\Phi((X_t^x)_{t \in [0, T]})|^p] + \mathbb{E} [|\Phi((\mathbb{X}_t)_{t \in [0, T]})|^p] \right) \\ &\leq \sup_{x \in [a, b]^d} \left( c^p \mathbb{E} [1 + \sup_{t \in [0, T]} \|X_t^x\|_{\mathbb{R}^d}^{cp}] + c^p \mathbb{E} [1 + \sup_{t \in [0, T]} \|\mathbb{X}_t\|_{\mathbb{R}^d}^{cp}] \right) \\ &\leq 2^{cp} c^p \left( \sup_{x \in [a, b]^d} \mathbb{E} [2 + \sup_{t \in [0, T]} \|X_t^x\|_{\mathbb{R}^d}^{cp} + \sup_{t \in [0, T]} \|\mathbb{X}_t\|_{\mathbb{R}^d}^{cp}] \right) < \infty. \end{aligned} \quad (54)$$

This establishes item (iii). In the next step we observe that (40) ensures that for every  $x, y \in [a, b]^d$ ,  $t \in [0, T]$  it holds  $\mathbb{P}$ -a.s. that

$$X_t^x - X_t^y = x - y + \int_0^t (\mu(X_s^x) - \mu(X_s^y)) ds + \int_0^t (\sigma(X_s^x) - \sigma(X_s^y)) dW_s. \quad (55)$$

The triangle inequality hence ensures that for every  $x, y \in [a, b]^d$ ,  $t \in [0, T]$  it holds  $\mathbb{P}$ -a.s. that

$$\begin{aligned} & \sup_{s \in [0, t]} \|X_s^x - X_s^y\|_{\mathbb{R}^d} \\ & \leq \|x - y\|_{\mathbb{R}^d} + \sup_{s \in [0, t]} \int_0^s \|\mu(X_r^x) - \mu(X_r^y)\|_{\mathbb{R}^d} dr + \sup_{s \in [0, t]} \left\| \int_0^s (\sigma(X_r^x) - \sigma(X_r^y)) dW_r \right\|_{\mathbb{R}^d} \\ & \leq \|x - y\|_{\mathbb{R}^d} + L \left[ \sup_{s \in [0, t]} \int_0^s \|X_r^x - X_r^y\|_{\mathbb{R}^d} dr \right] + \sup_{s \in [0, t]} \left\| \int_0^s (\sigma(X_r^x) - \sigma(X_r^y)) dW_r \right\|_{\mathbb{R}^d} \\ & = \|x - y\|_{\mathbb{R}^d} + L \int_0^t \|X_r^x - X_r^y\|_{\mathbb{R}^d} dr + \sup_{s \in [0, t]} \left\| \int_0^s (\sigma(X_r^x) - \sigma(X_r^y)) dW_r \right\|_{\mathbb{R}^d}. \end{aligned} \quad (56)$$

Therefore, we obtain for every  $p \in [1, \infty)$ ,  $x, y \in [a, b]^d$ ,  $t \in [0, T]$  that

$$\begin{aligned} \left( \mathbb{E} \left[ \sup_{s \in [0, t]} \|X_s^x - X_s^y\|_{\mathbb{R}^d}^p \right] \right)^{1/p} & \leq \|x - y\|_{\mathbb{R}^d} + L \int_0^t (\mathbb{E} [\|X_r^x - X_r^y\|_{\mathbb{R}^d}^p])^{1/p} dr \\ & \quad + \left( \mathbb{E} \left[ \sup_{s \in [0, t]} \left\| \int_0^s (\sigma(X_r^x) - \sigma(X_r^y)) dW_r \right\|_{\mathbb{R}^d}^p \right] \right)^{1/p}. \end{aligned} \quad (57)$$

The Burkholder-Davis-Gundy type inequality in Da Prato & Zabczyk [13, Lemma 7.2] hence shows that for every  $p \in [2, \infty)$ ,  $x, y \in [a, b]^d$ ,  $t \in [0, T]$  it holds that

$$\begin{aligned} \left( \mathbb{E} \left[ \sup_{s \in [0, t]} \|X_s^x - X_s^y\|_{\mathbb{R}^d}^p \right] \right)^{1/p} & \leq \|x - y\|_{\mathbb{R}^d} + L \int_0^t (\mathbb{E} [\|X_r^x - X_r^y\|_{\mathbb{R}^d}^p])^{1/p} dr \\ & \quad + p \left[ \int_0^t (\mathbb{E} [\|\sigma(X_r^x) - \sigma(X_r^y)\|_{HS(\mathbb{R}^d, \mathbb{R}^m)}^p])^{2/p} dr \right]^{1/2}. \end{aligned} \quad (58)$$

This demonstrates that for every  $p \in [2, \infty)$ ,  $x, y \in [a, b]^d$ ,  $t \in [0, T]$  it holds that

$$\begin{aligned} \left( \mathbb{E} \left[ \sup_{s \in [0, t]} \|X_s^x - X_s^y\|_{\mathbb{R}^d}^p \right] \right)^{1/p} & \leq \|x - y\|_{\mathbb{R}^d} + L \int_0^t (\mathbb{E} [\|X_r^x - X_r^y\|_{\mathbb{R}^d}^p])^{1/p} dr \\ & \quad + Lp \left[ \int_0^t (\mathbb{E} [\|X_r^x - X_r^y\|_{\mathbb{R}^d}^p])^{2/p} dr \right]^{1/2}. \end{aligned} \quad (59)$$

Hölder's inequality hence proves that for every  $p \in [2, \infty)$ ,  $x, y \in [a, b]^d$ ,  $t \in [0, T]$  it holds that

$$\begin{aligned}
\left( \mathbb{E} \left[ \sup_{s \in [0, t]} \|X_s^x - X_s^y\|_{\mathbb{R}^d}^p \right] \right)^{1/p} &\leq \|x - y\|_{\mathbb{R}^d} + L\sqrt{t} \left[ \int_0^t (\mathbb{E} [\|X_r^x - X_r^y\|_{\mathbb{R}^d}^p])^{2/p} dr \right]^{1/2} \\
&\quad + Lp \left[ \int_0^t (\mathbb{E} [\|X_r^x - X_r^y\|_{\mathbb{R}^d}^p])^{2/p} dr \right]^{1/2} \\
&\leq \|x - y\|_{\mathbb{R}^d} + L[p + \sqrt{T}] \left[ \int_0^t (\mathbb{E} [\|X_r^x - X_r^y\|_{\mathbb{R}^d}^p])^{2/p} dr \right]^{1/2}.
\end{aligned} \tag{60}$$

The fact that  $\forall v, w \in \mathbb{R}: |v + w|^2 \leq 2v^2 + 2w^2$  therefore shows that for every  $p \in [2, \infty)$ ,  $x, y \in [a, b]^d$ ,  $t \in [0, T]$  it holds that

$$\begin{aligned}
&\left( \mathbb{E} \left[ \sup_{s \in [0, t]} \|X_s^x - X_s^y\|_{\mathbb{R}^d}^p \right] \right)^{2/p} \\
&\leq 2\|x - y\|_{\mathbb{R}^d}^2 + 2L^2[p + \sqrt{T}]^2 \int_0^t (\mathbb{E} [\|X_r^x - X_r^y\|_{\mathbb{R}^d}^p])^{2/p} dr \\
&\leq 2\|x - y\|_{\mathbb{R}^d}^2 + 2L^2[p + \sqrt{T}]^2 \int_0^t \left( \mathbb{E} \left[ \sup_{s \in [0, r]} \|X_s^x - X_s^y\|_{\mathbb{R}^d}^p \right] \right)^{2/p} dr.
\end{aligned} \tag{61}$$

Combining the Gronwall inequality (cf., e.g., Andersson et al. [2, Lemma 2.6] (with  $\alpha = 0$ ,  $\beta = 0$ ,  $a = 2\|x - y\|_{\mathbb{R}^d}^2$ ,  $b = 3L^2[p + \sqrt{T}]$ ,  $e = ([0, T] \ni t \mapsto (\mathbb{E} [\sup_{s \in [0, t]} \|X_s^x - X_s^y\|_{\mathbb{R}^d}^p])^{2/p} \in [0, \infty)$  in the notation of Lemma 2.6)) and (53) hence establishes that for every  $p \in [2, \infty)$ ,  $x, y \in [a, b]^d$ ,  $t \in [0, T]$  it holds that

$$\left( \mathbb{E} \left[ \sup_{s \in [0, t]} \|X_s^x - X_s^y\|_{\mathbb{R}^d}^p \right] \right)^{2/p} \leq 2\|x - y\|_{\mathbb{R}^d}^2 \exp\left(2L^2t[p + \sqrt{T}]^2\right). \tag{62}$$

Therefore, we obtain that for every  $p \in [2, \infty)$ ,  $x, y \in [a, b]^d$  it holds that

$$\left( \mathbb{E} \left[ \sup_{t \in [0, T]} \|X_t^x - X_t^y\|_{\mathbb{R}^d}^p \right] \right)^{1/p} \leq \sqrt{2} \exp\left(L^2T[p + \sqrt{T}]^2\right) \|x - y\|_{\mathbb{R}^d}. \tag{63}$$

This establishes item (ii). Next observe that item (ii) and Jensen's inequality imply that for every  $p \in (0, \infty)$  it holds that

$$\sup_{x, y \in [a, b]^d, x \neq y} \left( \frac{(\mathbb{E} [\sup_{t \in [0, T]} \|X_t^x - X_t^y\|_{\mathbb{R}^d}^p])^{1/p}}{\|x - y\|_{\mathbb{R}^d}} \right) < \infty. \tag{64}$$



The hypothesis that the function  $\Phi: C([0, T], \mathbb{R}^d) \rightarrow \mathbb{R}$  is continuous and Lemma 2.5 hence ensure that for every  $\varepsilon \in (0, \infty)$ ,  $(x_n)_{n \in \mathbb{N}_0} \subseteq [a, b]^d$  with  $\limsup_{n \rightarrow \infty} \|x_0 - x_n\|_{\mathbb{R}^d} = 0$  it holds that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|\Phi((X_t^{x_0})_{t \in [0, T]}) - \Phi((X_t^{x_n})_{t \in [0, T]})| \geq \varepsilon) = 0. \quad (65)$$

Combining (54) with, e.g., Hutzenthaler et al. [35, Proposition 4.5] therefore implies that for every  $(x_n)_{n \in \mathbb{N}_0} \subseteq [a, b]^d$  with  $\limsup_{n \rightarrow \infty} \|x_0 - x_n\|_{\mathbb{R}^d} = 0$  it holds that

$$\limsup_{n \rightarrow \infty} \mathbb{E}[|\Phi((X_t^{x_0})_{t \in [0, T]}) - \Phi((X_t^{x_n})_{t \in [0, T]})|] = 0. \quad (66)$$

This establishes item (iv). In the next step we observe that (49) and (50) ensure that for every  $\varepsilon \in (0, \infty)$ ,  $x \in [a, b]^d$  it holds that

$$\limsup_{N \rightarrow \infty} \left[ \mathbb{P} \left( \sup_{t \in [0, T]} \|\Psi_{x, W}^N(t) - X_t^x\|_{\mathbb{R}^d} \geq \varepsilon \right) + \mathbb{P} \left( \sup_{t \in [0, T]} \|\Psi_{\xi, W}^N(t) - \mathbb{X}_t\|_{\mathbb{R}^d} \geq \varepsilon \right) \right] = 0. \quad (67)$$

The hypothesis that the function  $\Phi$  is continuous and Lemma 2.5 therefore demonstrate that for every  $\varepsilon \in (0, \infty)$ ,  $x \in [a, b]^d$  it holds that

$$\limsup_{N \rightarrow \infty} \left[ \mathbb{P}(|\Phi(\Psi_{x, W}^N) - \Phi((X_t^x)_{t \in [0, T]})| + |\Phi(\Psi_{\xi, W}^N) - \Phi((\mathbb{X}_t)_{t \in [0, T]})| \geq \varepsilon) \right] = 0. \quad (68)$$

Next observe that (44) assures that for every  $N \in \mathbb{N}$ ,  $x \in [a, b]^d$ ,  $p \in (0, \infty)$  it holds that

$$\begin{aligned} & \mathbb{E}[|\Phi(\Psi_{x, W}^N) - \Phi((X_t^x)_{t \in [0, T]})|^p + |\Phi(\Psi_{\xi, W}^N) - \Phi((\mathbb{X}_t)_{t \in [0, T]})|^p] \\ & \leq 2^p \mathbb{E}[|\Phi(\Psi_{x, W}^N)|^p + |\Phi((X_t^x)_{t \in [0, T]})|^p] + 2^p \mathbb{E}[|\Phi(\Psi_{\xi, W}^N)|^p + |\Phi((\mathbb{X}_t)_{t \in [0, T]})|^p] \\ & \leq 2^p c^p \mathbb{E} \left[ |1 + \sup_{t \in [0, T]} \|\Psi_{x, W}^N(t)\|_{\mathbb{R}^d}|^{cp} + |1 + \sup_{t \in [0, T]} \|X_t^x\|_{\mathbb{R}^d}|^{cp} \right] \\ & \quad + 2^p c^p \mathbb{E} \left[ |1 + \sup_{t \in [0, T]} \|\Psi_{\xi, W}^N(t)\|_{\mathbb{R}^d}|^{cp} + |1 + \sup_{t \in [0, T]} \|\mathbb{X}_t\|_{\mathbb{R}^d}|^{cp} \right] \\ & \leq 4^p c^p \mathbb{E} \left[ 2 + \sup_{t \in [0, T]} \|\Psi_{x, W}^N(t)\|_{\mathbb{R}^d}^{cp} + \sup_{t \in [0, T]} \|X_t^x\|_{\mathbb{R}^d}^{cp} \right] \\ & \quad + 4^p c^p \mathbb{E} \left[ 2 + \sup_{t \in [0, T]} \|\Psi_{\xi, W}^N(t)\|_{\mathbb{R}^d}^{cp} + \sup_{t \in [0, T]} \|\mathbb{X}_t\|_{\mathbb{R}^d}^{cp} \right]. \end{aligned} \quad (69)$$

Combining (48) and (53) hence shows that for every  $p \in (0, \infty)$  it holds that

$$\sup_{N \in \mathbb{N}} \sup_{x \in [a, b]^d} \left( \mathbb{E}[|\Phi(\Psi_{x, W}^N) - \Phi((X_t^x)_{t \in [0, T]})|^p + |\Phi(\Psi_{\xi, W}^N) - \Phi((\mathbb{X}_t)_{t \in [0, T]})|^p] \right) < \infty. \quad (70)$$

This, (68), and, e.g., Hutzenthaler et al. [35, Proposition 4.5] imply that for every  $x \in [a, b]^d$  it holds that

$$\limsup_{N \rightarrow \infty} \left( \mathbb{E}[|\Phi(\Psi_{x, W}^N) - \Phi((X_t^x)_{t \in [0, T]})|] + \mathbb{E}[|\Phi(\Psi_{\xi, W}^N) - \Phi((\mathbb{X}_t)_{t \in [0, T]})|] \right) = 0. \quad (71)$$

Combining (70) with Lebesgue's dominated convergence theorem therefore demonstrates that

$$\limsup_{N \rightarrow \infty} \left( \int_{[a,b]^d} \mathbb{E} [|\Phi(\Psi_{x,W}^N) - \Phi((X_t^x)_{t \in [0,T]})|] dx \right) = 0. \quad (72)$$

In addition, observe that (69), (48), and (53) prove that for all  $p \in (0, \infty)$  it holds that

$$\sup_{N \in \mathbb{N}} \sup_{x \in [a,b]^d} \mathbb{E} [|\Phi(\Psi_{x,W}^N)|^p] < \infty. \quad (73)$$

Next observe that (71) and the fact that  $\xi$  and  $W$  are independent imply that

$$\begin{aligned} \mathbb{E} [\Phi((X_t)_{t \in [0,T]})] &= \lim_{N \rightarrow \infty} \mathbb{E} [\Phi(\Psi_{\xi, W}^N)] \\ &= \lim_{N \rightarrow \infty} \left[ \int_{\Omega} \Phi(\Psi_{\xi(\omega), (W_t(\omega))_{t \in [0,T]}}^N) \mathbb{P}(d\omega) \right] \\ &= \lim_{N \rightarrow \infty} \left[ \int_{[a,b]^d \times C([0,T], \mathbb{R}^d)} \Phi(\Psi_{x,W}^N) ((\xi, W)(\mathbb{P}))(dx, dw) \right] \\ &= \lim_{N \rightarrow \infty} \left[ \int_{[a,b]^d \times C([0,T], \mathbb{R}^d)} \Phi(\Psi_{x,w}^N) ((\xi(\mathbb{P})) \otimes (W(\mathbb{P})))(dx, dw) \right]. \end{aligned} \quad (74)$$

