# MULTILEVEL PICARD APPROXIMATIONS AND DEEP NEURAL NETWORKS WITH RELU, LEAKY RELU, AND SOFTPLUS ACTIVATION OVERCOME THE CURSE OF DIMENSIONALITY WHEN APPROXIMATING SEMILINEAR PARABOLIC PARTIAL DIFFERENTIAL EQUATIONS IN $L^p$-SENSE

ARIEL NEUFELD[1] AND TUAN ANH NGUYEN[2]

ABSTRACT. We prove that multilevel Picard approximations and deep neural networks with ReLU, leaky ReLU, and softplus activation are capable of approximating solutions of semilinear Kolmogorov PDEs in $L^\mathfrak{p}$-sense, $\mathfrak{p} \in [2, \infty)$, in the case of gradient-independent, Lipschitz-continuous nonlinearities, while the computational effort of the multilevel Picard approximations and the required number of parameters in the neural networks grow at most polynomially in both dimension $d \in \mathbb{N}$ and reciprocal of the prescribed accuracy $\epsilon$.



## 1. INTRODUCTION

Partial differential equations (PDEs) are important tools to analyze many real world phenomena, e.g., in financial engineering, economics, quantum mechanics, or statistical physics to name but a few. In most of the cases such high-dimensional nonlinear PDEs cannot be solved explicitly. It is one of the most challenging problems in applied mathematics to approximately solve high-dimensional nonlinear PDEs. In particular, it is very difficult to find approximation schemata for nonlinear PDEs for which one can rigorously prove that they do overcome the so-called *curse of dimensionality* in the sense that the computational complexity only grows polynomially in the space dimension $d$ of the PDE and the reciprocal $\frac{1}{\varepsilon}$ of the accuracy $\varepsilon$.

In recent years, there are two types of approximation methods which are quite successful in the numerical approximation of solutions of high-dimensional nonlinear PDEs: neural network based approximation methods for PDEs, cf., [3, 4, 5, 6, 11, 13, 15, 17, 19, 20, 21, 23, 24, 25, 27, 28, 29, 30, 31, 32, 32, 33, 34, 35, 42, 48, 49, 50, 52, 53, 59, 62, 63, 64, 65, 66, 68, 69] and multilevel Monte-Carlo based approximation methods for PDEs, cf., [8, 9, 12, 26, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 57, 60, 61].

For multilevel Monte-Carlo based algorithms it is often possible to provide a complete convergence and complexity analysis. It has been proven that under some suitable assumptions, e.g., Lipschitz continuity on the linear part, the nonlinear part, and the initial (or terminal) condition function of the PDE under consideration, the multilevel Picard approximation algorithms can overcome the curse of dimensionality in the sense that the number of computational operations of the proposed Monte-Carlo based approximation method grows at most polynomially in both the reciprocal $\frac{1}{\varepsilon}$ of the prescribed approximation accuracy $\varepsilon \in (0, 1)$ and the PDE dimension $d \in \mathbb{N}$. More precisely, [38] considers smooth semilinear parabolic heat equations. Later, [40] extends [38] to a more general setting, namely, semilinear heat equations which are not necessary smooth. [9] considers semilinear heat equation with more general nonlinearities, namely locally Lipschitz nonlinearities. [39, 45] considers semilinear heat equations with gradient-dependent Lipschitz nonlinearities and [57, 61] extends them to semilinear PDEs with general drift and diffusion coefficients. [44] studies Black-Scholes-types semilinear PDEs. [41] consider semilinear parabolic PDEs with nonconstant drift and diffusion coefficients. [47] considers a slightly more general setting than [41], namely semilinear PDEs with locally monotone coefficient functions. [60] studies semilinear partial integro-differential equations. [46]

2010 *Mathematics Subject Classification.* 65C99, 68T05.

Financial support by the Nanyang Assistant Professorship Grant (NAP Grant) *Machine Learning based Algorithms in Finance and Insurance* is gratefully acknowledged.

considers McKean-Vlasov stochastic differential equations (SDEs) with constant diffusion coefficients. [8] studies a special type of elliptic equations. Almost all the works listed above prove $L^2$-error estimates except [43], which draws its attention to $L^{\mathfrak{p}}$-error estimates, $\mathfrak{p} \in [2, \infty)$.