Combining Fubini's theorem, (72), and (73) with Lebesgue's dominated convergence theorem therefore assures that

$$\begin{aligned} \mathbb{E} [\Phi((X_t)_{t \in [0,T]})] &= \lim_{N \rightarrow \infty} \left[ \int_{[a,b]^d} \left( \int_{C([0,T], \mathbb{R}^d)} \Phi(\Psi_{x,w}^N) (W(\mathbb{P}))(dw) \right) (\xi(\mathbb{P}))(dx) \right] \\ &= \lim_{N \rightarrow \infty} \left[ \int_{[a,b]^d} \left( \int_{\Omega} \Phi(\Psi_{x,w}^N) \mathbb{P}(dw) \right) (\xi(\mathbb{P}))(dx) \right] \\ &= \lim_{N \rightarrow \infty} \left[ \int_{[a,b]^d} \mathbb{E} [\Phi(\Psi_{x,W}^N)] (\xi(\mathbb{P}))(dx) \right] \\ &= \int_{[a,b]^d} \lim_{N \rightarrow \infty} \mathbb{E} [\Phi(\Psi_{x,W}^N)] (\xi(\mathbb{P}))(dx) \\ &= \int_{[a,b]^d} \mathbb{E} [\Phi((X_t^x)_{t \in [0,T]})] (\xi(\mathbb{P}))(dx) \\ &= \frac{1}{(b-a)^d} \left( \int_{[a,b]^d} \mathbb{E} [\Phi((X_t^x)_{t \in [0,T]})] dx \right). \end{aligned} \quad (75)$$

This establishes item (v). The proof of Lemma 2.6 is thus completed.  $\square$

**Proposition 2.7.** *Let  $d, m \in \mathbb{N}$ ,  $T \in (0, \infty)$ ,  $a \in \mathbb{R}$ ,  $b \in (a, \infty)$ , let  $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$  be globally Lipschitz continuous functions, let  $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}$  be a function, let  $u = (u(t, x))_{(t,x) \in [0,T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  be a function with at most polynomially*

growing partial derivatives which satisfies for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  that  $u(0, x) = \varphi(x)$  and

$$\frac{\partial u}{\partial t}(t, x) = \frac{1}{2} \text{Trace}_{\mathbb{R}^d}(\sigma(x)[\sigma(x)]^*(\text{Hess}_x u)(t, x)) + \langle \mu(x), (\nabla_x u)(t, x) \rangle_{\mathbb{R}^d}, \quad (76)$$

let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space with a normal filtration  $(\mathbb{F}_t)_{t \in [0, T]}$ , let  $W: [0, T] \times \Omega \rightarrow \mathbb{R}^m$  be a standard  $(\mathbb{F}_t)_{t \in [0, T]}$ -Brownian motion, let  $\xi: \Omega \rightarrow [a, b]^d$  be a continuous uniformly distributed  $\mathbb{F}_0/\mathcal{B}([a, b]^d)$ -measurable random variable, and let  $\mathbb{X} = (\mathbb{X}_t)_{t \in [0, T]}: [0, T] \times \Omega \rightarrow \mathbb{R}^d$  be an  $(\mathbb{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous sample paths which satisfies that for every  $t \in [0, T]$  it holds  $\mathbb{P}$ -a.s. that

$$\mathbb{X}_t = \xi + \int_0^t \mu(\mathbb{X}_s) ds + \int_0^t \sigma(\mathbb{X}_s) dW_s. \quad (77)$$

Then

(i) it holds that the function  $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}$  is twice continuously differentiable with at most polynomially growing derivatives,

(ii) it holds that there exists a unique continuous function  $U: [a, b]^d \rightarrow \mathbb{R}$  such that

$$\mathbb{E}[|\varphi(\mathbb{X}_T) - U(\xi)|^2] = \inf_{v \in C([a, b]^d, \mathbb{R})} \mathbb{E}[|\varphi(\mathbb{X}_T) - v(\xi)|^2], \quad (78)$$

and

(iii) it holds for every  $x \in [a, b]^d$  that  $U(x) = u(T, x)$ .

*Proof of Proposition 2.7.* Throughout this proof let  $X^x = (X_t^x)_{t \in [0, T]}: [0, T] \times \Omega \rightarrow \mathbb{R}^d$ ,  $x \in [a, b]^d$ , be  $(\mathbb{F}_t)_{t \in [0, T]}$ -adapted stochastic processes with continuous sample paths

a) which satisfy that for every  $t \in [0, T]$ ,  $x \in [a, b]^d$  it holds  $\mathbb{P}$ -a.s. that

$$X_t^x = x + \int_0^t \mu(X_s^x) ds + \int_0^t \sigma(X_s^x) dW_s \quad (79)$$

and

b) which satisfy that for every  $\omega \in \Omega$  it holds that the function  $[a, b]^d \ni x \mapsto X_T^x(\omega) \in \mathbb{R}^d$  is continuous (cf., for example, Cox et al. [11, Theorem 3.5] and item (ii) in Lemma 2.6).

Note that the assumption that  $\forall x \in \mathbb{R}^d: u(0, x) = \varphi(x)$  and the assumption that  $u \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  has at most polynomially growing partial derivatives establish item (i). Next note that item (i) and the fact that for every  $p \in (0, \infty)$ ,  $x \in [a, b]^d$  it holds that

$$\sup_{t \in [0, T]} \mathbb{E}[\|X_t^x\|_{\mathbb{R}^d}^p] < \infty \quad (80)$$

assure that for every  $x \in [a, b]^d$  it holds that

$$\mathbb{E}[|\varphi(X_T^x)|^2] < \infty. \quad (81)$$

Item (i), the assumption that for every  $\omega \in \Omega$  it holds that the function  $[a, b]^d \ni x \mapsto X_T^x(\omega) \in \mathbb{R}^d$  is continuous, and, e.g., Hutzenthaler et al. [35, Proposition 4.5] hence ensure that the function

$$[a, b]^d \ni x \mapsto \mathbb{E}[\varphi(X_T^x)] \in \mathbb{R} \quad (82)$$

is continuous. In the next step we combine the fact that for every  $x \in [a, b]^d$  it holds that the function  $\Omega \ni \omega \mapsto \varphi(X_T^x(\omega)) \in \mathbb{R}$  is  $\mathcal{F}/\mathcal{B}(\mathbb{R})$ -measurable, the fact that for every  $\omega \in \Omega$  it holds that the function  $[a, b]^d \ni x \mapsto \varphi(X_T^x(\omega)) \in \mathbb{R}$  is continuous, and Lemma 2.4 to obtain that the function

$$[a, b]^d \times \Omega \ni (x, \omega) \mapsto \varphi(X_T^x(\omega)) \in \mathbb{R} \quad (83)$$

is  $(\mathcal{B}([a, b]^d) \otimes \mathcal{F})/\mathcal{B}(\mathbb{R})$ -measurable. Combining this, (81), (82), and Proposition 2.2 demonstrates

A) that there exists a unique continuous function  $U: [a, b]^d \rightarrow \mathbb{R}$  which satisfies that

$$\int_{[a, b]^d} \mathbb{E}[|\varphi(X_T^x) - U(x)|^2] dx = \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \int_{[a, b]^d} \mathbb{E}[|\varphi(X_T^x) - v(x)|^2] dx \right) \quad (84)$$

and

B) that it holds for every  $x \in [a, b]^d$  that

$$U(x) = \mathbb{E}[\varphi(X_T^x)]. \quad (85)$$

Next note that for every continuous function  $V: [a, b]^d \rightarrow \mathbb{R}$  it holds that

$$\sup_{x \in [a, b]^d} |V(x)| < \infty. \quad (86)$$

Item (i) hence implies that for every continuous function  $V: [a, b]^d \rightarrow \mathbb{R}$  it holds that

$$C([0, T], \mathbb{R}^d) \ni (z_t)_{t \in [0, T]} \mapsto |\varphi(z_T) - V(z_0)|^2 \in \mathbb{R} \quad (87)$$

is an at most polynomially growing continuous function. Combining Lemma 2.6 (with  $\Phi = (C([0, T], \mathbb{R}^d) \ni (z_t)_{t \in [0, T]} \mapsto |\varphi(z_T) - V(z_0)|^2 \in \mathbb{R})$  for  $V \in C([a, b]^d, \mathbb{R}^d)$  in the notation of Lemma 2.6), (77), item (i), and (79) hence ensures that for every continuous function  $V: [a, b]^d \rightarrow \mathbb{R}$  it holds that

$$\mathbb{E}[|\varphi(X_T) - V(\xi)|^2] = \frac{1}{(b-a)^d} \left[ \int_{[a, b]^d} \mathbb{E}[|\varphi(X_T^x) - V(x)|^2] dx \right]. \quad (88)$$

Hence, we obtain that for every continuous function  $V: [a, b]^d \rightarrow \mathbb{R}$  with  $\mathbb{E}[|\varphi(\mathbb{X}_T) - V(\xi)|^2] = \inf_{v \in C([a, b]^d, \mathbb{R})} \mathbb{E}[|\varphi(\mathbb{X}_T) - v(\xi)|^2]$  it holds that

$$\begin{aligned}
& \int_{[a, b]^d} \mathbb{E}[|\varphi(X_T^x) - V(x)|^2] dx \\
&= (b-a)^d \left( \frac{1}{(b-a)^d} \left[ \int_{[a, b]^d} \mathbb{E}[|\varphi(X_T^x) - V(x)|^2] dx \right] \right) \\
&= (b-a)^d (\mathbb{E}[|\varphi(\mathbb{X}_T) - V(\xi)|^2]) \\
&= (b-a)^d \left( \inf_{v \in C([a, b]^d, \mathbb{R})} \mathbb{E}[|\varphi(\mathbb{X}_T) - v(\xi)|^2] \right) \\
&= (b-a)^d \left( \inf_{v \in C([a, b]^d, \mathbb{R})} \left( \frac{1}{(b-a)^d} \left[ \int_{[a, b]^d} \mathbb{E}[|\varphi(X_T^x) - v(x)|^2] dx \right] \right) \right) \\
&= \inf_{v \in C([a, b]^d, \mathbb{R})} \left[ \int_{[a, b]^d} \mathbb{E}[|\varphi(X_T^x) - v(x)|^2] dx \right].
\end{aligned} \tag{89}$$

Combining this with (84) proves that for every continuous function  $V: [a, b]^d \rightarrow \mathbb{R}$  with  $\mathbb{E}[|\varphi(\mathbb{X}_T) - V(\xi)|^2] = \inf_{v \in C([a, b]^d, \mathbb{R})} \mathbb{E}[|\varphi(\mathbb{X}_T) - v(\xi)|^2]$  it holds that

$$U = V. \tag{90}$$

Next observe that (88) and (84) demonstrate that

$$\begin{aligned}
\mathbb{E}[|\varphi(\mathbb{X}_T) - U(\xi)|^2] &= \frac{1}{(b-a)^d} \left( \int_{[a, b]^d} \mathbb{E}[|\varphi(X_T^x) - U(x)|^2] dx \right) \\
&= \inf_{v \in C([a, b]^d, \mathbb{R})} \left[ \frac{1}{(b-a)^d} \left( \int_{[a, b]^d} \mathbb{E}[|\varphi(X_T^x) - v(x)|^2] dx \right) \right] \\
&= \inf_{v \in C([a, b]^d, \mathbb{R})} \mathbb{E}[|\varphi(\mathbb{X}_T) - v(\xi)|^2].
\end{aligned} \tag{91}$$

Combining this with (90) proves item (ii). Next note that (76), (85), and the Feynman-Kac formula (cf., for example, Hairer et al. [22, Corollary 4.17]) imply that for every  $x \in [a, b]^d$  it holds that  $U(x) = \mathbb{E}[\varphi(X_T^x)] = u(T, x)$ . This establishes item (iii). The proof of Proposition 2.7 is thus completed.  $\square$

In the next step we use Proposition 2.7 to obtain a minimization problem which is uniquely solved by the function  $[a, b]^d \ni x \mapsto u(T, x) \in \mathbb{R}$ . More specifically, let  $\xi: \Omega \rightarrow [a, b]^d$  be a continuously uniformly distributed  $\mathbb{F}_0/\mathcal{B}([a, b]^d)$ -measurable random variable, and let  $\mathbb{X}: [0, T] \times \Omega \rightarrow \mathbb{R}^d$  be an  $(\mathbb{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous sample paths which satisfies that for every  $t \in [0, T]$  it holds  $\mathbb{P}$ -a.s. that

$$\mathbb{X}_t = \xi + \int_0^t \mu(\mathbb{X}_s) ds + \int_0^t \sigma(\mathbb{X}_s) dW_s. \tag{92}$$

Proposition 2.7 then guarantees that the function  $[a, b]^d \ni x \mapsto u(T, x) \in \mathbb{R}$  is the unique global minimizer of the function

$$C([a, b]^d, \mathbb{R}) \ni v \mapsto \mathbb{E}[|\varphi(\mathbb{X}_T) - v(\xi)|^2] \in \mathbb{R}. \quad (93)$$

In the following two subsections we derive an approximated minimization problem by discretizing the stochastic process  $\mathbb{X}: [0, T] \times \Omega \rightarrow \mathbb{R}^d$  (see Subsection 2.4 below) and by employing a deep neural network approximation for the function  $\mathbb{R}^d \ni x \mapsto u(T, x) \in \mathbb{R}$  (see Subsection 2.5 below).

## 2.4 Discretization of the stochastic differential equation

In this subsection we use the Euler-Maruyama scheme (cf., for example, Kloeden & Platen [41] and Maruyama [44]) to temporally discretize the solution process  $\mathbb{X}$  of the SDE (92).

More specifically, let  $N \in \mathbb{N}$ , let  $t_0, t_1, \dots, t_N \in [0, \infty)$  be real numbers which satisfy that

$$0 = t_0 < t_1 < \dots < t_N = T. \quad (94)$$

Note that (92) implies that for every  $n \in \{0, 1, \dots, N-1\}$  it holds  $\mathbb{P}$ -a.s. that

$$\mathbb{X}_{t_{n+1}} = \mathbb{X}_{t_n} + \int_{t_n}^{t_{n+1}} \mu(\mathbb{X}_s) ds + \int_{t_n}^{t_{n+1}} \sigma(\mathbb{X}_s) dW_s. \quad (95)$$

This suggests that for sufficiently small mesh size  $\sup_{n \in \{0, 1, \dots, N-1\}} (t_{n+1} - t_n)$  it holds that

$$\mathbb{X}_{t_{n+1}} \approx \mathbb{X}_{t_n} + \mu(\mathbb{X}_{t_n}) (t_{n+1} - t_n) + \sigma(\mathbb{X}_{t_n}) (W_{t_{n+1}} - W_{t_n}). \quad (96)$$

Let  $\mathcal{X}: \{0, 1, \dots, N\} \times \Omega \rightarrow \mathbb{R}^d$  be the stochastic process which satisfies for every  $n \in \{0, 1, \dots, N-1\}$  that  $\mathcal{X}_0 = \xi$  and

$$\mathcal{X}_{n+1} = \mathcal{X}_n + \mu(\mathcal{X}_n) (t_{n+1} - t_n) + \sigma(\mathcal{X}_n) (W_{t_{n+1}} - W_{t_n}). \quad (97)$$

Observe that (96) and (97) suggest, in turn, that for every  $n \in \{0, 1, 2, \dots, N\}$  it holds that

$$\mathcal{X}_n \approx \mathbb{X}_{t_n} \quad (98)$$

(cf., for example, Theorem 2.8 below for a strong convergence result for the Euler-Maruyama scheme).