Numerical experiments indicate that deep learning methods work exceptionally well when approximating solutions of high-dimensional PDEs and that they do not suffer from the curse of dimensionality. However, there exist only few theoretical results proving that deep learning based approximations of solutions of PDEs do not suffer from the curse of dimensionality. More precisely, [14] shows that empirical risk minimization over deep neural network (DNN) hypothesis classes overcomes the curse of dimensionality for the numerical solution of linear Kolmogorov equations with affine coefficients. Next, [22] considers the pricing problem of a European best-of-call option on a basket of $d$ assets within the Black–Scholes model and proves that the solution to the $d$-variate option pricing problem can be approximated up to an $\epsilon$-error by a deep ReLU network with depth $\mathcal{O}(\ln(d)\ln(\epsilon^{-1}) + (\ln(d))^2)$ and $\mathcal{O}(d^{2+\frac{1}{n}}\epsilon^{-\frac{1}{n}})$ nonzero weights, where $n \in \mathbb{N}$ is arbitrary (with the constant implied in $\mathcal{O}(\cdot)$ depending on $n$). Furthermore, [28] investigates the use of random neural networks for learning Kolmogorov partial integro-differential equations (PIDEs) associated to Black-Scholes and more general exponential Lévy models. Here, random neural networks are single-hidden-layer feedforward neural networks in which the input weights are randomly generated and only the output weights are trained. In addition, [55] proves that rectified deep neural networks overcome the curse of dimensionality when approximating solutions of McKean–Vlasov stochastic differential equations. Moreover, [29] studies the expression rates of DNNs for option prices written on baskets of $d$ risky assets whose log-returns are modelled by a multivariate Lévy process with general correlation structure of jumps. Note that the PIDEs studied by [29] are also Black-Scholes-type PIDEs (see [29, Display (2.3)]). Next, [30] proves that DNNs with ReLU activation function are able to express viscosity solutions of Kolmogorov linear PIDEs on state spaces of possibly high dimension $d$. Furthermore, [31] proves that DNNs overcome the curse of dimensionality when approximating the solutions to Black-Scholes PDEs and [52] proves that DNNs overcome the curse of dimensionality in the numerical approximation of linear Kolmogorov PDEs with constant diffusion and nonlinear drift coefficients. In addition, [58] proves that the solution of the linear heat equation can be approximated by a random neural network whose amount of neurons only grow polynomially in the space dimension of the PDE and the reciprocal of the accuracy, hence overcoming the curse of dimensionality when approximating such an equation. Moreover, [42] proves that DNNs overcome the curse of dimensionality in the numerical approximation of semilinear heat equations and [1] extends [42] to estimates with respect to $L^{\mathfrak{p}}$-norms, $\mathfrak{p} \in [2, \infty)$, when approximating the semilinear heat equation. Furthermore, [2] demonstrates space-time $L^{\mathfrak{p}}$-error estimates, $\mathfrak{p} \in [2, \infty)$, when approximating the semilinear heat equation. Next, [17] extends [42] to semilinear PDEs with general drift and diffusion coefficients and [56] extends [42] to semilinear PIDEs. Note that except [1, 2] all the works mentioned in this paragraph establish $L^2$-error estimates, but not $L^{\mathfrak{p}}$-estimates for general $\mathfrak{p} \in [2, \infty)$.

The main novelty of our paper is the following:

(A) We extend the $L^2$-complexity analysis in [41] to an $L^{\mathfrak{p}}$-complexity analysis, $\mathfrak{p} \in [2, \infty)$. More precisely, in our first main result, Theorem 1.1 below, we prove that the MLP algorithms introduced by [41] overcome the curse of dimensionality when approximating semilinear parabolic PDEs in $L^{\mathfrak{p}}$-sense, $\mathfrak{p} \in [2, \infty)$.

(B) We extend the result by [17] to an $L^{\mathfrak{p}}$-sense, $\mathfrak{p} \in [2, \infty)$, and to DNNs with ReLU, leaky ReLU, or softplus activation, see Theorem 1.3 below, which is our second main result. More precisely, we show that for every $\mathfrak{p} \in [2, \infty)$ we have that solutions of semilinear PDEs with Lipschitz continuous nonlinearities can be approximated in the $L^{\mathfrak{p}}$-sense by DNNs with ReLU, leaky ReLU, or softplus activation without the curse of dimensionality.

1.1. **Notations.** Throughout this paper we use the following notations. Let $\mathbb{R}$ denote the set of all real numbers. Let $\mathbb{Z}, \mathbb{N}_0, \mathbb{N}$ denote the sets which satisfy that $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$,

$\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Let $\nabla$ denote the gradient and $\mathrm{Hess}$ denote the Hessian matrix. For every matrix $A$ let $A^\top$ denote the transpose of $A$ and let $\mathrm{trace}(A)$ denote the trace of $A$ when $A$ is a square matrix. For every probability space $(\Omega, \mathcal{F}, \mathbb{P})$, every random variable $X : \Omega \to \mathbb{R}$, and every $s \in [1, \infty)$ let $\|X\|_s \in [0, \infty]$ satisfy that $\|X\|_s = (\mathbb{E}[|X|^s])^{\frac{1}{s}}$. For every $d \in \mathbb{N}$ let $\|\cdot\|, |||\cdot||| : \mathbb{R}^d \to [0, \infty)$ satisfy for all $x = (x_i)_{i \in [1,d] \cap \mathbb{Z}} \in \mathbb{R}^d$ that $\|x\| = \sqrt{\sum_{i=1}^d |x_i|^2}$ and $|||x||| = \sup_{i \in [1,d] \cap \mathbb{Z}} |x_i|$. For every $d \in \mathbb{N}$ let $\langle \cdot, \cdot \rangle : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x = (x_i)_{i \in [1,d] \cap \mathbb{Z}}$, $y = (y_i)_{i \in [1,d] \cap \mathbb{Z}}$ that $\langle x, y \rangle = \sum_{i=1}^d x_i y_i$. For every $d \in \mathbb{N}$ let $\|\cdot\| : \mathbb{R}^{d \times d} \to [0, \infty)$ satisfy for all $a = (a_{ij})_{i,j \in [1,d] \cap \mathbb{Z}} \in \mathbb{R}^{d \times d}$ that $\|a\| = \sqrt{\sum_{i=1}^d \sum_{i=1}^d |a_{ij}|^2}$. When applying a result we often use a phrase like "Lemma 3.8 with $d \curvearrowleft (d-1)$" that should be read as "Lemma 3.8 applied with $d$ (in the notation of Lemma 3.8) replaced by $(d-1)$ (in the current notation)".