**Theorem 2.8** (Strong convergence rate for the Euler-Maruyama scheme). *Let  $T \in (0, \infty)$ ,  $d \in \mathbb{N}$ ,  $p \in [2, \infty)$ , let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space with a normal filtration  $(\mathbb{F}_t)_{t \in [0, T]}$ , let  $W: [0, T] \times \Omega \rightarrow \mathbb{R}^d$  be a standard  $(\mathbb{F}_t)_{t \in [0, T]}$ -Brownian motion, let  $\xi: \Omega \rightarrow \mathbb{R}^d$  be a random variable which satisfies that  $\mathbb{E}[\|\xi\|_{\mathbb{R}^d}^p] < \infty$ , let  $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  be Lipschitz*

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

$$\mathbb{X}_t = \xi + \int_0^t \mu(\mathbb{X}_s) ds + \int_0^t \sigma(\mathbb{X}_s) dW_s, \quad (99)$$

for every  $N \in \mathbb{N}$  let  $t_0^N, t_1^N, \dots, t_N^N \in [0, T]$  be real numbers which satisfy that

$$0 = t_0^N < t_1^N < \dots < t_N^N = T, \quad (100)$$

and for every  $N \in \mathbb{N}$  let  $\mathcal{X}^N: \{0, 1, \dots, N\} \times \Omega \rightarrow \mathbb{R}^d$  be the stochastic process which satisfies for every  $n \in \{0, 1, \dots, N-1\}$  that  $\mathcal{X}_0^N = \xi_0$  and

$$\mathcal{X}_{n+1}^N = \mathcal{X}_n^N + \mu(\mathcal{X}_n^N)(t_{n+1}^N - t_n^N) + \sigma(\mathcal{X}_n^N)(W_{t_{n+1}^N} - W_{t_n^N}). \quad (101)$$

Then there exists a real number  $C \in (0, \infty)$  such that for every  $N \in \mathbb{N}$  it holds that

$$\sup_{n \in \{0, 1, \dots, N\}} \left( \mathbb{E} [\|\mathbb{X}_{t_n^N} - \mathcal{X}_n^N\|_{\mathbb{R}^d}^p] \right)^{1/p} \leq C \left[ \max_{n \in \{0, 1, \dots, N-1\}} |t_{n+1}^N - t_n^N| \right]^{1/2}. \quad (102)$$

The proof of Theorem 2.4 is well-known in the literature (cf., for instance, Kloeden & Platen [41], Milstein [46], Hofmann, Müller-Gronbach, & Ritter [30], Müller-Gronbach & Ritter [49], and the references mentioned therein).

## 2.5 Deep artificial neural network approximations

In this subsection we employ suitable approximations for the solution  $\mathbb{R}^d \ni x \mapsto u(T, x) \in \mathbb{R}$  of the PDE (1) at time  $T$ .

More specifically, let  $\nu \in \mathbb{N}$  and let  $\mathbb{U} = (\mathbb{U}(\theta, x))_{(\theta, x) \in \mathbb{R}^\nu \times \mathbb{R}^d}: \mathbb{R}^\nu \times \mathbb{R}^d \rightarrow \mathbb{R}$  be a continuous function. For every suitable  $\theta \in \mathbb{R}^\nu$  and every  $x \in [a, b]^d$  we think of  $\mathbb{U}(\theta, x) \in \mathbb{R}$  as an appropriate approximation

$$\mathbb{U}(\theta, x) \approx u(T, x) \quad (103)$$

of  $u(T, x)$ . We suggest to choose the function  $\mathbb{U}: \mathbb{R}^\nu \times \mathbb{R}^d \rightarrow \mathbb{R}$  as a deep neural network (cf., for example, Bishop [6]). For instance, let  $\mathcal{L}_d: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be the function which satisfies for every  $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$  that

$$\mathcal{L}_d(x) = \left( \frac{\exp(x_1)}{\exp(x_1) + 1}, \frac{\exp(x_2)}{\exp(x_2) + 1}, \dots, \frac{\exp(x_d)}{\exp(x_d) + 1} \right) \quad (104)$$

(multidimensional version of the standard logistic function), for every  $k, l \in \mathbb{N}$ ,  $v \in \mathbb{N}_0 = \{0\} \cup \mathbb{N}$ ,  $\theta = (\theta_1, \dots, \theta_\nu) \in \mathbb{R}^\nu$  with  $v + l(k+1) \leq \nu$  let  $A_{k,l}^{\theta, v}: \mathbb{R}^k \rightarrow \mathbb{R}^l$  be the function

which satisfies for every  $x = (x_1, \dots, x_k) \in \mathbb{R}^k$  that

$$A_{k,l}^{\theta,v}(x) = \begin{pmatrix} \theta_{v+1} & \theta_{v+2} & \dots & \theta_{v+k} \\ \theta_{v+k+1} & \theta_{v+k+2} & \dots & \theta_{v+2k} \\ \theta_{v+2k+1} & \theta_{v+2k+2} & \dots & \theta_{v+3k} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{v+(l-1)k+1} & \theta_{v+(l-1)k+2} & \dots & \theta_{v+lk} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_k \end{pmatrix} + \begin{pmatrix} \theta_{v+kl+1} \\ \theta_{v+kl+2} \\ \theta_{v+kl+3} \\ \vdots \\ \theta_{v+kl+l} \end{pmatrix}, \quad (105)$$

let  $s \in \{3, 4, 5, 6, \dots\}$ , assume that  $(s-1)d(d+1) + d + 1 \leq \nu$ , and let  $\mathbb{U}: \mathbb{R}^\nu \times \mathbb{R}^d \rightarrow \mathbb{R}$  be the function which satisfies for every  $\theta \in \mathbb{R}^\nu$ ,  $x \in \mathbb{R}^d$  that

$$\mathbb{U}(\theta, x) = (A_{d,1}^{\theta,(s-1)d(d+1)} \circ \mathcal{L}_d \circ A_{d,d}^{\theta,(s-2)d(d+1)} \circ \dots \circ \mathcal{L}_d \circ A_{d,d}^{\theta,d(d+1)} \circ \mathcal{L}_d \circ A_{d,d}^{\theta,0})(x). \quad (106)$$

The function  $\mathbb{U}: \mathbb{R}^\nu \times \mathbb{R}^d \rightarrow \mathbb{R}$  in (106) describes an artificial neural network with  $s+1$  layers (1 input layer with  $d$  neurons,  $s-1$  hidden layers with  $d$  neurons each, and 1 output layer with  $d$  neurons) and standard logistic functions as activation functions (cf., for instance, Bishop [6]).

## 2.6 Stochastic gradient descent-type minimization

As described in Subsection 2.5 for every *suitable*  $\theta \in \mathbb{R}^\nu$  and every  $x \in [a, b]^d$  we think of  $\mathbb{U}(\theta, x) \in \mathbb{R}$  as an appropriate approximation of  $u(T, x) \in \mathbb{R}$ . In this subsection we intend to find a *suitable*  $\theta \in \mathbb{R}^\nu$  as an approximate minimizer of the function

$$\mathbb{R}^\nu \ni \theta \mapsto \mathbb{E}[|\varphi(\mathcal{X}_N) - \mathbb{U}(\theta, \xi)|^2] \in \mathbb{R}. \quad (107)$$

To be more specific, we intend to find an approximate minimizer of the function in (107) through a stochastic gradient descent-type minimization algorithm (cf., for instance, Ruder [53, Section 4], Jentzen et al. [37], and the references mentioned therein). For this we approximate the derivative of the function in (107) by means of the Monte Carlo method.

More precisely, let  $\xi^{(m)}: \Omega \rightarrow [a, b]^d$ ,  $m \in \mathbb{N}_0$ , be independent continuously uniformly distributed  $\mathbb{F}_0/\mathcal{B}([a, b]^d)$ -measurable random variables, let  $W^{(m)}: [0, T] \times \Omega \rightarrow \mathbb{R}^d$ ,  $m \in \mathbb{N}_0$ , be independent standard  $(\mathbb{F}_t)_{t \in [0, T]}$ -Brownian motions, for every  $m \in \mathbb{N}_0$  let  $\mathcal{X}^{(m)} = (\mathcal{X}_n^{(m)})_{n \in \{0, 1, \dots, N\}}: \{0, 1, \dots, N\} \times \Omega \rightarrow \mathbb{R}^d$  be the stochastic process which satisfies for every  $n \in \{0, 1, \dots, N-1\}$  that  $\mathcal{X}_0^{(m)} = \xi^{(m)}$  and

$$\mathcal{X}_{n+1}^{(m)} = \mathcal{X}_n^{(m)} + \mu(\mathcal{X}_n^{(m)}) (t_{n+1} - t_n) + \sigma(\mathcal{X}_n^{(m)}) (W_{t_{n+1}}^{(m)} - W_{t_n}^{(m)}), \quad (108)$$

let  $\gamma \in (0, \infty)$ , and let  $\Theta: \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R}^\nu$  be a stochastic process which satisfies for every  $m \in \mathbb{N}_0$  that

$$\Theta_{m+1} = \Theta_m - 2\gamma \cdot (\mathbb{U}(\Theta_m, \xi^{(m)}) - \varphi(\mathcal{X}_N^{(m)})) \cdot (\nabla_\theta \mathbb{U})(\Theta_m, \xi^{(m)}). \quad (109)$$



Under appropriate hypotheses we think for every sufficiently large  $m \in \mathbb{N}$  of the random variable  $\Theta_m: \Omega \rightarrow \mathbb{R}^\nu$  as a suitable approximation of a local minimum point of the function (107) and we think for every sufficiently large  $m \in \mathbb{N}$  of the random function  $[a, b]^d \ni x \mapsto \mathbb{U}(\Theta_m, x) \in \mathbb{R}$  as a suitable approximation of the function  $[a, b]^d \ni x \mapsto u(T, x) \in \mathbb{R}$ .

## 2.7 Description of the algorithm in a special case

In this subsection we give a description of the proposed approximation method in a special case, that is, we describe the proposed approximation method in the specific case where a particular neural network approximation is chosen and where the plain-vanilla stochastic gradient descent method with a constant learning rate is the employed stochastic minimization algorithm (cf. (109) above). For a more general description of the proposed approximation method we refer the reader to Subsection 2.8 below.

**Framework 2.9.** *Let  $T, \gamma \in (0, \infty)$ ,  $a \in \mathbb{R}$ ,  $b \in (a, \infty)$ ,  $d, N \in \mathbb{N}$ ,  $s \in \{3, 4, 5, \dots\}$ , let  $\nu = sd(d+1)$ , let  $t_0, t_1, \dots, t_N \in [0, T]$  be real numbers with*

$$0 = t_0 < t_1 < \dots < t_N = T, \quad (110)$$

let  $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  be continuous functions, let  $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t \in [0, T]})$  be a filtered probability space, let  $\xi^{(m)}: \Omega \rightarrow [a, b]^d$ ,  $m \in \mathbb{N}_0$ , be independent continuously uniformly distributed  $\mathbb{F}_0/\mathcal{B}([a, b]^d)$ -measurable random variables, let  $W^{(m)}: [0, T] \times \Omega \rightarrow \mathbb{R}^d$ ,  $m \in \mathbb{N}_0$ , be i.i.d. standard  $(\mathbb{F}_t)_{t \in [0, T]}$ -Brownian motions, for every  $m \in \mathbb{N}_0$  let  $\mathcal{X}^{(m)}: \{0, 1, \dots, N\} \times \Omega \rightarrow \mathbb{R}^d$  be the stochastic process which satisfies for every  $n \in \{0, 1, \dots, N-1\}$  that  $\mathcal{X}_0^{(m)} = \xi^{(m)}$  and

$$\mathcal{X}_{n+1}^{(m)} = \mathcal{X}_n^{(m)} + \mu(\mathcal{X}_n^{(m)}) (t_{n+1} - t_n) + \sigma(\mathcal{X}_n^{(m)}) (W_{t_{n+1}}^{(m)} - W_{t_n}^{(m)}), \quad (111)$$

let  $\mathcal{L}_d: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be the function which satisfies for every  $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$  that

$$\mathcal{L}_d(x) = \left( \frac{\exp(x_1)}{\exp(x_1) + 1}, \frac{\exp(x_2)}{\exp(x_2) + 1}, \dots, \frac{\exp(x_d)}{\exp(x_d) + 1} \right), \quad (112)$$

for every  $k, l \in \mathbb{N}$ ,  $v \in \mathbb{N}_0 = \{0\} \cup \mathbb{N}$ ,  $\theta = (\theta_1, \dots, \theta_\nu) \in \mathbb{R}^\nu$  with  $v + l(k+1) \leq \nu$  let  $A_{k,l}^{\theta, v}: \mathbb{R}^k \rightarrow \mathbb{R}^l$  be the function which satisfies for every  $x = (x_1, \dots, x_k) \in \mathbb{R}^k$  that

$$A_{k,l}^{\theta, v}(x) = \begin{pmatrix} \theta_{v+1} & \theta_{v+2} & \dots & \theta_{v+k} \\ \theta_{v+k+1} & \theta_{v+k+2} & \dots & \theta_{v+2k} \\ \theta_{v+2k+1} & \theta_{v+2k+2} & \dots & \theta_{v+3k} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{v+(l-1)k+1} & \theta_{v+(l-1)k+2} & \dots & \theta_{v+lk} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_k \end{pmatrix} + \begin{pmatrix} \theta_{v+kl+1} \\ \theta_{v+kl+2} \\ \theta_{v+kl+3} \\ \vdots \\ \theta_{v+kl+l} \end{pmatrix}, \quad (113)$$

let  $\mathbb{U}: \mathbb{R}^\nu \times \mathbb{R}^d \rightarrow \mathbb{R}$  be the function which satisfies for every  $\theta \in \mathbb{R}^\nu$ ,  $x \in \mathbb{R}^d$  that

$$\mathbb{U}(\theta, x) = (A_{d,1}^{\theta,(s-1)d(d+1)} \circ \mathcal{L}_d \circ A_{d,d}^{\theta,(s-2)d(d+1)} \circ \dots \circ \mathcal{L}_d \circ A_{d,d}^{\theta,d(d+1)} \circ \mathcal{L}_d \circ A_{d,d}^{\theta,0})(x), \quad (114)$$

and let  $\Theta: \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R}^\nu$  be a stochastic process which satisfies for every  $m \in \mathbb{N}_0$  that

$$\Theta_{m+1} = \Theta_m - 2\gamma \cdot (\mathbb{U}(\Theta_m, \xi^{(m)}) - \varphi(\mathcal{X}_N^{(m)})) \cdot (\nabla_\theta \mathbb{U})(\Theta_m, \xi^{(m)}) \quad (115)$$

Under appropriate hypotheses we think for every sufficiently large  $m \in \mathbb{N}$  and every  $x \in [a, b]^d$  of the random variable  $\mathbb{U}(\Theta_m, x): \Omega \rightarrow \mathbb{R}$  in Framework 2.9 as a suitable approximation  $\mathbb{U}(\Theta_m, x) \approx u(T, x)$  of  $u(T, x) \in \mathbb{R}$  where  $u = u(t, x)_{(t,x) \in [0,T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  is a function with at most polynomially growing partial derivatives which satisfies for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  that  $u(0, x) = \varphi(x)$  and

$$\frac{\partial u}{\partial t}(t, x) = \frac{1}{2} \text{Trace}_{\mathbb{R}^d}(\sigma(x)[\sigma(x)]^*(\text{Hess}_x u)(t, x)) + \langle \mu(x), (\nabla_x u)(t, x) \rangle_{\mathbb{R}^d} \quad (116)$$

(cf. (1) above).

## 2.8 Description of the algorithm in the general case

In this subsection we provide a general framework which covers the approximation method derived in Subsections 2.1–2.6 above and which allows, in addition, to incorporate other minimization algorithms (cf., for example, Kingma & Ba [38], Ruder [53], E et al. [14], Han et al. [24], and Beck et al. [3]) than just the plain vanilla stochastic gradient descent method. The proposed approximation algorithm is an extension of the approximation algorithm in E et al. [14], Han et al. [24], and Beck et al. [3] in the special case of linear Kolmogorov partial differential equations.