### 1.2. MLP approximations overcome the curse of dimensionality when approximating semilinear parabolic PDEs in $L^{\mathfrak{p}}$-sense.

**Theorem 1.1.** *Let $T, \mathbf{k} \in (0, \infty)$, $\mathfrak{p} \in [2, \infty)$, $c \in [\mathfrak{p}^2, \infty)$. Let $M : \mathbb{N} \to \mathbb{N}$ satisfy for all $n \in \mathbb{N}$ that $M_n = \max\{k \in \mathbb{N} : k \leqslant \exp(|\ln(n)|^{1/2})\}$. For every $d \in \mathbb{N}$ let $g^d \in C(\mathbb{R}^d, \mathbb{R})$, $f \in C(\mathbb{R}, \mathbb{R})$, $\mu^d \in C(\mathbb{R}^d, \mathbb{R}^d)$, $\sigma^d \in C(\mathbb{R}^d, \mathbb{R}^{d \times d})$. Assume for all $x, y \in \mathbb{R}^d$, $v, w \in \mathbb{R}^d$ that*

$$\max\{|Tf(0)|, |g^d(0)|, \|\mu^d(0)\|, \|\sigma^d(0)\|\} \leqslant cd^c, \quad |g(x)| \leqslant c(d^c + \|x\|^2)^{\frac{1}{2}}, \tag{1}$$

$$\max\{\sqrt{T}|g^d(x) - g^d(y)|, \|\mu^d(x) - \mu^d(y)\|, \|\sigma^d(x) - \sigma^d(y)\|\} \leqslant c\|x - y\|, \tag{2}$$

$$|f(w) - f(v)| \leqslant c|w - v|. \tag{3}$$

*Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t \in [0,T]})$ be a filtered probability space which satisfies the usual conditions. Let $\Theta = \bigcup_{n \in \mathbb{N}} \mathbb{Z}^n$. Let $\mathfrak{t}^\theta : \Omega \to [0, 1]$, $\theta \in \Theta$, be identically distributed and independent random variables. Assume for all $t \in (0, 1)$ that $\mathbb{P}(\mathfrak{t}^0 \leqslant t) = t$. For every $d \in \mathbb{N}$ let $W^{d,\theta} : [0, T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, be independent standard $(\mathbb{F}_t)_{t \in [0,T]}$-Brownian motions. Assume that $(\mathfrak{t}^\theta)_{\theta \in \Theta}$ and $(W^{d,\theta})_{d \in \mathbb{N}, \theta \in \Theta}$ are independent. For every $K \in \mathbb{N}$ let $\lfloor \cdot \rfloor_K : \mathbb{R} \to \mathbb{R}$ satisfy for all $t \in \mathbb{R}$ that $\lfloor t \rfloor_K = \max(\{0, \frac{T}{K}, \ldots, \frac{(K-1)T}{T}, T\} \cap ((-\infty, t) \cup \{0\}))$. For every $d, K \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$ let $Y^{d,\theta,K,t,x} = (Y_s^{d,\theta,K,t,x})_{s \in [t,T]} : [t, T] \times \Omega \to \mathbb{R}^d$ satisfy for all $s \in [t, T]$ that $Y_t^{d,\theta,K,t,x} = x$ and*

$$Y_s^{d,\theta,K,t,x} = Y_{\max\{t,\lfloor s \rfloor_K\}}^{d,\theta,K,t,x} + \mu^d(Y_{\max\{t,\lfloor s \rfloor_K\}}^{d,\theta,K,t,x})(s - \max\{t, \lfloor s \rfloor_K\}) + \sigma^d(Y_{\max\{t,\lfloor s \rfloor_K\}}^{d,\theta,K,t,x})(W_s^{d,\theta} - W_{\max\{t,\lfloor s \rfloor_K\}}^{d,\theta}). \tag{4}$$

*Let $U_{n,m}^{d,\theta,K} : [0, T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$, $n \in \mathbb{Z}$, $d, K, m \in \mathbb{N}$, $\theta \in \Theta$, satisfy for all $\theta \in \Theta$, $d, K, m \in \mathbb{N}$, $n \in \mathbb{N}_0$, $t \in [0, T]$, $x \in \mathbb{R}^d$ that $U_{-1,m}^{d,\theta,K}(t, x) = U_{0,m}^{d,\theta,K}(t, x) = 0$ and*

$$\begin{aligned} U_{n,m}^{d,\theta,K}(t, x) &= \frac{1}{m^n} \sum_{i=1}^{m^n} g^d(Y_T^{d,(\theta,0,-i),K,t,x}) \\ &+ \sum_{\ell=0}^{n-1} \frac{T - t}{m^{n-\ell}} \sum_{i=1}^{m^{n-\ell}} \left(f \circ U_{\ell,m}^{d,(\theta,\ell,i),K} - \mathbb{1}_{\mathbb{N}}(\ell) f \circ U_{\ell-1,m}^{d,(\theta,-\ell,i),K}\right)\left(t + (T - t)\mathfrak{t}^{(\theta,\ell,i)}, Y_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}}^{d,(\theta,\ell,i),K,t,x}\right). \end{aligned} \tag{5}$$