**Framework 2.10.** Let  $T \in (0, \infty)$ ,  $N, d, \varrho, \nu, \varsigma \in \mathbb{N}$ , let  $H: [0, T]^2 \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}$  be functions, let  $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t \in [0, T]})$  be a filtered probability space, let  $W^{m,j}: [0, T] \times \Omega \rightarrow \mathbb{R}^d$ ,  $m \in \mathbb{N}_0$ ,  $j \in \mathbb{N}$ , be independent standard  $(\mathbb{F}_t)_{t \in [0, T]}$ -Brownian motions on  $(\Omega, \mathcal{F}, \mathbb{P})$ , let  $\xi^{m,j}: \Omega \rightarrow \mathbb{R}^d$ ,  $m \in \mathbb{N}_0$ ,  $j \in \mathbb{N}$ , be i.i.d.  $\mathbb{F}_0/\mathcal{B}(\mathbb{R}^d)$ -measurable random variables, let  $t_0, t_1, \dots, t_N \in [0, T]$  be real numbers with  $0 = t_0 < t_1 < \dots < t_N = T$ , for every  $\theta \in \mathbb{R}^\nu$ ,  $j \in \mathbb{N}$ ,  $\mathbf{s} \in \mathbb{R}^\varsigma$  let  $\mathbb{U}^{\theta,j,\mathbf{s}}: \mathbb{R}^d \rightarrow \mathbb{R}$  be a function, for every  $m \in \mathbb{N}_0$ ,  $j \in \mathbb{N}$  let  $\mathcal{X}^{m,j} = (\mathcal{X}_n^{m,j})_{n \in \{0,1,\dots,N\}}: \{0, 1, \dots, N\} \times \Omega \rightarrow \mathbb{R}^d$  be a stochastic process which satisfies for every  $n \in \{0, 1, \dots, N-1\}$  that  $\mathcal{X}_0^{m,j} = \xi^{m,j}$  and

$$\mathcal{X}_{n+1}^{m,j} = H(t_n, t_{n+1}, \mathcal{X}_n^{m,j}, W_{t_{n+1}}^{m,j} - W_{t_n}^{m,j}), \quad (117)$$

let  $(J_m)_{m \in \mathbb{N}_0} \subseteq \mathbb{N}$  be a sequence, for every  $m \in \mathbb{N}_0$ ,  $\mathbf{s} \in \mathbb{R}^\varsigma$  let  $\phi^{m,\mathbf{s}}: \mathbb{R}^\nu \times \Omega \rightarrow \mathbb{R}$  be the function which satisfies for every  $(\theta, \omega) \in \mathbb{R}^\nu \times \Omega$  that

$$\phi^{m,\mathbf{s}}(\theta, \omega) = \frac{1}{J_m} \sum_{j=1}^{J_m} [\mathbb{U}^{\theta,j,\mathbf{s}}(\xi^{m,j}(\omega)) - \varphi(\mathcal{X}_N^{m,j}(\omega))]^2, \quad (118)$$

for every  $m \in \mathbb{N}_0$ ,  $\mathbf{s} \in \mathbb{R}^s$  let  $\Psi^{m,\mathbf{s}}: \mathbb{R}^\nu \times \Omega \rightarrow \mathbb{R}^\nu$  be a function which satisfies for every  $\omega \in \Omega$ ,  $\theta \in \{\eta \in \mathbb{R}^\nu: \phi^{m,\mathbf{s}}(\cdot, \omega): \mathbb{R}^\nu \rightarrow \mathbb{R} \text{ is differentiable at } \eta\}$  that

$$\Psi^{m,\mathbf{s}}(\theta, \omega) = (\nabla_\theta \phi^{m,\mathbf{s}})(\theta, \omega), \quad (119)$$

let  $\mathcal{S}: \mathbb{R}^s \times \mathbb{R}^\nu \times (\mathbb{R}^d)^\mathbb{N} \rightarrow \mathbb{R}^s$  be a function, for every  $m \in \mathbb{N}_0$  let  $\Phi_m: \mathbb{R}^e \rightarrow \mathbb{R}^\nu$  and  $\Psi_m: \mathbb{R}^e \times \mathbb{R}^\nu \rightarrow \mathbb{R}^e$  be functions, let  $\Theta: \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R}^\nu$ ,  $\mathbb{S}: \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R}^s$ , and  $\Xi: \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R}^e$  be stochastic processes which satisfy for every  $m \in \mathbb{N}_0$  that

$$\mathbb{S}_{m+1} = \mathcal{S}(\mathbb{S}_m, \Theta_m, (\mathcal{X}_N^{m,i})_{i \in \mathbb{N}}), \quad \Xi_{m+1} = \Psi_m(\Xi_m, \Psi^{m,\mathbb{S}_{m+1}}(\Theta_m)), \quad (120)$$

$$\text{and} \quad \Theta_{m+1} = \Theta_m - \Phi_m(\Xi_{m+1}). \quad (121)$$

Under appropriate hypotheses we think for every sufficiently large  $m \in \mathbb{N}$  and every  $x \in [a, b]^d$  of the random variable  $\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x): \Omega \rightarrow \mathbb{R}$  in Framework 2.10 as a suitable approximation  $\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x) \approx u(T, x)$  of  $u(T, x) \in \mathbb{R}$  where  $u = u(t, x)_{(t,x) \in [0, T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  is a function with at most polynomially growing partial derivatives which satisfies for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  that  $u(0, x) = \varphi(x)$  and

$$\frac{\partial u}{\partial t}(t, x) = \frac{1}{2} \text{Trace}_{\mathbb{R}^d}(\sigma(x)[\sigma(x)]^*(\text{Hess}_x u)(t, x)) + \langle \mu(x), (\nabla_x u)(t, x) \rangle_{\mathbb{R}^d}, \quad (122)$$

where  $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  are sufficiently regular functions (cf. (1) above).

### 3 Examples

In this section we test the proposed approximation algorithm (see Section 2 above) in the case of several examples of SDEs and Kolmogorov PDEs, respectively. In particular, in this section we apply the proposed approximation algorithm to the heat equation (cf. Subsection 3.2 below), to independent geometric Brownian motions (cf. Subsection 3.3 below), to the Black-Scholes model (cf. Subsection 3.4 below), to stochastic Lorenz equations (cf. Subsection 3.5 below), and to the Heston model (cf. Subsection 3.6 below). In the case of each of the examples below we employ the general approximation algorithm in Framework 2.10 above in conjunction with the Adam optimizer (cf. Kingma & Ba [38]) with mini-batches of size 8192 in each iteration step (see Subsection 3.1 below for a precise description). Moreover, we employ a fully-connected feedforward neural network with one input layer, two hidden layers, and one one-dimensional output layer in our implementations in the case of each of these examples. We also use batch normalization (cf. Ioffe & Szegedy [36]) just before the first linear transformation, just before each of the two nonlinear activation functions in front of the hidden layers as well as just after the last linear transformation. For the two nonlinear activation functions we employ the multidimensional version of the function  $\mathbb{R} \ni x \mapsto \tanh(x) \in (-1, 1)$ . All weights in the neural network are initialized by means of the Xavier initialization (cf. Glorot & Bengio [18]). All computations were

performed in single precision (float32) on a NVIDIA GeForce GTX 1080 GPU with 1974 MHz core clock and 8 GB GDDR5X memory with 1809.5 MHz clock rate. The underlying system consisted of an Intel Core i7-6800K CPU with 64 GB DDR4-2133 memory running Tensorflow 1.5 on Ubuntu 16.04.

### 3.1 Setting

**Framework 3.1.** Assume Framework 2.10, let  $\varepsilon = 10^{-8}$ ,  $\beta_1 = \frac{9}{10}$ ,  $\beta_2 = \frac{999}{1000}$ ,  $(\gamma_m)_{m \in \mathbb{N}_0} \subseteq (0, \infty)$ , let  $\text{Pow}_r: \mathbb{R}^\nu \rightarrow \mathbb{R}^\nu$ ,  $r \in (0, \infty)$ , be the functions which satisfy for every  $r \in (0, \infty)$ ,  $x = (x_1, \dots, x_\nu) \in \mathbb{R}^\nu$  that

$$\text{Pow}_r(x) = (|x_1|^r, \dots, |x_\nu|^r), \quad (123)$$

assume for every  $m \in \mathbb{N}_0$ ,  $i \in \{0, 1, \dots, N\}$  that  $J_m = 8192$ ,  $t_i = \frac{iT}{N}$ ,  $\varrho = 2\nu$ ,  $T = 1$ ,  $\gamma_m = 10^{-3} \mathbb{1}_{[0, 250000]}(m) + 10^{-4} \mathbb{1}_{(250000, 500000]}(m) + 10^{-5} \mathbb{1}_{(500000, \infty)}(m)$ , assume for every  $m \in \mathbb{N}_0$ ,  $x = (x_1, \dots, x_\nu)$ ,  $y = (y_1, \dots, y_\nu)$ ,  $\eta = (\eta_1, \dots, \eta_\nu) \in \mathbb{R}^\nu$  that

$$\Psi_m(x, y, \eta) = (\beta_1 x + (1 - \beta_1)\eta, \beta_2 y + (1 - \beta_2)\text{Pow}_2(\eta)) \quad (124)$$

and

$$\psi_m(x, y) = \left( \left[ \sqrt{\frac{|y_1|}{1 - (\beta_2)^m}} + \varepsilon \right]^{-1} \frac{\gamma_m x_1}{1 - (\beta_1)^m}, \dots, \left[ \sqrt{\frac{|y_\nu|}{1 - (\beta_2)^m}} + \varepsilon \right]^{-1} \frac{\gamma_m x_\nu}{1 - (\beta_1)^m} \right). \quad (125)$$

Equations (124) and (125) in Framework 3.1 describe the Adam optimizer (cf. Kingma & Ba [38], e.g., E et al. [24, (32)–(33) in Section 4.2 and (90)–(91) in Section 5.2], and line 84 in PYTHON code 1 in Section 4 below).

### 3.2 Heat equation

In this subsection we apply the proposed approximation algorithm to the heat equation (see (126) below).

Assume Framework 2.10, assume for every  $s, t \in [0, T]$ ,  $x, w \in \mathbb{R}^d$ ,  $m \in \mathbb{N}_0$  that  $N = 1$ ,  $d = 100$ ,  $\nu = d(2d) + (2d)^2 + 2d = 2d(3d + 1)$ ,  $\varphi(x) = \|x\|_{\mathbb{R}^d}^2$ ,  $H(s, t, x, w) = x + \sqrt{2} \text{Id}_{\mathbb{R}^d} w$ , assume that  $\xi^{0,1}: \Omega \rightarrow \mathbb{R}^d$  is continuous uniformly distributed on  $[0, 1]^d$ , and let  $u = (u(t, x))_{(t, x) \in [0, T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  be an at most polynomially growing function which satisfies for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  that  $u(0, x) = \varphi(x)$  and

$$\left( \frac{\partial u}{\partial t} \right)(t, x) = (\Delta_x u)(t, x). \quad (126)$$

Combining, e.g., Lemma 3.2 below with, e.g., Hairer et al. [22, Corollary 4.17 and Remark 4.1] shows that for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  it holds that

$$u(t, x) = \|x\|_{\mathbb{R}^d}^2 + t d. \quad (127)$$

Table 1 approximately presents the relative  $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (128) below), the relative  $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (129) below), and the relative  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (130) below) against  $m \in \{0, 10000, 50000, 100000, 150000, 200000, 500000, 750000\}$  (cf. PYTHON code 2 in Subsection 4.2 below). Figure 1 approximately depicts the relative  $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (128) below), the relative  $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (129) below), and the relative  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (130) below) against  $m \in \{0, 100, 200, 300, \dots, 299800, 299900, 300000\}$  (cf. PYTHON code 2 in Subsection 4.2 below). In our numerical simulations for Table 1 and Figure 1 we calculated the exact solution of the PDE (126) by means of Lemma 3.2 below (see (127) above), we approximately calculated the relative  $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right| dx \quad (128)$$

for  $m \in \{0, 10000, 50000, 100000, 150000, 200000, 500000, 750000\}$  by means of Monte Carlo approximations with 10240000 samples, we approximately calculated the relative  $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\sqrt{\int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right|^2 dx} \quad (129)$$

for  $m \in \{0, 10000, 50000, 100000, 150000, 200000, 500000, 750000\}$  by means of Monte Carlo approximations with 10240000 samples, and we approximately calculated the relative  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\sup_{x \in [0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right| \quad (130)$$

for  $m \in \{0, 10000, 50000, 100000, 150000, 200000, 500000, 750000\}$  by means of Monte Carlo approximations with 10240000 samples (see Lemma 3.5 below). Table 2 approximately presents the relative  $L^1(\mathbb{P}; L^1(\lambda_{[0,1]^d}; \mathbb{R}))$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (131) below), the relative  $L^2(\mathbb{P}; L^2(\lambda_{[0,1]^d}; \mathbb{R}))$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (132) below), and the relative  $L^2(\mathbb{P}; L^\infty(\lambda_{[0,1]^d}; \mathbb{R}))$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (133) below), against  $m \in \{0, 10000, 50000, 100000, 150000, 200000, 500000, 750000\}$  (cf. PYTHON code 2 in Subsection 4.2 below). In our numerical simulations for Table 2 we calculated the exact solution of the PDE (126) by means of Lemma 3.2 below (see (127) above), we approximately calculated the relative  $L^1(\mathbb{P}; L^1(\lambda_{[0,1]^d}; \mathbb{R}))$ -approximation error

$$\mathbb{E} \left[ \int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right| dx \right] \quad (131)$$

for  $m \in \{0, 10000, 50000, 100000, 150000, 200000, 500000, 750000\}$  by means of Monte Carlo approximations with 10240000 samples for the Lebesgue integral and 5 samples for the expectation, we approximately calculated the relative  $L^2(\mathbb{P}; L^2(\lambda_{[0,1]^d}; \mathbb{R}))$ -approximation error

$$\left( \mathbb{E} \left[ \int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right|^2 dx \right] \right)^{1/2} \quad (132)$$

for  $m \in \{0, 10000, 50000, 100000, 150000, 200000, 500000, 750000\}$  by means of Monte Carlo approximations with 10240000 samples for the Lebesgue integral and 5 samples for the expectation, and we approximately calculated the relative  $L^2(\mathbb{P}; L^\infty(\lambda_{[0,1]^d}; \mathbb{R}))$ -approximation error

$$\left( \mathbb{E} \left[ \sup_{x \in [0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right|^2 \right] \right)^{1/2} \quad (133)$$

for  $m \in \{0, 10000, 50000, 100000, 150000, 200000, 500000, 750000\}$  by means of Monte Carlo approximations with 10240000 samples for the supremum (see Lemma 3.5 below) and 5 samples for the expectation. The following elementary result, Lemma 3.2 below, specifies the explicit solution of the PDE (126) above (cf. (127) above). For completeness we also provide here a proof for Lemma 3.2.

Number of steps	Relative $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Relative $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Relative $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Runtime in seconds
0	0.998253	0.998254	1.003524	0.5
10000	0.957464	0.957536	0.993083	44.6
50000	0.786743	0.786806	0.828184	220.8
100000	0.574013	0.574060	0.605283	440.8
150000	0.361564	0.361594	0.384105	661.0
200000	0.150346	0.150362	0.164140	880.8
500000	0.000882	0.001112	0.007360	2200.7
750000	0.000822	0.001036	0.007423	3300.6

Table 1: Approximative presentations of the relative approximation errors in (128)–(130) for the heat equation in (126).

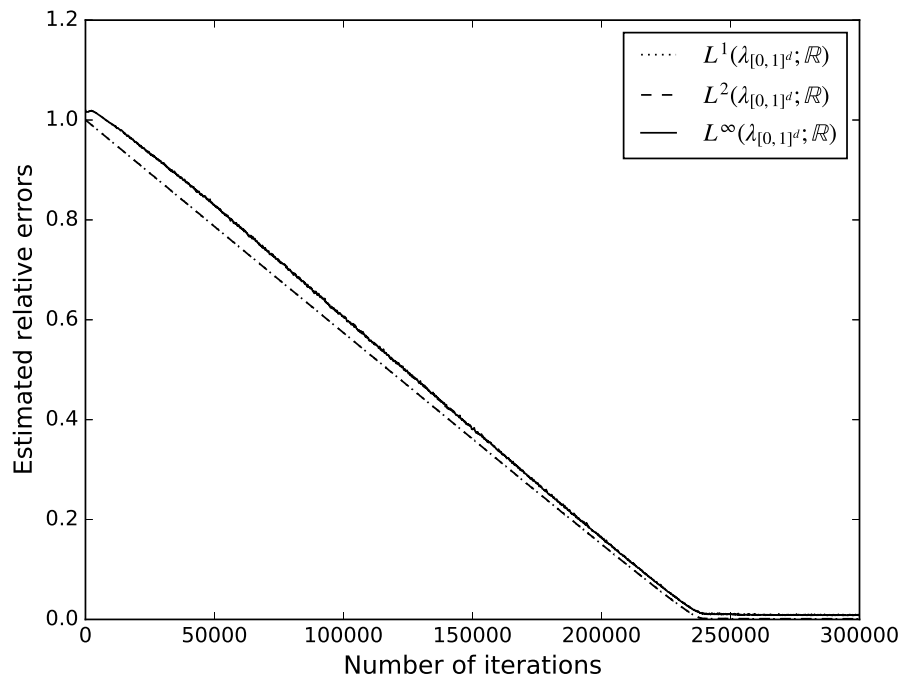


Figure 1: Approximative plots of the relative approximation errors in (128)–(130) for the heat equation in (126).

Number of steps	Relative $L^1(\mathbb{P}; L^1(\lambda_{[0,1]^d}; \mathbb{R}))$ -error	Relative $L^2(\mathbb{P}; L^2(\lambda_{[0,1]^d}; \mathbb{R}))$ -error	Relative $L^2(\mathbb{P}; L^\infty(\lambda_{[0,1]^d}; \mathbb{R}))$ -error	Mean runtime in seconds
0	1.000310	1.000311	1.005674	0.6
10000	0.957481	0.957554	0.993097	44.7
50000	0.786628	0.786690	0.828816	220.4
100000	0.573867	0.573914	0.605587	440.5
150000	0.361338	0.361369	0.382967	660.8
200000	0.001387	0.001741	0.010896	880.9
500000	0.000883	0.001112	0.008017	2201.0
750000	0.000822	0.001038	0.007547	3300.4

Table 2: Approximative presentations of the relative approximation errors in (131)–(133) for the heat equation in (126).