*Let $(C_{n,m}^{d,K})_{d,K \in \mathbb{N}, n,m \in \mathbb{Z}} \subseteq \mathbb{N}_0$ satisfy for all $d, K \in \mathbb{N}$, $m, n \in \mathbb{N}$ that*

$$C_{0,m}^{d,K} = 0, \quad C_{n,m}^{d,K} \leqslant (cd^c + cd^c K)m^n + \sum_{\ell=0}^{n-1} m^{n-\ell}\left(2cd^c + cd^c K + C_{\ell,m}^{d,K} + C_{\ell-1,m}^{d,K}\right). \tag{6}$$

*Then the following items are true.*

*(i) For every $d \in \mathbb{N}$ there exists a unique at most polynomially growing viscosity solution $u^d$ of*

$$\frac{\partial u^d}{\partial t}(t, x) + \frac{1}{2}\mathrm{trace}(\sigma^d(x)(\sigma^d(x))^\top(\mathrm{Hess}_x u^d(t, x))) + \langle \mu^d(x), (\nabla_x u^d)(t, x) \rangle + f(u^d(t, x)) = 0 \tag{7}$$

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

*(ii) There exist* $(C_\delta)_{\delta\in(0,1)} \subseteq (0,\infty)$, $(n(d,\epsilon))_{d\in\mathbb{N},\epsilon\in(0,1)} \subseteq \mathbb{N}$ *such that for all* $d \in \mathbb{N}$, $\epsilon \in (0,1)$ *it holds that*

$$\sup_{t\in[0,T],x\in[0,\mathbf{k}]^d} \left\| U^{d,0,(M_{n(d,\varepsilon)})^{n(d,\varepsilon)}}_{n(d,\varepsilon),M_{n(d,\varepsilon)}}(t,x) - u^d(t,x) \right\|_{\mathfrak{p}} \leqslant \epsilon \quad \text{and} \quad C^{d,(M_{n(d,\epsilon)})^{n(d,\epsilon)}}_{n(d,\epsilon),M_{n(d,\epsilon)}} \leqslant \eta d^\eta \epsilon^{-(4+\delta)}. \tag{8}$$

The proof of Theorem 1.1 is presented directly after the proof of Lemma 2.4. Let us comment on the mathematical objects in Theorem 1.1. Our goal here in Theorem 1.1 is to approximately solve the family of semilinear parabolic PDEs in (7) indexed by $d \in \mathbb{N}$. The functions $\mu^d$ and $\sigma^d$ are the drift and diffusion coefficients of the linear part of the PDEs. The function $f$ is the nonlinear part of the PDEs. The functions $g^d$ is the terminal condition at time $T$ of the PDEs. Next, (1)–(3) are usual regularity properties for the coefficients of the PDEs, which assure that the PDEs has unique viscosity solutions. The filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ in Theorem 1.1 above is the probability space on which we introduce the stochastic MLP approximations which we employ to approximate the solutions $u^d$ of the PDEs in (7). The set $\Theta$ in Theorem 1.1 is used as an index set to introduce sufficiently many independent random variables. The functions $\mathfrak{t}^\theta$ are independent random variables which are uniformly distributed on $[0,1]$. The functions $W^{d,\theta}$ describe independent standard Brownian motions which we use as random input sources for the MLP approximations. The functions $Y^{d,\theta,K,t,x}$ in (4) above describe Euler-Mayurama approximations which we use in the MLP approximations in (5) above as discretizations of the underlying Itô processes associated to the linear parts of the PDEs in (7). The function $U^{d,\theta,K}_{n,m}$ in (5) describe the MLP approximations which we employ to approximately compute the solutions $u^d$ to the PDEs (7). Let us discuss the computational effort of the MLP approximations in (5). We assume that the computational effort of $f$, $g^d$, $(\mu^d, \sigma^d)$ plus the effort to simulate an arbitrary $d$-dimensional Brownian increments is bounded by $cd^c$, which is a polynomial of $d$. Each $C^{d,K}_{n,m}$ in (6) is the computational effort to compute a realization of $U^{d,\theta,K}_{n,m}(t,x,\omega)$. Due to (49) and (5) the family $(C^{d,K}_{n,m})$ satisfies the recursive inequality (6) above. Theorem 1.1 establishes that the solutions $u^d$ of the PDEs in (7) can be approximated by the MLP approximations $U^{d,\theta,K}_{n,m}$ in (5) with the number of involved function evaluations and the number of involved scalar random variables growing at most polynomially in the reciprocal $1/\epsilon$ of the prescribed approximation accuracy $\epsilon \in (0,1)$ and at most polynomially in the PDE dimension $d \in \mathbb{N}$. In other words, Theorem 1.1 states that MLP approximations overcome the curse of dimensionality when approximating the semilinear parabolic PDEs in (7).

1.3. **A mathematical framework for DNNs.** In order to formulate our second main result, Theorem 1.3, we first need to introduce a mathematical frame work for DNNs.