**Lemma 3.2.** *Let  $T \in (0, \infty)$ ,  $d \in \mathbb{N}$ , let  $C \in \mathbb{R}^{d \times d}$  be a strictly positive and symmetric matrix, and let  $u: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  be the function which satisfies for every  $(t, x) \in [0, T] \times \mathbb{R}^d$  that*

$$u(t, x) = \|x\|_{\mathbb{R}^d}^2 + t \operatorname{Trace}_{\mathbb{R}^d}(C). \quad (134)$$

Then

(i) *it holds that  $u \in C^\infty([0, T] \times \mathbb{R}^d, \mathbb{R})$  is at most polynomially growing and*

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

$$\left(\frac{\partial u}{\partial t}\right)(t, x) = \frac{1}{2} \operatorname{Trace}_{\mathbb{R}^d}(C(\operatorname{Hess}_x u)(t, x)). \quad (135)$$

*Proof of Lemma 3.2.* First, note that  $u$  is a polynomial. This establishes item (i). Moreover, note that for every  $(t, x) \in [0, T] \times \mathbb{R}^d$  it holds that

$$\left(\frac{\partial u}{\partial t}\right)(t, x) = \operatorname{Trace}_{\mathbb{R}^d}(C), \quad (\nabla_x u)(t, x) = 2x, \quad (136)$$

$$\text{and} \quad (\operatorname{Hess}_x u)(t, x) = \left(\frac{\partial}{\partial x}(\nabla_x u)\right)(t, x) = 2 \operatorname{Id}_{\mathbb{R}^d}. \quad (137)$$

Hence, we obtain for every  $(t, x) \in [0, T] \times \mathbb{R}^d$  that

$$\left(\frac{\partial u}{\partial t}\right)(t, x) - \frac{1}{2} \operatorname{Trace}_{\mathbb{R}^d}(C(\operatorname{Hess}_x u)(t, x)) = \operatorname{Trace}_{\mathbb{R}^d}(C) - \frac{1}{2} \operatorname{Trace}_{\mathbb{R}^d}(2C) = 0. \quad (138)$$

This proves item (ii). The proof of Lemma 3.2 is thus completed.  $\square$



Lemma 3.5 below discloses the strategy how we approximatively calculate the  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -errors in (130) and (133) above. Our proof of Lemma 3.5 employs the elementary auxiliary results in Lemma 3.3 and Lemma 3.4 below. For completeness we also include proofs for Lemma 3.3 and Lemma 3.4 here.

**Lemma 3.3.** *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and let  $A, \mathcal{O} \in \mathcal{F}$  satisfy that  $\mathbb{P}(\mathcal{O}) = 1$ . Then it holds that*

$$\mathbb{P}(A) = \mathbb{P}(A \cap \mathcal{O}). \quad (139)$$

*Proof of Lemma 3.3.* Observe that the monotonicity of  $\mathbb{P}$  ensures that

$$\begin{aligned} \mathbb{P}(A \cap \mathcal{O}) &\leq \mathbb{P}(A) = \mathbb{P}([A \cap \mathcal{O}] \cup [A \setminus (A \cap \mathcal{O})]) \\ &= \mathbb{P}(A \cap \mathcal{O}) + \mathbb{P}(A \setminus (A \cap \mathcal{O})) \\ &= \mathbb{P}(A \cap \mathcal{O}) + \mathbb{P}(A \setminus \mathcal{O}) \\ &\leq \mathbb{P}(A \cap \mathcal{O}) + \mathbb{P}(\Omega \setminus \mathcal{O}) \\ &= \mathbb{P}(A \cap \mathcal{O}) + \mathbb{P}(\Omega) - \mathbb{P}(\mathcal{O}) = \mathbb{P}(A \cap \mathcal{O}). \end{aligned} \quad (140)$$

Hence, we obtain that for every  $A, \mathcal{O} \in \mathcal{F}$  with  $\mathbb{P}(\mathcal{O}) = 1$  it holds that  $\mathbb{P}(A) = \mathbb{P}(A \cap \mathcal{O})$ . The Proof of Lemma 3.3 is thus completed.  $\square$

**Lemma 3.4.** *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $X_n: \Omega \rightarrow [0, \infty)$ ,  $n \in \mathbb{N}_0$ , be random variables, assume for every  $n \in \mathbb{N}$  that  $\mathbb{P}(X_n \geq X_{n+1}) = 1$ , and assume for every  $\varepsilon \in (0, \infty)$  that*

$$\limsup_{n \rightarrow \infty} \mathbb{P}(X_n > \varepsilon) = 0. \quad (141)$$

*Then*

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} X_n = 0\right) = 1. \quad (142)$$

*Proof of Lemma 3.4.* Throughout this proof let  $\mathcal{O} \subseteq \Omega$  be the set given by

$$\mathcal{O} = \bigcap_{n=1}^{\infty} \{X_n \geq X_{n+1}\} = \{\forall n \in \mathbb{N}: X_n \geq X_{n+1}\}. \quad (143)$$

Observe that Lemma 3.3 and the hypothesis that for every  $n \in \mathbb{N}$  it holds that  $\mathbb{P}(X_n \geq X_{n+1}) = 1$  assure that for every  $N \in \mathbb{N}$  it holds that

$$\begin{aligned} \mathbb{P}(\bigcap_{n=1}^N \{X_n \geq X_{n+1}\}) &= \mathbb{P}([\bigcap_{n=1}^{N-1} \{X_n \geq X_{n+1}\}] \cap \{X_N \geq X_{N+1}\}) \\ &= \mathbb{P}(\bigcap_{n=1}^{N-1} \{X_n \geq X_{n+1}\}). \end{aligned} \quad (144)$$

This implies that for every  $N \in \mathbb{N}$  it holds that

$$\begin{aligned} \mathbb{P}(\bigcap_{n=1}^N \{X_n \geq X_{n+1}\}) &= \mathbb{P}(\bigcap_{n=1}^0 \{X_n \geq X_{n+1}\}) \\ &= \mathbb{P}(\Omega) = 1. \end{aligned} \quad (145)$$

The fact that the measure  $\mathbb{P}$  is continuous from above hence demonstrates that

$$\begin{aligned}\mathbb{P}(\mathcal{O}) &= \mathbb{P}(\cap_{n=1}^{\infty} \{X_n \geq X_{n+1}\}) = \mathbb{P}(\cap_{N=1}^{\infty} [\cap_{n=1}^N \{X_n \geq X_{n+1}\}]) \\ &= \lim_{N \rightarrow \infty} \mathbb{P}(\cap_{n=1}^N \{X_n \geq X_{n+1}\}) = 1.\end{aligned}\tag{146}$$

Next note that

$$\begin{aligned}\mathbb{P}\left(\limsup_{n \rightarrow \infty} X_n > 0\right) &= \mathbb{P}\left(\exists k \in \mathbb{N}: \left[\limsup_{n \rightarrow \infty} X_n > \frac{1}{k}\right]\right) \\ &= \mathbb{P}\left(\exists k \in \mathbb{N}: \forall m \in \mathbb{N}: \exists n \in \mathbb{N} \cap [m, \infty): [X_n > \frac{1}{k}]\right) \\ &= \mathbb{P}\left(\cup_{k \in \mathbb{N}} \cap_{m \in \mathbb{N}} \cup_{n \in \mathbb{N} \cap [m, \infty)} \{X_n > \frac{1}{k}\}\right) \\ &\leq \sum_{k=1}^{\infty} \mathbb{P}\left(\cap_{m=1}^{\infty} \cup_{n=m}^{\infty} \{X_n > \frac{1}{k}\}\right) \\ &= \sum_{k=1}^{\infty} \left[\limsup_{m \rightarrow \infty} \mathbb{P}(\cup_{n=m}^{\infty} \{X_n > \frac{1}{k}\})\right].\end{aligned}\tag{147}$$

Lemma 3.3 and (146) therefore ensure that

$$\begin{aligned}\mathbb{P}\left(\limsup_{n \rightarrow \infty} X_n > 0\right) &\leq \sum_{k=1}^{\infty} \left[\limsup_{m \rightarrow \infty} \mathbb{P}\left([\cup_{n=m}^{\infty} \{X_n > \frac{1}{k}\}] \cap \mathcal{O}\right)\right] \\ &\leq \sum_{k=1}^{\infty} \left[\limsup_{m \rightarrow \infty} \mathbb{P}\left([\cup_{n=m}^{\infty} \{X_n > \frac{1}{k}\}] \cap \{\forall n \in \mathbb{N} \cap [m, \infty): X_m \geq X_n\}\right)\right] \\ &= \sum_{k=1}^{\infty} \left[\limsup_{m \rightarrow \infty} \mathbb{P}\left(\{X_m > \frac{1}{k}\} \cap \{\forall n \in \mathbb{N} \cap [m, \infty): X_m \geq X_n\}\right)\right] \\ &\leq \sum_{k=1}^{\infty} \left[\limsup_{m \rightarrow \infty} \mathbb{P}(X_m > \frac{1}{k})\right].\end{aligned}\tag{148}$$

Combining this and (141) establishes (142). The proof of Lemma 3.4 is thus completed.  $\square$

**Lemma 3.5.** *Let  $d \in \mathbb{N}$ ,  $a \in \mathbb{R}$ ,  $b \in (a, \infty)$ , let  $f: [a, b]^d \rightarrow \mathbb{R}$  be a continuous function, let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $X_n: \Omega \rightarrow [a, b]^d$ ,  $n \in \mathbb{N}$ , be i.i.d. random variables, and assume that  $X_1$  is continuous uniformly distributed on  $[a, b]^d$ . Then*

(i) it holds that

$$\mathbb{P} \left( \limsup_{N \rightarrow \infty} \left| \left[ \max_{1 \leq n \leq N} f(X_n) \right] - \left[ \sup_{x \in [a, b]^d} f(x) \right] \right| = 0 \right) = 1 \quad (149)$$

and

(ii) it holds for every  $p \in (0, \infty)$  that

$$\limsup_{N \rightarrow \infty} \mathbb{E} \left[ \left| \left[ \max_{1 \leq n \leq N} f(X_n) \right] - \left[ \sup_{x \in [a, b]^d} f(x) \right] \right|^p \right] = 0. \quad (150)$$

*Proof of Lemma 3.5.* First, observe that the fact that  $f: [a, b]^d \rightarrow \mathbb{R}$  is a continuous function and the fact that  $[a, b]^d \subseteq \mathbb{R}^d$  is a compact set demonstrate that there exists  $\xi \in [a, b]^d$  which satisfies that  $f(\xi) = \sup_{x \in [a, b]^d} f(x)$ . Next note that the fact that for every  $N \in \mathbb{N}$ ,  $n \in \{1, 2, \dots, N\}$  it holds that  $f(X_n) \leq \sup_{x \in [a, b]^d} f(x)$  implies that for every  $N \in \mathbb{N}$  it holds that  $\max_{1 \leq n \leq N} f(X_n) \leq \sup_{x \in [a, b]^d} f(x)$ . Hence, we obtain that for every  $N \in \mathbb{N}$  it holds that

$$\left| \left[ \max_{1 \leq n \leq N} f(X_n) \right] - \left[ \sup_{x \in [a, b]^d} f(x) \right] \right| = \left[ \sup_{x \in [a, b]^d} f(x) \right] - \left[ \max_{1 \leq n \leq N} f(X_n) \right]. \quad (151)$$

Combining this with the fact that  $f(\xi) = \sup_{x \in \mathbb{R}^d} f(x)$  ensures that for every  $\varepsilon \in (0, \infty)$ ,  $N \in \mathbb{N}$  it holds that

$$\begin{aligned} & \left\{ \left| \max_{1 \leq n \leq N} f(X_n) - \sup_{x \in [a, b]^d} f(x) \right| \leq \varepsilon \right\} \\ &= \left\{ \max_{1 \leq n \leq N} f(X_n) \geq \sup_{x \in [a, b]^d} f(x) - \varepsilon \right\} \\ &= \bigcup_{n=1}^N \left\{ f(X_n) \geq \sup_{x \in [a, b]^d} f(x) - \varepsilon \right\} = \bigcup_{n=1}^N \left\{ |f(X_n) - \sup_{x \in [a, b]^d} f(x)| \leq \varepsilon \right\} \\ &= \bigcup_{n=1}^N \left\{ |f(X_n) - f(\xi)| \leq \varepsilon \right\}. \end{aligned} \quad (152)$$

In the next step we observe that the fact that  $f: [a, b]^d \rightarrow \mathbb{R}$  is continuous ensures that for every  $\varepsilon \in (0, \infty)$  there exists  $\delta \in (0, \infty)$  such that for every  $x \in [a, b]^d$  with  $\|x - \xi\|_{\mathbb{R}^d} \leq \delta$  it holds that  $|f(x) - f(\xi)| \leq \varepsilon$ . Combining this and (152) shows that for every  $\varepsilon \in (0, \infty)$  there exists  $\delta \in (0, \infty)$  such that for every  $N \in \mathbb{N}$  it holds that

$$\begin{aligned} & \mathbb{P} \left( \left| \max_{1 \leq n \leq N} f(X_n) - \sup_{x \in [a, b]^d} f(x) \right| \leq \varepsilon \right) = \mathbb{P} \left( \bigcup_{n=1}^N \left\{ |f(X_n) - f(\xi)| \leq \varepsilon \right\} \right) \\ & \geq \mathbb{P} \left( \bigcup_{n=1}^N \left\{ \|X_n - \xi\|_{\mathbb{R}^d} \leq \delta \right\} \right) = 1 - \mathbb{P} \left( \bigcap_{n=1}^N \left\{ \|X_n - \xi\|_{\mathbb{R}^d} > \delta \right\} \right). \end{aligned} \quad (153)$$

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

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \mathbb{P} \left( \left| \max_{1 \leq n \leq N} f(X_n) - \sup_{x \in [a, b]^d} f(x) \right| \leq \varepsilon \right) \\ & \geq 1 - \liminf_{N \rightarrow \infty} \mathbb{P} \left( \bigcap_{n=1}^N \left\{ \|X_n - \xi\|_{\mathbb{R}^d} > \delta \right\} \right). \end{aligned} \quad (154)$$

Next observe that the fact that the random variables  $X_n: \Omega \rightarrow [a, b]^d$ ,  $n \in \{1, 2, \dots, N\}$ , are i.i.d. ensures that for every  $\delta \in (0, \infty)$ ,  $N \in \mathbb{N}$  it holds that

$$\begin{aligned} \mathbb{P}\left(\bigcap_{n=1}^N \{\|X_n - \xi\|_{\mathbb{R}^d} > \delta\}\right) &= \prod_{n=1}^N \mathbb{P}(\|X_n - \xi\|_{\mathbb{R}^d} > \delta) \\ &= [\mathbb{P}(\|X_1 - \xi\|_{\mathbb{R}^d} > \delta)]^N. \end{aligned} \quad (155)$$

In addition, note that the fact that for every  $\delta \in (0, \infty)$  it holds that the set  $\{x \in [a, b]^d: \|x - \xi\|_{\mathbb{R}^d} \leq \delta\} \subseteq \mathbb{R}^d$  has strictly positive  $d$ -dimensional Lebesgue measure and the fact that  $X_1$  is continuous uniformly distributed on  $[a, b]^d$  ensure that for every  $\delta \in (0, \infty)$  it holds that

$$\mathbb{P}(\|X_1 - \xi\|_{\mathbb{R}^d} > \delta) = 1 - \mathbb{P}(\|X_1 - \xi\|_{\mathbb{R}^d} \leq \delta) < 1. \quad (156)$$

Hence, we obtain that for every  $\delta \in (0, \infty)$  it holds that

$$\limsup_{N \rightarrow \infty} \left( [\mathbb{P}(\|X_1 - \xi\|_{\mathbb{R}^d} > \delta)]^N \right) = 0. \quad (157)$$

Combining this with (155) demonstrates that for every  $\delta \in (0, \infty)$  it holds that

$$\limsup_{N \rightarrow \infty} \mathbb{P}\left(\bigcap_{n=1}^N \{\|X_n - \xi\|_{\mathbb{R}^d} > \delta\}\right) = 0. \quad (158)$$

This and (154) assure that for every  $\varepsilon \in (0, \infty)$  it holds that

$$\liminf_{N \rightarrow \infty} \mathbb{P}\left(|\max_{1 \leq n \leq N} f(X_n) - \sup_{x \in [a, b]^d} f(x)| \leq \varepsilon\right) = 1. \quad (159)$$

Therefore, we obtain that for every  $\varepsilon \in (0, \infty)$  it holds that

$$\limsup_{N \rightarrow \infty} \mathbb{P}\left(|\max_{1 \leq n \leq N} f(X_n) - \sup_{x \in [a, b]^d} f(x)| > \varepsilon\right) = 0. \quad (160)$$

Combining this with with Lemma 3.4 establishes item (i). It thus remains to prove item (ii). For this note that the fact that  $f: [a, b]^d \rightarrow \mathbb{R}$  is globally bounded, item (i), and Lebesgue's dominated convergence theorem ensure that for every  $p \in (0, \infty)$  it holds that

$$\limsup_{N \rightarrow \infty} \mathbb{E}\left[\left|\max_{1 \leq i \leq N} f(X_i) - \sup_{x \in [a, b]^d} f(x)\right|^p\right] = 0. \quad (161)$$

This establishes item (ii). The proof of Lemma 3.5 is thus completed.  $\square$

### 3.3 Geometric Brownian motions

In this subsection we apply the proposed approximation algorithm to a Black-Scholes PDE with independent underlying geometric Brownian motions.

Assume Framework 2.10, let  $r = \frac{1}{20}$ ,  $\mu = r - \frac{1}{10} = -\frac{1}{20}$ ,  $\sigma_1 = \frac{1}{10} + \frac{1}{200}$ ,  $\sigma_2 = \frac{1}{10} + \frac{2}{200}, \dots, \sigma_{100} = \frac{1}{10} + \frac{100}{200}$ , assume for every  $s, t \in [0, T]$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $w = (w_1, w_2, \dots, w_d) \in \mathbb{R}^d$ ,  $m \in \mathbb{N}_0$  that  $N = 1$ ,  $d = 100$ ,  $\varphi(x) = \exp(-rT) \max\{\max_{i \in \{1, 2, \dots, d\}} x_i - 100, 0\}$ , and

$$H(s, t, x, w) = \left( x_1 \exp\left(\left(\mu_1 - \frac{|\sigma_1|^2}{2}\right)(t-s) + \sigma_1 w_1\right), \dots, x_d \exp\left(\left(\mu_d - \frac{|\sigma_d|^2}{2}\right)(t-s) + \sigma_d w_d\right) \right), \quad (162)$$

assume that  $\xi^{0,1}: \Omega \rightarrow \mathbb{R}^d$  is continuous uniformly distributed on  $[90, 110]^d$ , and let  $u = (u(t, x))_{(t,x) \in [0, T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  be an at most polynomially growing function which satisfies for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  that  $u(0, x) = \varphi(x)$  and

$$\left(\frac{\partial u}{\partial t}\right)(t, x) = \frac{1}{2} \sum_{i=1}^d |\sigma_i x_i|^2 \left(\frac{\partial^2 u}{\partial x_i^2}\right)(t, x) + \sum_{i=1}^d \mu_i x_i \left(\frac{\partial u}{\partial x_i}\right)(t, x). \quad (163)$$

The Feynman-Kac formula (cf., for example, Hairer et al. [22, Corollary 4.17]) shows that for every standard Brownian motion  $\mathcal{W} = (\mathcal{W}^{(1)}, \dots, \mathcal{W}^{(d)}): [0, T] \times \Omega \rightarrow \mathbb{R}^d$  and every  $t \in [0, T]$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$u(t, x) = \mathbb{E} \left[ \varphi \left( x_1 \exp \left( \sigma_1 \mathcal{W}_t^{(1)} + \left( \mu_1 - \frac{|\sigma_1|^2}{2} \right) t \right), \dots, x_d \exp \left( \sigma_d \mathcal{W}_t^{(d)} + \left( \mu_d - \frac{|\sigma_d|^2}{2} \right) t \right) \right) \right]. \quad (164)$$

Table 3 approximately presents the relative  $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (165) below), the relative  $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (166) below), and the relative  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (167) below) against  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  (cf. PYTHON code 3 in Subsection 4.3 below). In our numerical simulations for Table 3 we approximately calculated the exact solution of the PDE (163) by means of (164) and Monte Carlo approximations with 1048576 samples, we approximately calculated the relative  $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right| dx \quad (165)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo approximations with 81920 samples, we approximately calculated the relative  $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\sqrt{\int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right|^2 dx} \quad (166)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo approximations with 81920 samples, and we approximately calculated the relative  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\sup_{x \in [0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, S_m}(x)}{u(T, x)} \right| \quad (167)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo approximations with 81920 samples (see Lemma 3.5 above).

Number of steps	Relative $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Relative $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Relative $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Runtime in seconds
0	1.004285	1.004286	1.009524	1
25000	0.842938	0.843021	0.87884	110.2
50000	0.684955	0.685021	0.719826	219.5
100000	0.371515	0.371551	0.387978	437.9
150000	0.064605	0.064628	0.072259	656.2
250000	0.001220	0.001538	0.010039	1092.6
500000	0.000949	0.001187	0.005105	2183.8
750000	0.000902	0.001129	0.006028	3275.1

Table 3: Approximative presentations of the relative approximation errors in (165)–(167) for the Black-Scholes PDE with independent underlying geometric Brownian motions in (163).

### 3.4 Black-Scholes model with correlated noise

In this subsection we apply the proposed approximation algorithm to a Black-Scholes PDE with correlated noise.

Assume Framework 2.10, let  $r = \frac{1}{20}$ ,  $\mu = r - \frac{1}{10} = -\frac{1}{20}$ ,  $\beta_1 = \frac{1}{10} + \frac{1}{200}$ ,  $\beta_2 = \frac{1}{10} + \frac{2}{200}, \dots, \beta_{100} = \frac{1}{10} + \frac{100}{200}$ ,  $Q = (Q_{i,j})_{(i,j) \in \{1,2,\dots,100\}}$ ,  $\Sigma = (\Sigma_{i,j})_{(i,j) \in \{1,2,\dots,100\}} \in \mathbb{R}^{100 \times 100}$ ,  $\varsigma_1, \varsigma_2, \dots, \varsigma_{100} \in \mathbb{R}^{100}$ , assume for every  $s, t \in [0, T]$ ,  $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ ,  $w = (w_1, w_2, \dots, w_d) \in \mathbb{R}^d$ ,  $m \in \mathbb{N}_0$ ,  $i, j, k \in \{1, 2, \dots, 100\}$  with  $i < j$  that  $N = 1$ ,  $d = 100$ ,  $\nu = d(2d) + (2d)^2 + 2d = 2d(3d + 1)$ ,  $Q_{k,k} = 1$ ,  $Q_{i,j} = Q_{j,i} = \frac{1}{2}$ ,  $\Sigma_{i,j} = 0$ ,  $\Sigma_{k,k} > 0$ ,  $\Sigma \Sigma^* = Q$  (cf., for example, Golub & Van Loan [19, Theorem 4.2.5]),  $\varsigma_k = (\Sigma_{k,1}, \dots, \Sigma_{k,100})$ ,  $\varphi(x) = \exp(-\mu T) \max\{110 - [\min_{i \in \{1,2,\dots,d\}} x_i], 0\}$ , and

$$H(s, t, x, w) = \left( x_1 \exp\left(\left(\mu - \frac{1}{2} \|\beta_1 \varsigma_1\|_{\mathbb{R}^d}^2\right)(t - s) + \langle \varsigma_1, w \rangle_{\mathbb{R}^d}\right), \dots, \right. \\ \left. x_d \exp\left(\left(\mu - \frac{1}{2} \|\beta_d \varsigma_d\|_{\mathbb{R}^d}^2\right)(t - s) + \langle \varsigma_d, w \rangle_{\mathbb{R}^d}\right) \right), \quad (168)$$

assume that  $\xi^{0,1}: \Omega \rightarrow \mathbb{R}^d$  is continuous uniformly distributed on  $[90, 110]^d$ , and let  $u = (u(t, x))_{t \in [0, T], x \in \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  be an at most polynomially growing continuous function which satisfies for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  that  $u(0, x) = \varphi(x)$  and

$$\left(\frac{\partial u}{\partial t}\right)(t, x) = \frac{1}{2} \sum_{i,j=1}^d x_i x_j \beta_i \beta_j \langle \varsigma_i, \varsigma_j \rangle_{\mathbb{R}^d} \left(\frac{\partial^2 u}{\partial x_i^2}\right)(t, x) + \sum_{i=1}^d \mu_i x_i \left(\frac{\partial u}{\partial x_i}\right)(t, x). \quad (169)$$

The Feynman-Kac formula (cf., for example, Hairer et al. [22, Corollary 4.17]) shows that for every standard Brownian motion  $\mathcal{W} = (\mathcal{W}^{(1)}, \dots, \mathcal{W}^{(d)}): [0, T] \times \Omega \rightarrow \mathbb{R}^d$  and every  $t \in [0, T]$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$u(t, x) = \mathbb{E} \left[ \varphi \left( x_1 \exp \left( \langle \varsigma_1, \mathcal{W}_t^{(1)} \rangle_{\mathbb{R}^d} + \left( \mu_1 - \frac{\|\beta_1 \varsigma_1\|_{\mathbb{R}^d}^2}{2} \right) t \right), \dots, \right. \right. \\ \left. \left. x_d \exp \left( \langle \varsigma_d, \mathcal{W}_t^{(d)} \rangle_{\mathbb{R}^d} + \left( \mu_d - \frac{\|\beta_d \varsigma_d\|_{\mathbb{R}^d}^2}{2} \right) t \right) \right) \right]. \quad (170)$$

Table 4 approximately presents the relative  $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (171) below), the relative  $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (172) below), and the relative  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (173) below) against  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  (cf. PYTHON code 4 in Subsection 3.4 below). In our numerical simulations for Table 4 we approximately calculated the exact solution of the PDE (169) by means of (170) and Monte Carlo approximations with 1048576 samples, we approximately calculated the relative  $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right| dx \quad (171)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo approximations with 81920 samples, we approximately calculated the relative  $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\sqrt{\int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right|^2 dx} \quad (172)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo approximations with 81920 samples, and we approximately calculated the relative  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\sup_{x \in [0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right| \quad (173)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo approximations with 81920 samples (see Lemma 3.5 above).

Number of steps	Relative $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Relative $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Relative $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Runtime in seconds
0	1.003383	1.003385	1.011662	0.8
25000	0.631420	0.631429	0.640633	112.1
50000	0.269053	0.269058	0.275114	223.3
100000	0.000752	0.000948	0.00553	445.8
150000	0.000694	0.00087	0.004662	668.2
250000	0.000604	0.000758	0.006483	1119.3
500000	0.000493	0.000615	0.002774	2292.8
750000	0.000471	0.00059	0.002862	3466.8

Table 4: Approximative presentations of the relative approximation errors in (171)–(173) for the Black-Scholes PDE with correlated noise in (169).

### 3.5 Stochastic Lorenz equations

In this subsection we apply the proposed approximation algorithm to the stochastic Lorenz equation.

Assume Framework 2.10, let  $\alpha_1 = 10$ ,  $\alpha_2 = 14$ ,  $\alpha_3 = \frac{8}{3}$ ,  $\beta = \frac{3}{20}$ , let  $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be a function, assume for every  $s, t \in [0, T]$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $w = (w_1, w_2, \dots, w_d) \in \mathbb{R}^d$ ,  $m \in \mathbb{N}_0$  that  $N = 100$ ,  $d = 3$ ,  $\nu = (d + 20)d + (d + 20)^2 + (d + 20) = (d + 20)(2d + 21)$ ,  $\mu(x) = (\alpha_1(x_2 - x_1), \alpha_2 x_1 - x_2 - x_1 x_3, x_1 x_2 - \alpha_3 x_3)$ ,  $\varphi(x) = \|x\|_{\mathbb{R}^d}^2$ , and

$$H(s, t, x, w) = x + \mu(x)(t - s)\mathbb{1}_{[0, N/T]}(\|\mu(x)\|_{\mathbb{R}^d}) + \beta w \quad (174)$$

(cf., for example, Hutzenthaler et al. [32], Hutzenthaler et al. [33], Hutzenthaler et al. [34], Milstein & Tretyakov [48], Sabanis [54, 55], and the references mentioned therein for related temporal numerical approximation schemes for SDEs), assume that  $\xi^{0,1}: \Omega \rightarrow \mathbb{R}^d$  is continuous uniformly distributed on  $[\frac{1}{2}, \frac{3}{2}] \times [8, 10] \times [10, 12]$ , and let  $u = (u(t, x))_{(t,x) \in [0, T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  be an at most polynomially growing function (cf., for example, Hairer et al. [22, Corollary 4.17] and Hörmander [31, Theorem 1.1]) which satisfies for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  that  $u(0, x) = \varphi(x)$  and

$$\begin{aligned} \left(\frac{\partial u}{\partial t}\right)(t, x) &= \frac{\beta^2}{2}(\Delta_x u)(t, x) + \alpha_1(x_2 - x_1)\left(\frac{\partial u}{\partial x_1}\right)(t, x) \\ &+ (\alpha_2 x_1 - x_2 - x_1 x_3)\left(\frac{\partial u}{\partial x_2}\right)(t, x) + (x_1 x_2 - \alpha_3 x_3)\left(\frac{\partial u}{\partial x_3}\right)(t, x). \end{aligned} \quad (175)$$

Table 5 approximately presents the relative  $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0, 1]^d}$  (see (176) below), the relative  $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0, 1]^d}$  (see (177) below), and the relative  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0, 1]^d}$  (see (178) below) against  $m \in \{0, 25000, 50000,$



100000, 150000, 250000, 500000, 750000} (cf. PYTHON code 5 in Subsection 4.5 below). In our numerical simulations for Table 5 we approximately calculated the exact solution of the PDE (175) by means of Monte Carlo approximations with 104857 samples and temporal SDE-discretizations based on (174) and 100 equidistant time steps, we approximately calculated the relative  $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right| dx \quad (176)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo approximations with 20480 samples, we approximately calculated the relative  $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\sqrt{\int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right|^2 dx} \quad (177)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo approximations with 20480 samples, and we approximately calculated the relative  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\sup_{x \in [0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right| \quad (178)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo approximations with 20480 samples (see Lemma 3.5 above).

Number of steps	Relative $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Relative $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Relative $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Runtime in seconds
0	0.995732	0.995732	0.996454	1.0
25000	0.905267	0.909422	1.247772	750.1
50000	0.801935	0.805497	1.115690	1461.7
100000	0.599847	0.602630	0.823042	2932.1
150000	0.392394	0.394204	0.542209	4423.3
250000	0.000732	0.000811	0.002865	7327.9
500000	0.000312	0.000365	0.003158	14753.0
750000	0.000187	0.000229	0.001264	21987.4

Table 5: Approximative presentations of the relative approximation errors in (176)–(178) for the stochastic Lorenz equation in (175).

### 3.6 Heston model

In this subsection we apply the proposed approximation algorithm to the Heston model in (180) below.