**Setting 1.2** (A mathematical framework for DNNs)**.** *Let* $a \in C(\mathbb{R},\mathbb{R})$. *Let* $\mathbf{A}_d\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, *satisfy for all* $d \in \mathbb{N}$, $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$ *that*

$$\mathbf{A}_d(x) = (a(x_1), a(x_2), \ldots, a(x_d)). \tag{9}$$

*Let* $\mathbf{D} = \cup_{H\in\mathbb{N}} \mathbb{N}^{H+2}$. *Let*

$$\mathbf{N} = \bigcup_{H\in\mathbb{N}} \bigcup_{(k_0,k_1,\ldots,k_{H+1})\in\mathbb{N}^{H+2}} \left[ \prod_{n=1}^{H+1} \left( \mathbb{R}^{k_n\times k_{n-1}} \times \mathbb{R}^{k_n} \right) \right]. \tag{10}$$

*Let* $\mathcal{D}\colon \mathbf{N} \to \mathbf{D}$, $\mathcal{P}\colon \mathbf{N} \to \mathbb{N}$, $\mathcal{R}\colon \mathbf{N} \to (\cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k,\mathbb{R}^l))$ *satisfy that for all* $H \in \mathbb{N}$, $k_0, k_1, \ldots, k_H, k_{H+1} \in \mathbb{N}$, $\Phi = ((W_1,B_1),\ldots,(W_{H+1},B_{H+1})) \in \prod_{n=1}^{H+1} \left( \mathbb{R}^{k_n\times k_{n-1}} \times \mathbb{R}^{k_n} \right)$, $x_0 \in \mathbb{R}^{k_0}, \ldots, x_H \in \mathbb{R}^{k_H}$ *with the property that* $\forall\, n \in \mathbb{N}\cap[1,H]\colon x_n = \mathbf{A}_{k_n}(W_n x_{n-1} + B_n)$ *we have that*

$$\mathcal{P}(\Phi) = \sum_{n=1}^{H+1} k_n(k_{n-1}+1), \quad \mathcal{D}(\Phi) = (k_0, k_1, \ldots, k_H, k_{H+1}), \tag{11}$$

$\mathcal{R}(\Phi) \in C(\mathbb{R}^{k_0}, \mathbb{R}^{k_{H+1}})$, *and*

$$(\mathcal{R}(\Phi))(x_0) = W_{H+1}x_H + B_{H+1}. \tag{12}$$

Let us comment on the mathematical objects in Setting 1.2. The function $a$ is called the activation function. An example of $a$ is the ReLU function $\mathbb{R} \ni x \mapsto \max\{x, 0\} \in \mathbb{R}$. However, in this paper we do not restrict ourselves in this function. For all $d \in \mathbb{N}$, $\mathbf{A}_d \colon \mathbb{R}^d \to \mathbb{R}^d$ refers to the componentwise activation function. By $\mathbf{N}$ we denote the set of all parameters characterizing artificial feed-forward DNNs. For every $H \in \mathbb{N}$, $k_0, k_1, \ldots, k_H, k_{H+1} \in \mathbb{N}$, $\Phi = ((W_1, B_1), \ldots, (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} \left(\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n}\right) \subseteq \mathbf{N}$ the natural number $H$ can be interpreted as the depth of the parameters characterizing artificial feed-forward DNN $\Phi$ and $(W_1, B_1), \ldots, (W_{H+1}, B_{H+1})$ can be interpreted as the parameters of $\Phi$. By $\mathcal{R}$ we denote the operator that maps each parameters characterizing a DNN to its corresponding function. By $\mathcal{P}$ we denote the function that counts the number of parameters of the corresponding DNN. By $\mathcal{D}$ we denote the function that maps the parameters characterizing a DNN to the vector of its layer dimensions.

### 1.4. DNNs overcome the curse of dimensionality when approximating semilinear parabolic PDEs in $L^{\mathfrak{p}}$-sense.

**Theorem 1.3.** *Assume Setting 1.2. Let* $\alpha \in [0, \infty)\backslash\{1\}$, $\mathfrak{a}_0, \mathfrak{a}_1 \in C(\mathbb{R}, \mathbb{R})$ *satisfy for all* $x \in \mathbb{R}$ *that* $\mathfrak{a}_0 = \max\{x, \alpha x\}$ *and* $\mathfrak{a}_1 = \ln(1 + e^x)$. *Assume that* $a \in \{\mathfrak{a}_0, \mathfrak{a}_1\}$. *Let* $\beta, \mathfrak{p} \in [2, \infty)$, $c \in [1, \infty)$. *For every* $d \in \mathbb{N}$, $\varepsilon \in (0, 1)$, $v \in \mathbb{R}^d$ *let* $\Phi_{\mu^d_\varepsilon}, \Phi_{\sigma^d_\varepsilon, v}, \Phi_{g^d_\varepsilon} \in \mathbf{N}$, $f \in C(\mathbb{R}, \mathbb{R})$, $g^d, g^d_\varepsilon \in C(\mathbb{R}^d, \mathbb{R})$, $\mu^d, \mu^d_\varepsilon \in C(\mathbb{R}^d, \mathbb{R}^d)$, $\sigma^d, \sigma^d_\varepsilon \in C(\mathbb{R}^{d\times d}, \mathbb{R}^d)$ *satisfy for all* $v \in \mathbb{R}^d$ *that* $\mu^d_\varepsilon = \mathcal{R}(\Phi_{\mu^d_\varepsilon})$, $\sigma^d_\varepsilon(\cdot)v = \mathcal{R}(\Phi_{\sigma^d_\varepsilon, v})$, $g^d_\varepsilon = \mathcal{R}(\Phi_{g^d_\varepsilon})$. *Assume for all* $d \in \mathbb{N}$, $\varepsilon \in (0, 1)$, $v \in \mathbb{R}^d$ *that* $\mathcal{D}(\Phi_{\sigma^d_\varepsilon, v}) = \mathcal{D}(\Phi_{\sigma^d_\varepsilon, 0})$. *Assume for all* $d \in \mathbb{N}$, $\varepsilon \in (0, 1)$, $v, w \in \mathbb{R}$, $x, y \in \mathbb{R}^d$ *that*