Assume Framework 2.10, let  $\delta = 25$ ,  $\alpha = \frac{1}{20}$ ,  $\kappa = \frac{6}{10}$ ,  $\theta = \frac{1}{25}$ ,  $\beta = \frac{1}{5}$ ,  $\varrho = -\frac{4}{5}$ , let  $e_i \in \mathbb{R}^{50}$ ,  $i \in \{1, 2, \dots, 50\}$ , be the vectors which satisfy that  $e_1 = (1, 0, 0, \dots, 0, 0) \in \mathbb{R}^{50}$ ,  $e_2 = (0, 1, 0, \dots, 0, 0) \in \mathbb{R}^{50}$ ,  $\dots$ ,  $e_{50} = (0, 0, 0, \dots, 0, 1) \in \mathbb{R}^{50}$ , assume for every  $s, t \in [0, T]$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $w = (w_1, w_2, \dots, w_d) \in \mathbb{R}^d$  that  $N = 100$ ,  $d = 2\delta = 50$ ,  $\nu = (d + 50)d + (d + 50)^2 + (d + 50) = (d + 50)(2d + 51)$ ,  $\varphi(x) = \exp(-\alpha T) \max\{110 - [\sum_{i=1}^{\delta} \frac{x_{2i-1}}{\delta}], 0\}$ , and

$$\begin{aligned}
H(s, t, x, w) = & \sum_{i=1}^{\delta} \left( \left[ x_{2i-1} \exp\left(\left(\alpha - \frac{x_{2i}}{2}\right)(t-s) + w_{2i-1} \sqrt{x_{2i}}\right) \right] e_{2i-1} \right. \\
& + \left[ \max\left\{ \left[ \max\left\{ \frac{\beta}{2} \sqrt{t-s}, \max\left\{ \frac{\beta}{2} \sqrt{t-s}, \sqrt{x_{2i}} \right\} + \frac{\beta}{2} (\rho w_{2i-1} + [1 - \rho^2]^{1/2} w_{2i}) \right\} \right]^2 \right. \right. \\
& \left. \left. + \left( \kappa \theta - \frac{\beta^2}{4} - \kappa x_{2i} \right) (t-s), 0 \right\} \right] e_{2i} \right) \quad (179)
\end{aligned}$$

(cf. Hefter & Herzwurm [25, Section 1]), assume that  $\xi^{0,1}: \Omega \rightarrow \mathbb{R}^d$  is continuous uniformly distributed on  $\times_{i=1}^{\delta} ([90, 110] \times [0.02, 0.2])$ , and let  $u = (u(t, x))_{(t,x) \in [0,T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  be an at most polynomially growing function (cf., for example, Alfonsi [1, Proposition 4.1]) which satisfies for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  that  $u(0, x) = \varphi(x)$  and

$$\begin{aligned}
\left(\frac{\partial u}{\partial t}\right)(t, x) = & \left[ \sum_{i=1}^{\delta} \left( \alpha x_{2i-1} \left(\frac{\partial u}{\partial x_{2i-1}}\right)(t, x) + \kappa(\theta - x_{2i}) \left(\frac{\partial u}{\partial x_{2i}}\right)(t, x) \right) \right] \\
& + \left[ \sum_{i=1}^{\delta} \frac{|x_{2i}|}{2} \left( |x_{2i-1}|^2 \left(\frac{\partial^2 u}{\partial x_{2i-1}^2}\right)(t, x) + 2x_{2i-1} \beta \varrho \left(\frac{\partial^2 u}{\partial x_{2i-1} \partial x_{2i}}\right)(t, x) + \beta^2 \left(\frac{\partial^2 u}{\partial x_{2i}^2}\right)(t, x) \right) \right]. \quad (180)
\end{aligned}$$

Table 6 approximately presents the relative  $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (181) below), the relative  $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (182) below), and the relative  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error associated to  $(\mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x))_{x \in [0,1]^d}$  (see (183) below) against  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  (cf. PYTHON code 6 in Subsection 4.6 below). In our numerical simulations for Table 6 we approximately calculated the exact solution of the PDE (180) by means of Monte Carlo approximations with 1048576 samples and temporal SDE-discretizations based on (179) and 100 equidistant time steps, we approximately calculated the relative  $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right| dx \quad (181)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo

approximations with 10240 samples, we approximately calculated the relative  $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\sqrt{\int_{[0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right|^2 dx} \quad (182)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo approximations with 10240 samples, and we approximately calculated the relative  $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -approximation error

$$\sup_{x \in [0,1]^d} \left| \frac{u(T, x) - \mathbb{U}^{\Theta_m, 1, \mathbb{S}_m}(x)}{u(T, x)} \right| \quad (183)$$

for  $m \in \{0, 25000, 50000, 100000, 150000, 250000, 500000, 750000\}$  by means of Monte Carlo approximations with 10240 samples (see Lemma 3.5 above).

Number of steps	Relative $L^1(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Relative $L^2(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Relative $L^\infty(\lambda_{[0,1]^d}; \mathbb{R})$ -error	Runtime in seconds
0	1.038045	1.038686	1.210235	1.0
25000	0.005691	0.007215	0.053298	688.4
50000	0.005115	0.006553	0.036513	1375.2
100000	0.004749	0.005954	0.032411	2746.8
150000	0.006465	0.008581	0.051907	4120.2
250000	0.005075	0.006378	0.024458	6867.5
500000	0.002082	0.002704	0.019604	13763.7
750000	0.00174	0.002233	0.012466	20758.8

Table 6: Approximative presentations of the relative approximation errors in (181)–(183) for the Heston model in (180).

## 4 PYTHON source codes

### 4.1 PYTHON source code for the algorithm

In Subsections 4.2–4.6 below we present PYTHON source codes associated to the numerical simulations in Subsections 3.2–3.6 above. The following PYTHON source code, PYTHON code 1 below, is employed in the case of each of the PYTHON source codes in Subsections 4.2–4.6 below.

PYTHON code 1: *common.py*

```
1 import numpy as np
2 import tensorflow as tf
```

```

3 import time
4 from tensorflow.python.ops import init_ops
5 from tensorflow.contrib.layers.python.layers import initializers
6 from tensorflow.python.training.moving_averages import assign_moving_average
7
8
9 def neural_net(x, neurons, is_training, name,
10               mv_decay=0.9, dtype=tf.float32):
11
12     def _batch_normalization(_x):
13         beta = tf.get_variable('beta', [_x.get_shape()[-1]],
14                                dtype, init_ops.zeros_initializer())
15         gamma = tf.get_variable('gamma', [_x.get_shape()[-1]],
16                                 dtype, init_ops.ones_initializer())
17         mv_mean = tf.get_variable('mv_mean', [_x.get_shape()[-1]],
18                                   dtype, init_ops.zeros_initializer(),
19                                   trainable=False)
20         mv_variance = tf.get_variable('mv_variance', [_x.get_shape()[-1]],
21                                       dtype, init_ops.ones_initializer(),
22                                       trainable=False)
23         mean, variance = tf.nn.moments(_x, [0], name='moments')
24         tf.add_to_collection(tf.GraphKeys.UPDATE_OPS,
25                              assign_moving_average(mv_mean, mean,
26                                                    mv_decay, True))
27         tf.add_to_collection(tf.GraphKeys.UPDATE_OPS,
28                              assign_moving_average(mv_variance, variance,
29                                                    mv_decay, False))
30         mean, variance = tf.cond(is_training,
31                                  lambda: (mean, variance),
32                                  lambda: (mv_mean, mv_variance))
33         return tf.nn.batch_normalization(_x, mean, variance,
34                                         beta, gamma, 1e-6)
35
36     def _layer(_x, out_size, activation_fn):
37         w = tf.get_variable('weights',
38                             [_x.get_shape().as_list()[-1], out_size],
39                             dtype, initializers.xavier_initializer())
40         return activation_fn(_batch_normalization(tf.matmul(_x, w)))
41
42     with tf.variable_scope(name):
43         x = _batch_normalization(x)
44         for i in range(len(neurons)):
45             with tf.variable_scope('layer_%i_' % (i + 1)):
46                 x = _layer(x, neurons[i],
47                             tf.nn.tanh if i < len(neurons)-1 else tf.identity)
48     return x
49
50
51 def kolmogorov_train_and_test(xi, x_sde, phi, u_reference, neurons,

```

```

52         lr_boundaries, lr_values, train_steps,
53         mc_rounds, mc_freq, file_name,
54         dtype=tf.float32):
55
56     def _approximate_errors():
57         lr, gs = sess.run([learning_rate, global_step])
58         l1_err, l2_err, li_err = 0., 0., 0.
59         rel_l1_err, rel_l2_err, rel_li_err = 0., 0., 0.
60         for _ in range(mc_rounds):
61             l1, l2, li, r1l1, r1l2, r1li \
62                 = sess.run([err_l1_l1, err_l1_l2, err_l1_inf,
63                             rel_err_l1_l1, rel_err_l1_l2, rel_err_l1_inf],
64                             feed_dict={is_training: False})
65             l1_err, l2_err, li_err = (l1_err + l1, l2_err + l2,
66                                     np.maximum(li_err, li))
67             rel_l1_err, rel_l2_err, rel_li_err \
68                 = (rel_l1_err + r1l1, rel_l2_err + r1l2,
69                   np.maximum(rel_li_err, r1li))
70             l1_err, l2_err = l1_err / mc_rounds, np.sqrt(l2_err / mc_rounds)
71             rel_l1_err, rel_l2_err \
72                 = rel_l1_err / mc_rounds, np.sqrt(rel_l2_err / mc_rounds)
73             t_mc = time.time()
74             file_out.write('%i, %f, %f, %f, %f, %f, %f, %f, '
75                             '%f, %f\n' % (gs, l1_err, l2_err, li_err,
76                                             rel_l1_err, rel_l2_err, rel_li_err, lr,
77                                             t1_train - t0_train, t_mc - t1_train))
78             file_out.flush()
79
80         t0_train = time.time()
81         is_training = tf.placeholder(tf.bool, [])
82         u_approx = neural_net(xi, neurons, is_training, 'u_approx', dtype=dtype)
83         loss = tf.reduce_mean(tf.squared_difference(u_approx, phi(x_sde)))
84
85         err = tf.abs(u_approx - u_reference)
86         err_l1_l1 = tf.reduce_mean(err)
87         err_l1_l2 = tf.reduce_mean(err ** 2)
88         err_l1_inf = tf.reduce_max(err)
89         rel_err = err / tf.maximum(u_reference, 1e-8)
90         rel_err_l1_l1 = tf.reduce_mean(rel_err)
91         rel_err_l1_l2 = tf.reduce_mean(rel_err ** 2)
92         rel_err_l1_inf = tf.reduce_max(rel_err)
93
94         global_step = tf.get_variable('global_step', [], tf.int32,
95                                     tf.constant_initializer(0),
96                                     trainable=False)
97         learning_rate = tf.train.piecewise_constant(global_step,
98                                                     lr_boundaries,
99                                                     lr_values)
100        optimizer = tf.train.AdamOptimizer(learning_rate)

```

```

101 update_ops = tf.get_collection(tf.GraphKeys.UPDATE_OPS, 'u_approx')
102 with tf.control_dependencies(update_ops):
103     train_op = optimizer.minimize(loss, global_step)
104
105 file_out = open(file_name, 'w')
106 file_out.write('step, l1_err, l2_err, li_err, l1_rel, '
107               'l2_rel, li_rel, learning_rate, time_train, time_mc\n')
108
109 with tf.Session() as sess:
110
111     sess.run(tf.global_variables_initializer())
112
113     for step in range(train_steps):
114         if step % mc_freq == 0:
115             t1_train = time.time()
116             _approximate_errors()
117             t0_train = time.time()
118             sess.run(train_op, feed_dict={is_training: True})
119             t1_train = time.time()
120             _approximate_errors()
121
122 file_out.close()

```

## 4.2 A PYTHON source code associated to the numerical simulations in Subsection 3.2

PYTHON code 2: *example3-2.py*

```

1 import numpy as np
2 import tensorflow as tf
3 from common import kolmogorov_train_and_test
4
5 tf.reset_default_graph()
6 dtype = tf.float32
7 T, N, d = 1., 1, 100
8 batch_size = 8192
9 neurons = [d + 100, d + 100, 1]
10 train_steps = 750000
11 mc_rounds, mc_freq = 1250, 100
12 lr_boundaries = [250001, 500001]
13 lr_values = [0.001, 0.0001, 0.00001]
14 xi = tf.random_uniform(shape=(batch_size, d), minval=0.,
15                          maxval=1., dtype=dtype)
16 x_sde = xi + tf.random_normal(shape=(batch_size, d),
17                                stddev=np.sqrt(2. * T / N), dtype=dtype)
18

```

```

19
20 def phi(x):
21     return tf.reduce_sum(x ** 2, axis=1, keepdims=True)
22
23
24 u_reference = phi(xi) + 2. * T * d
25
26 kolmogorov_train_and_test(xi, x_sde, phi, u_reference, neurons,
27                           lr_boundaries, lr_values, train_steps,
28                           mc_rounds, mc_freq, 'example3_1.csv', dtype)

```

### 4.3 A PYTHON source code associated to the numerical simulations in Subsection 3.3

PYTHON code 3: *example3\_3.py*

```

1 import numpy as np
2 import tensorflow as tf
3 from common import kolmogorov_train_and_test
4
5 tf.reset_default_graph()
6 dtype = tf.float32
7 T, N, d = 1., 1, 100
8 r, c, K = 0.05, 0.1, 100.
9 sigma = tf.constant(0.1 + 0.5 * np.linspace(start=1. / d, stop=1., num=d,
10                                           endpoint=True), dtype=dtype)
11 batch_size = 8192
12 neurons = [d + 100, d + 100, 1]
13 train_steps = 750000
14 mc_rounds, mc_freq = 10, 25000
15 mc_samples_ref, mc_rounds_ref = 1024, 1024
16 lr_boundaries = [250001, 500001]
17 lr_values = [0.001, 0.0001, 0.00001]
18 xi = tf.random_uniform((batch_size, d), minval=90., maxval=110., dtype=dtype)
19
20
21 def phi(x, axis=1):
22     return np.exp(-r * T) \
23             * tf.maximum(tf.reduce_max(x, axis=axis, keepdims=True) - K, 0.)
24
25
26 def mc_body(idx, p):
27     _x = xi * tf.exp((r - c - 0.5 * sigma ** 2) * T + sigma
28                    * tf.random_normal((mc_samples_ref, batch_size, d),
29                                       stddev=np.sqrt(T / N), dtype=dtype))
30     return idx + 1, p + tf.reduce_mean(phi(_x, 2), axis=0)

```

```

31
32
33 x_sde = xi * tf.exp((r - c - 0.5 * sigma ** 2) * T
34                + sigma * tf.random_normal((batch_size, d),
35                stddev=np.sqrt(T / N),
36                dtype=dtype))
37 _, u = tf.while_loop(lambda idx, p: idx < mc_rounds_ref, mc_body,
38                (tf.constant(0), tf.zeros((batch_size, 1), dtype)))
39 u_reference = u / tf.cast(mc_rounds_ref, tf.float32)
40
41 kolmogorov_train_and_test(xi, x_sde, phi, u_reference, neurons,
42                lr_boundaries, lr_values, train_steps,
43                mc_rounds, mc_freq, 'example3_2.csv', dtype)

```

#### 4.4 A PYTHON source code associated to the numerical simulations in Subsection 3.4

PYTHON code 4: *example3\_4.py*

```

1 import numpy as np
2 import tensorflow as tf
3 from common import kolmogorov_train_and_test
4
5 tf.reset_default_graph()
6 dtype = tf.float32
7 T, N, d = 1., 1, 100
8 r, c, K = 0.05, 0.1, 110.
9 Q = np.ones([d, d]) * 0.5
10 np.fill_diagonal(Q, 1.)
11 L = np.linalg.cholesky(Q).transpose()
12 sigma_norms = tf.constant(np.linalg.norm(L, axis=0), dtype=dtype)
13 sigma = tf.constant(L, dtype=dtype)
14 beta = tf.constant(0.1 + 0.5 * np.linspace(start=1. / d, stop=1., num=d,
15                endpoint=True), dtype=dtype)
16 batch_size = 8192
17 neurons = [d + 100, d + 100, 1]
18 train_steps = 750000
19 mc_rounds, mc_freq = 10, 25000
20 mc_samples_ref, mc_rounds_ref = 1024, 1024
21 lr_boundaries = [250001, 500001]
22 lr_values = [0.001, 0.0001, 0.00001]
23 xi = tf.random_uniform((batch_size, d), minval=90., maxval=110., dtype=dtype)
24
25
26 def phi(x, axis=1):
27     return np.exp(-r * T) \

```



```

28         * tf.maximum(K - tf.reduce_min(x, axis=axis, keepdims=True), 0.)
29
30
31 def mc_body(idx, p):
32     _w = tf.matmul(tf.random_normal((mc_samples_ref * batch_size, d),
33                                   stddev=np.sqrt(T / N), dtype=dtype),
34                  sigma)
35     _w = tf.reshape(_w, (mc_samples_ref, batch_size, d))
36     _x = xi * tf.exp((r - c - 0.5 * (beta * sigma_norms) ** 2) * T
37                    + beta * _w)
38     return idx + 1, p + tf.reduce_mean(phi(_x, 2), axis=0)
39
40
41 x_sde = xi * tf.exp((r - c - 0.5 * (beta * sigma_norms) ** 2) * T + beta
42                   * tf.matmul(tf.random_normal((batch_size, d),
43                                                 stddev=np.sqrt(T / N),
44                                                 dtype=dtype),
45                              sigma))
46 _, u = tf.while_loop(lambda idx, p: idx < mc_rounds_ref, mc_body,
47                     (tf.constant(0), tf.zeros((batch_size, 1), dtype)))
48 u_reference = u / tf.cast(mc_rounds_ref, tf.float32)
49
50 kolmogorov_train_and_test(xi, x_sde, phi, u_reference, neurons,
51                          lr_boundaries, lr_values, train_steps,
52                          mc_rounds, mc_freq, 'example3_3.csv', dtype)

```

## 4.5 A PYTHON source code associated to the numerical simulations in Subsection 3.5

PYTHON code 5: *example3-5.py*

```

1 import numpy as np
2 import tensorflow as tf
3 from common import kolmogorov_train_and_test
4
5
6 tf.reset_default_graph()
7 dtype = tf.float32
8 batch_size = 1024
9
10 T, N, d = 1., 100, 3
11 alpha_1, alpha_2, alpha_3 = 10., 14., 8./3.
12 beta = tf.constant([0.15, 0.15, 0.15], dtype=dtype)
13 h = T / N
14 neurons = [d + 20, d + 20, 1]
15 train_steps = 750000