$$\max\{\|\mu^d_\varepsilon(x) - \mu^d_\varepsilon(y)\|, \|\sigma^d_\varepsilon(x) - \sigma^d_\varepsilon(y)\|\} \leqslant c\|x - y\|, \tag{13}$$

$$|g^d_\varepsilon(x) - g^d_\varepsilon(y)| \leqslant c\frac{(d^c + \|x\|)^\beta + (d^c + \|y\|)^\beta}{2\sqrt{T}}\|x - y\|, \tag{14}$$

$$|g^d_\varepsilon(x)| \leqslant c(d^c + \|x\|)^\beta, \quad \max\left\{\|\mu^d_\varepsilon(0)\|, \|\sigma^d_\varepsilon(0)\|, |Tf(0)|, |g^d_\varepsilon(0)|\right\} \leqslant cd^c, \tag{15}$$

$$\max\{\|\mu^d_\varepsilon(x) - \mu^d(x)\|, \|\sigma^d_\varepsilon(x) - \sigma^d(x)\|, \|g^d_\varepsilon(x) - g^d(x)\|\} \leqslant \varepsilon c d^c(d^c + \|x\|)^\beta, \tag{16}$$

$$\max\left\{\mathcal{P}(\Phi_{g^d_\varepsilon}), \mathcal{P}(\Phi_{\mu^d_\varepsilon}), \mathcal{P}(\Phi_{\sigma^d_\varepsilon, 0})\right\} \leqslant cd^c\varepsilon^{-c}. \tag{17}$$

*Then the following items are true.*

*(i) For every* $d \in \mathbb{N}$ *there exists a unique at most polynomially growing viscosity solution* $u^d$ *of*

$$\frac{\partial u^d}{\partial t}(t, x) + \frac{1}{2}\mathrm{trace}(\sigma^d(x)(\sigma^d(x))^\top(\mathrm{Hess}_x u^d(t, x))) + \langle \mu^d(x), (\nabla_x u^d)(t, x)\rangle + f(u^d(t, x)) = 0 \tag{18}$$

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

*(ii) There exists* $(C_\delta)_{\delta\in(0,1)} \subseteq (0, \infty)$, $\eta \in (0, \infty)$, $(\Psi_{d,\epsilon})_{d\in\mathbb{N}, \epsilon\in(0,1)} \subseteq \mathbf{N}$ *such that for all* $d \in \mathbb{N}$, $\epsilon \in (0, 1)$ *we have that* $\mathcal{R}(\Psi_{d,\epsilon}) \in C(\mathbb{R}^d, \mathbb{R})$,

$$\mathcal{P}(\Psi_{d,\epsilon}) \leqslant C_\delta \eta d^\eta \epsilon^{-(4+\delta)-6c}, \quad \textit{and} \quad \left(\int_{[0,1]^d} \left|(\mathcal{R}(\Psi_{d,\epsilon}))(x) - u^d(0, x)\right|^{\mathfrak{p}} dx\right)^{\frac{1}{\mathfrak{p}}} < \epsilon. \tag{19}$$

Let us make some comments on the mathematical objects in Theorem 1.3. First of all, in Theorem 1.3 we consider different types of activation functions. The function $\mathfrak{a}_\nu$ is the ReLU activation if $\nu = \alpha = 0$, the leaky ReLU activation if $\nu = 0$ and $\alpha \in (0, 1)$, or the softplus activation if $\nu = 1$. Next, the assumptions above (13) ensure that the functions $g^d_\varepsilon, \mu^d_\varepsilon, \sigma^d_\varepsilon$, which approximate the terminal condition and the linear part of the PDE are DNNs. The bound $cd^c\varepsilon^{-c}$ in (17), which is a polynomial of $d$ and $\varepsilon^{-1}$, ensures that the functions $\mu^d_\varepsilon, \sigma^d_\varepsilon, g^d_\varepsilon$ are DNNs whose corresponding numbers of parameters grow without the curse of dimensionality. Under these assumptions Theorem 1.3 states that, roughly speaking, if DNNs can approximate the terminal condition and the linear part of the PDE in (18) without the curse of dimensionality, then they can also approximate its solution without the curse of dimensionality. More precisely, we show in (19) that for every dimension $d \in \mathbb{N}$ and for every accuracy $\epsilon \in (0,1)$ the $L^{\mathfrak{p}}(dx)$-expression error of the unique viscosity solution of the nonlinear PDE (18) is $\epsilon$ and the number of parameters of the DNNs is upper bounded polynominally in $d$ and $\epsilon^{-1}$. Therefore, the approximation rates are free from the curse of dimensionality. We refer to [1, 16, 17, 31, 42, 52, 54] for similar results obtained for PDEs without any non-local/ jump term.

1.5. **Sketch of the proofs.** Since (ii) in Theorem 1.1 contains an $L^{\mathfrak{p}}$-estimate we first need to prove $L^{\mathfrak{p}}$-estimates for MLP approximations (cf. Theorem 2.3 and Lemma 2.4), which, to the best of our knowledge, still do not appear in the scientific literature for general $\mathfrak{p} \in [2, \infty)$. The main tool to get $L^{\mathfrak{p}}$-estimates is the Marcinkiewicz-Zygmund inequality (see [67, Theorem 2.1]). In addition, Lemma 2.4 is the $L^{\mathfrak{p}}$-version of [41, Proposition 4.1]. Our first main result, Theorem 1.1, is a direct consequence of Theorem 2.3 and Lemma 2.4 and its proof is presented directly after the proof of Lemma 2.4. From the technical point of view, the main novelty of Theorem 1.1 is the sequence $(M_n)_{n\in\mathbb{N}}$. In the $L^2$-case we can simply choose $M_n = n$.

Theorem 1.3 follows from Theorem 4.1 and Lemmas 3.1, 3.2, 3.4, and 3.5. We present the proof of Theorem 1.3 after the proof of Theorem 4.1. Let us sketch the proof of Theorem 4.1. Although the result presented in Theorem 4.1 is purely deterministic, we use probabilistic arguments to prove its statement. More precisely, we employ the theory of full history recursive MLP approximations, which are numerical approximation methods for which it is now (cf. Theorem 1.1) known that they overcome the curse of dimensionality. We refer to [60] for the convergence analysis of MLP algorithms for semilinear PIDEs and to [8, 9, 26, 37, 38, 39, 40, 41, 44, 45, 47, 57] for corresponding results proving that MLP algorithms overcome the curse of dimensionality for PDEs without any non-local/ jump term in $L^2$-sense.

The main strategy of the proof of Theorem 4.1, roughly speaking, is to demonstrate that these MLP approximations can be represented by DNNs, if the coefficients determining the linear part, the terminal condition, and the nonlinear part are corresponding DNNs (cf. Lemma 3.15). Such ideas have been successfully applied to prove that DNNs overcome the curse of dimensionality in the numerical approximations of semilinear heat equations (see [1,42]) as well as semilinear Kolmogorov PDEs (see [17]). We also refer to [31, 52] for results proving that DNNs overcome the curse of dimensionality when approximating linear PDEs.

More precisely, we represent $u^d$ as solution of the stochastic fixed point equation (SFPE) (176) where the forward processes $(X_s^{d,\theta,t,x})$ are defined by (172) with drift $\mu^d$ and diffusion $\sigma^d$. We define the MLP approximations in (170) involving the Euler-Maruyama approximations in (169). Each $U_{n,m}^{d,\theta,K,\varepsilon}$ can be considered as approximation of the solution $u^d$ to the PDE (18). In order to estimate the approximation error $U_{n,m}^{d,\theta,K,\varepsilon} - u^d$ we decompose $U_{n,m}^{d,\theta,K,\varepsilon} - u^d = U_{n,m}^{d,\theta,K,\varepsilon} - u^{d,\varepsilon} + u^{d,\varepsilon} - u^d$ where $u^{d,\varepsilon}$ is defined by SFPE (177) where the forward processes $X^{d,\theta,\varepsilon,t,x}$ here are defined by (171) with drift $\mu_\varepsilon^d$ and diffusion $\sigma_\varepsilon^d$, which are DNN functions. The error $U_{n,m}^{d,\theta,K,\varepsilon} - u^{d,\varepsilon}$ is the error bound for an MLP approximation involving Euler-Maruyama approximations and therefore can be established in Lemma 2.4 (see (187)). The error $u^{d,\varepsilon} - u^d$ can be estimated, as in the $L^2$ case, by the perturbation result in [17, Lemma 2.3]. The main difficulty in the case of leaky ReLU and softplus activation is, compared to the case with ReLU, that here we have another definition of the operator $\odot$ than that in, e.g., [17] (see Setting 3.6 and Lemma 3.11) and as a consequence we need to rebuild the whole DNN calculus.

The paper is organized as follows. In Section 2 we establish $L^{\mathfrak{p}}$-estimates for MLP approximations and prove our first main result, Theorem 1.1. In Section 3 we study DNN representations for MLP approximations for PDEs of the form (18). In Section 4 we use the main representations in Section 3 to prove our second main result, Theorem 1.3.

## 2. MLP APPROXIMATIONS

2.1. **Error bounds for abstract MLP approximations.** In this section we establish $L^{\mathfrak{p}}$-estimate for MLP approximations. More precisely, we extend [41, Corollary 3.12] and [41, Proposition 4.1] to $L^{\mathfrak{p}}$-estimates. First of all, we work with an abstract MLP setting, Setting 2.1, and prove $L^{\mathfrak{p}}$-error estimates, see Theorem 2.3. The main difference between the general $L^{\mathfrak{p}}$-case and the $L^2$-case is that in the $L^{\mathfrak{p}}$-case we appeal to the Marcinkiewicz-Zygmund inequality (see [67, Theorem 2.1]). Having proven the $L^{\mathfrak{p}}$-error estimate we easily prove the $L^{\mathfrak{p}}$-error estimate for MLP approximations involving Euler-Maruyama approximations, see Lemma 2.4.