```

```

16 mc_rounds, mc_freq = 20, 25000
17 mc_samples_ref, mc_rounds_ref = 1024, 1024
18 lr_boundaries = [250001, 500001]
19 lr_values = [0.001, 0.0001, 0.00001]
20 xi = tf.stack([tf.random_uniform((batch_size,), minval=0.5,
21                                 maxval=2.5, dtype=dtype),
22               tf.random_uniform((batch_size,), minval=8.,
23                                 maxval=10., dtype=dtype),
24               tf.random_uniform((batch_size,), minval=10.,
25                                 maxval=12., dtype=dtype)], axis=1)
26
27
28 def phi(x, axis=1):
29     return tf.reduce_sum(x ** 2, axis=axis, keepdims=True)
30
31
32 def mu(x):
33     x_1 = tf.expand_dims(x[:, :, 0], axis=2)
34     x_2 = tf.expand_dims(x[:, :, 1], axis=2)
35     x_3 = tf.expand_dims(x[:, :, 2], axis=2)
36     return tf.concat([alpha_1 * (x_2 - x_1),
37                      alpha_2 * x_1 - x_2 - x_1 * x_3,
38                      x_1 * x_2 - alpha_3 * x_3], axis=2)
39
40
41 def sde_body(idx, s, samples):
42     return tf.add(idx, 1), s \
43           + tf.cast(T / N * tf.sqrt(phi(mu(s), 2 if samples > 1 else 1))
44                 <= 1., dtype) * mu(s) * T / N \
45           + beta * tf.random_normal((samples, batch_size, d),
46                                     stddev=np.sqrt(T / N), dtype=dtype)
47
48
49 def mc_body(idx, p):
50     _, _x = tf.while_loop(lambda _idx, s: _idx < N,
51                           lambda _idx, s: sde_body(_idx, s,
52                                                       mc_samples_ref),
53                           loop_var_mc)
54     return idx + 1, p + tf.reduce_mean(phi(_x, 2), axis=0)
55
56
57 loop_var_mc = (tf.constant(0),
58               tf.ones((mc_samples_ref, batch_size, d), dtype) * xi)
59 loop_var = (tf.constant(0), tf.ones((1, batch_size, d), dtype) * xi)
60 _, x_sde = tf.while_loop(lambda idx, s: idx < N,
61                           lambda idx, s: sde_body(idx, s, 1),
62                           loop_var)
63 _, u = tf.while_loop(lambda idx, p: idx < mc_rounds_ref,
64                       mc_body,

```

```

65         (tf.constant(0), tf.zeros((batch_size, 1), dtype)))
66 u_reference = u / tf.cast(mc_rounds_ref, tf.float32)
67
68 kolmogorov_train_and_test(xi, tf.squeeze(x_sde, axis=0), phi, u_reference,
69                          neurons, lr_boundaries, lr_values, train_steps,
70                          mc_rounds, mc_freq, 'example3_4.csv', dtype)

```

## 4.6 A PYTHON source code associated to the numerical simulations in Subsection 3.6

PYTHON code 6: *example3\_6.py*

```

1  import numpy as np
2  import tensorflow as tf
3  from common import kolmogorov_train_and_test
4
5  tf.reset_default_graph()
6  dtype = tf.float32
7  batch_size = 1024
8
9  T, N, d = 1., 100, 50
10 alpha, K = 0.05, 110.
11 kappa, sigma = 0.6, 0.2
12 theta, rho = 0.04, -0.8
13 S_0 = tf.random_uniform((batch_size, d / 2),
14                        minval=90., maxval=110., dtype=dtype)
15 V_0 = tf.random_uniform((batch_size, d / 2),
16                        minval=0.02, maxval=0.2, dtype=dtype)
17 h = T / N
18 neurons = [d + 50, d + 50, 1]
19 train_steps = 750000
20 mc_rounds, mc_freq = 10, 25000
21 mc_samples_ref, mc_rounds_ref = 256, 4096
22 lr_boundaries = [250001, 500001]
23 lr_values = [0.001, 0.0001, 0.00001]
24 xi = tf.reshape(tf.stack([S_0, V_0], axis=2),
25                (batch_size, d))
26
27
28 def phi(x, axis=1):
29     return np.exp(-alpha * T) \
30         * tf.maximum(K - tf.reduce_mean(tf.exp(x), axis=axis,
31                                       keepdims=True), 0.)
32
33
34 def sde_body(idx, s, v, samples):

```

```

35 _sqrt_v = tf.sqrt(v)
36 dw_1 = tf.random_normal(shape=(samples, batch_size, d / 2),
37                          stddev=np.sqrt(h), dtype=dtype)
38 dw_2 = rho * dw_1 + np.sqrt(1. - rho ** 2) \
39       * tf.random_normal(shape=(samples, batch_size, d / 2),
40                          stddev=np.sqrt(h), dtype=dtype)
41 return tf.add(idx, 1), s + (alpha - v / 2.) * h + _sqrt_v * dw_1, \
42       tf.maximum(tf.maximum(np.float32(sigma / 2. * np.sqrt(h)),
43                          tf.maximum(np.float32(sigma / 2. * np.sqrt(h)),
44                          _sqrt_v)
45                          + sigma / 2. * dw_2) ** 2
46       + (kappa * theta - sigma ** 2 / 4. - kappa * v) * h, 0.)
47
48
49 def mc_body(idx, p):
50     _, _x, _v = tf.while_loop(lambda _idx, s, v: _idx < N,
51                             lambda _idx, s, v: sde_body(_idx, s, v,
52                                                         mc_samples_ref),
53                             loop_var_mc)
54     return idx + 1, p + tf.reduce_mean(phi(_x, 2), axis=0)
55
56
57 loop_var_mc = (tf.constant(0), tf.ones((mc_samples_ref, batch_size, d / 2),
58                                       dtype) * tf.log(S_0),
59              tf.ones((mc_samples_ref, batch_size, d / 2),
60                      dtype) * V_0)
61 loop_var = (tf.constant(0), tf.ones((1, batch_size, d / 2),
62                                    dtype) * tf.log(S_0),
63            tf.ones((1, batch_size, d / 2), dtype) * V_0)
64 _, x_sde, _v = tf.while_loop(lambda idx, s, v: idx < N,
65                             lambda idx, s, v: sde_body(idx, s, v, 1),
66                             loop_var)
67 _, u = tf.while_loop(lambda idx, p: idx < mc_rounds_ref, mc_body,
68                    (tf.constant(0), tf.zeros((batch_size, 1), dtype)))
69 u_reference = u / tf.cast(mc_rounds_ref, tf.float32)
70
71 kolmogorov_train_and_test(xi, tf.squeeze(x_sde, axis=0), phi, u_reference,
72                          neurons, lr_boundaries, lr_values, train_steps,
73                          mc_rounds, mc_freq, 'example3_5.csv', dtype)

```

## References

- [1] ALFONSI, A. On the discretization schemes for the CIR (and Bessel squared) processes. *Monte Carlo Methods and Applications mcma* 11, 4 (2005), 355–384.

- [2] ANDERSSON, ADAM AND JENTZEN, ARNULF AND KURNIAWAN, RYAN. Existence, uniqueness, and regularity for stochastic evolution equations with irregular initial values. *arXiv:1512.06899* (2015), 35 pages. Revision requested from *J. Math. Anal. Appl.*
- [3] 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. *arXiv:1709.05963* (2017), 56 pages. Revision requested from *J. Nonlinear Sci.*
- [4] BECKER, S., CHERIDITO, P., AND JENTZEN, A. Deep optimal stopping. *arXiv:1804.05394* (2018), 18 pages.
- [5] BELLMAN, R. E. *Dynamic Programming*. Princeton University Press, 1957.
- [6] BISHOP, C. M. *Pattern recognition and machine learning*. Information Science and Statistics. Springer, New York, 2006.
- [7] BRENNAN, M. J., AND SCHWARTZ, E. S. The valuation of american put options. *The Journal of Finance* 32, 2 (1977), 449–462.
- [8] BRENNAN, M. J., AND SCHWARTZ, E. S. Finite difference methods and jump processes arising in the pricing of contingent claims: A synthesis. *The Journal of Financial and Quantitative Analysis* 13, 3 (1978), 461–474.
- [9] BRENNER, S., AND SCOTT, R. *The mathematical theory of finite element methods*, vol. 15. Springer Science & Business Media, 2007.
- [10] CIARLET, P. G. Basic error estimates for elliptic problems.
- [11] COX, S., HUTZENTHALER, M., AND JENTZEN, A. Local Lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations. *arXiv:1309.5595* (2013), 84 pages.
- [12] COX, SONJA AND JENTZEN, ARNULF AND KURNIAWAN, RYAN AND PUŠNIK, PRIMOŽ. On the mild Itô formula in Banach spaces. *arXiv:1612.03210* (2016), 27 pages. Accepted in *Discrete Contin. Dyn. Syst. Ser. B*.
- [13] DA PRATO, G., AND ZABCZYK, J. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2008.
- [14] 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* (2017), 349–380.

- [15] E, W., AND YU, B. The Deep Ritz method: A deep learning-based numerical algorithm for solving variational problems. *arXiv:1710.00211* (2017), 14 pages.
- [16] FUJII, M., TAKAHASHI, A., AND TAKAHASHI, M. Asymptotic Expansion as Prior Knowledge in Deep Learning Method for high dimensional BSDEs. *arXiv:1710.07030* (2017), 16 pages.
- [17] GILES, M. B. Multilevel Monte Carlo path simulation. *Oper. Res.* 56, 3 (2008), 607–617.
- [18] GLOROT, X., AND BENGIO, Y. Understanding the difficulty of training deep feed-forward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (2010), pp. 249–256.
- [19] GOLUB, G. H., AND VAN LOAN, C. F. *Matrix computations*, fourth ed. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 2013.
- [20] GRAHAM, C., AND TALAY, D. *Stochastic simulation and Monte Carlo methods*, vol. 68 of *Stochastic Modelling and Applied Probability*. Springer, Heidelberg, 2013. Mathematical foundations of stochastic simulation.
- [21] GYÖNGY, I. A note on Euler’s approximations. *Potential Anal.* 8, 3 (1998), 205–216.
- [22] HAIRER, M., HUTZENTHALER, M., AND JENTZEN, A. Loss of regularity for Kolmogorov equations. *Ann. Probab.* 43, 2 (2015), 468–527.
- [23] HAN, H., AND WU, X. A fast numerical method for the black–scholes equation of american options. *SIAM Journal on Numerical Analysis* 41, 6 (2003), 2081–2095.
- [24] HAN, J., JENTZEN, A., AND E, W. Overcoming the curse of dimensionality: Solving high-dimensional partial differential equations using deep learning. *arXiv:1707.02568* (2017), 13 pages.
- [25] HEFTER, M., AND HERZWURM, A. Strong convergence rates for Cox-Ingersoll-Ross processes—full parameter range. *J. Math. Anal. Appl.* 459, 2 (2018), 1079–1101.
- [26] HENRY-LABORDERE, P. Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM. 16 pages. Available at SSRN: <https://ssrn.com/abstract=3071506>.
- [27] HIGHAM., D. J. An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations. *SIAM Review* 43, 3 (2001), 525–546.

- [28] HIGHAM, D. J. Stochastic ordinary differential equations in applied and computational mathematics. *IMA journal of applied mathematics* 76, 3 (2011), 449–474.
- [29] HIGHAM, D. J., MAO, X., AND STUART, A. M. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.* 40, 3 (2002), 1041–1063.
- [30] HOFMANN, N., MÜLLER-GRONBACH, T., AND RITTER, K. Optimal approximation of stochastic differential equations by adaptive step-size control. *Math. Comp.* 69, 231 (2000), 1017–1034.
- [31] HÖRMANDER, L. Hypoelliptic second order differential equations. *Acta Math.* 119 (1967), 147–171.
- [32] HUTZENTHALER, M., AND JENTZEN, A. Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients. *Mem. Amer. Math. Soc.* 236, 1112 (2015), v+99.
- [33] HUTZENTHALER, M., JENTZEN, A., AND KLOEDEN, P. E. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *Ann. Appl. Probab.* 22, 4 (2012), 1611–1641.
- [34] HUTZENTHALER, M., JENTZEN, A., AND WANG, X. Exponential integrability properties of numerical approximation processes for nonlinear stochastic differential equations. *Math. Comp.* 87, 311 (2018), 1353–1413.
- [35] HUTZENTHALER, MARTIN AND JENTZEN, ARNULF AND SALIMOVA, DIYORA. Strong convergence of full-discrete nonlinearity-truncated accelerated exponential Euler-type approximations for stochastic Kuramoto-Sivashinsky equations. *arXiv:1604.02053* (2016), 40 pages. Accepted in *Comm. Math. Sci.*
- [36] IOFFE, S., AND SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv:1502.03167* (2015), 11 pages.
- [37] JENTZEN, A., KUCKUCK, B., NEUFELD, A., AND VON WURSTEMBERGER, P. Strong  $L^p$ -error analysis for stochastic gradient descent optimization algorithms. (2018), 51 pages.
- [38] KINGMA, D., AND BA, J. *Adam: a method for stochastic optimization*. Proceedings of the International Conference on Learning Representations (ICLR), 2015.
- [39] KLENKE, A. *Probability theory*, second ed. Universitext. Springer, London, 2014. A comprehensive course.

- [40] KLOEDEN, P. E. The systematic derivation of higher order numerical schemes for stochastic differential equations. *Milan J. Math.* 70 (2002), 187–207.
- [41] KLOEDEN, P. E., AND PLATEN, E. *Numerical solution of stochastic differential equations*, vol. 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.
- [42] KLOEDEN, P. E., PLATEN, E., AND SCHURZ, H. *Numerical solution of SDE through computer experiments*. Springer Science & Business Media, 2012.
- [43] KUSHNER, H. Finite difference methods for the weak solutions of the kolmogorov equations for the density of both diffusion and conditional diffusion processes. *Journal of Mathematical Analysis and Applications* 53, 2 (1976), 251 – 265.
- [44] MARUYAMA, G. Continuous Markov processes and stochastic equations. *Rend. Circ. Mat. Palermo (2)* 4 (1955), 48–90.
- [45] MILSTEIN, G. N. Approximate integration of stochastic differential equations. *Teor. Veroyatnost. i Primenen.* 19 (1974), 583–588.
- [46] MILSTEIN, G. N. *Numerical integration of stochastic differential equations*, vol. 313 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1995. Translated and revised from the 1988 Russian original.
- [47] MILSTEIN, G. N., AND TRETYAKOV, M. V. *Stochastic numerics for mathematical physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.
- [48] MILSTEIN, G. N., AND TRETYAKOV, M. V. Numerical Integration of Stochastic Differential Equations with Nonglobally Lipschitz Coefficients. *SIAM Journal on Numerical Analysis* 43, 3 (2005), 1139–1154.
- [49] MÜLLER-GRONBACH, T., AND RITTER, K. Minimal errors for strong and weak approximation of stochastic differential equations. In *Monte Carlo and quasi-Monte Carlo methods 2006*. Springer, Berlin, 2008, pp. 53–82.
- [50] ØKSENDAL, B. *Stochastic differential equations*, sixth ed. Universitext. Springer-Verlag, Berlin, 2003. An introduction with applications.
- [51] RAISSI, M. Forward-backward stochastic neural networks: Deep learning of high-dimensional partial differential equations. *arXiv:1804.07010* (2018).
- [52] RÖSSLER, A. *Runge-Kutta Methods for the Strong Approximation of Solutions of Stochastic Differential Equations*. Shaker, Aachen, 2009.



- [53] RUDER, S. An overview of gradient descent optimization algorithms. *arXiv:1609.04747* (2016), 14 pages.
- [54] SABANIS, S. A note on tamed Euler approximations. *Electron. Commun. Probab.* 18 (2013), no. 47, 10.
- [55] SABANIS, S. Euler approximations with varying coefficients: the case of superlinearly growing diffusion coefficients. *Ann. Appl. Probab.* 26, 4 (2016), 2083–2105.
- [56] SCHWARTZ, E. S. The valuation of warrants: Implementing a new approach. *Journal of Financial Economics* 4, 1 (1977), 79 – 93.
- [57] SIRIGNANO, J., AND SPILIOPOULOS, K. DGM: A deep learning algorithm for solving partial differential equations. *arXiv:1708.07469* (2017), 16 pages.
- [58] ZHAO, J., DAVISON, M., AND CORLESS, R. M. Compact finite difference method for american option pricing. *Journal of Computational and Applied Mathematics* 206, 1 (2007), 306 – 321.
- [59] ZIENKIEWICZ, O. C., TAYLOR, R. L., ZIENKIEWICZ, O. C., AND TAYLOR, R. L. *The finite element method*, vol. 3. McGraw-hill London, 1977.