**Setting 2.1.** *Let $d \in \mathbb{N}$, $p_v \in [1,\infty)$, $c, T \in (0,\infty)$, $f \in C(\mathbb{R},\mathbb{R})$, $g \in C(\mathbb{R}^d,\mathbb{R})$, $V \in C([0,T]\times\mathbb{R}^d,[1,\infty))$, $\Theta = \bigcup_{n\in\mathbb{N}}\mathbb{Z}^n$. Let $\mathfrak{t}^\theta\colon \Omega \to [0,1]$, $\theta\in\Theta$, be identically distributed and independent random variables which satisfy for all $t\in[0,1]$ that $\mathbb{P}(\mathfrak{t}^0 \leqslant t) = t$. Let $(X_t^{\theta,s,x})_{s\in[0,T],t\in[s,T],x\in\mathbb{R}^d}\colon \{(\sigma,\tau)\in[0,T]^2\colon \sigma\leqslant\tau\}\times\mathbb{R}^d\times\Omega\to\mathbb{R}^d$, $\theta\in\Theta$, be measurable and identically distributed and independent. Assume that $(X_t^{\theta,s,x})_{s\in[0,T],t\in[s,T],x\in\mathbb{R}^d,\theta\in\Theta}$ and $(\mathfrak{t}^\theta)_{\theta\in\Theta}$ are independent. Assume for all $s\in[0,T]$, $t\in[s,T]$, $x\in\mathbb{R}^d$, $w_1,w_2\in\mathbb{R}$ that*

$$|g(x)| \leqslant V(T,x), \quad |Tf(0)| \leqslant V(s,x), \tag{20}$$

$$|f(w_1)-f(w_2)| \leqslant c|w_1-w_2|, \tag{21}$$

$$\left\|V(t,X_t^{0,s,x})\right\|_{p_v} \leqslant V(s,x). \tag{22}$$

*Let $U_{n,m}^\theta\colon[0,T]\times\mathbb{R}^d\times\Omega\to\mathbb{R}$, $n,m\in\mathbb{Z}$, $\theta\in\Theta$, satisfy for all $n,m\in\mathbb{N}$, $\theta\in\Theta$, $t\in[0,T]$, $x\in\mathbb{R}^d$ that $U_{-1,m}^\theta(t,x) = U_{0,m}^\theta(t,x) = 0$ and*

$$\begin{aligned} U_{n,m}^\theta(t,x) &= \frac{1}{m^n}\sum_{i=1}^{m^n} g(X_T^{(\theta,0,-i),t,x}) \\ &+ \sum_{\ell=0}^{n-1}\frac{T-t}{m^{n-\ell}}\sum_{i=1}^{m^{n-\ell}}\left(f\circ U_{\ell,m}^{(\theta,\ell,i)} - \mathbb{1}_{\mathbb{N}}(\ell) f\circ U_{\ell-1,m}^{(\theta,-\ell,i)}\right)\left(t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}, X_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}}^{(\theta,\ell,i),t,x}\right). \end{aligned} \tag{23}$$

**Lemma 2.2** (Independence and distributional properties). *Assume Setting 2.1. Then*

*(i) it holds for all $n\in\mathbb{N}_0$, $m\in\mathbb{N}$, $\theta\in\Theta$ that $U_{n,m}^\theta$ and $f\circ U_{n,m}^\theta$ are measurable,*

*(ii) it holds[1] for all $n\in\mathbb{N}_0$, $m\in\mathbb{N}$, $\theta\in\Theta$ that*

$$\mathfrak{S}((U_{n,m}^\theta(t,x))_{t\in[0,T],x\in\mathbb{R}^d}) \subseteq \mathfrak{S}((\mathfrak{t}^{(\theta,\vartheta)})_{\vartheta\in\Theta}, (X_t^{(\theta,\vartheta),s,x})_{\vartheta\in\Theta,s\in[0,T],t\in[s,T],x\in\mathbb{R}^d}), \tag{24}$$

*(iii) it holds for all $n\in\mathbb{N}_0$, $m\in\mathbb{N}$, $\theta\in\Theta$ that $(U_{n,m}^\theta(t,x))_{t\in[0,T],x\in\mathbb{R}^d}$, $(X_t^{\theta,s,x})_{s\in[0,T],t\in[s,T],x\in\mathbb{R}^d}$, and $\mathfrak{t}^\theta$ are independent,*

*(iv) it holds for all $n,m\in\mathbb{N}_0$, $m\in\mathbb{N}$, $i,j,k,\ell,\nu\in\mathbb{Z}$, $\theta\in\Theta$ with $(i,j)\neq(k,l)$ that $(U_{n,m}^{(\theta,i,j)}(t,x))_{t\in[0,T],x\in\mathbb{R}^d}$, $(U_{n,m}^{(\theta,k,\ell)}(t,x))_{t\in[0,T],x\in\mathbb{R}^d}$, $\mathfrak{t}^{(\theta,i,j)}$, and $(X_t^{(\theta,i,j),s,x})_{s\in[0,T],t\in[s,T],x\in\mathbb{R}^d}$ are independent, and*

*(v) it holds for all $n\in\mathbb{N}_0$, $m\in\mathbb{N}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ that $U_{n,m}^\theta(t,x)$, $\theta\in\Theta$, are identically distributed.*

*Proof of Lemma 2.2.* See [41, Lemma 3.2]. □

**Theorem 2.3** ($L^p$-error estimates, $p\in[2,\infty)$, for MLP approximations). *Assume Setting 2.1. Let $p\in[2,\infty)$, $q_1\in[1,\infty)$ satisfy that $pq_1\leqslant p_v$. Then*

*(i) there exists a unique measurable $u\colon[0,T]\times\mathbb{R}^d\to\mathbb{R}$ which satisfies for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that that $\mathbb{E}[|g(X_T^{0,t,x})|] + \int_t^T \mathbb{E}[|f(u(s,X_s^{0,t,x}))|]\,ds + \sup_{y\in\mathbb{R}^d,s\in[0,T]}\frac{|u(s,y)|}{V(s,y)} < \infty$ and*

$$u(t,x) = \mathbb{E}[g(X_T^{0,t,x})] + \int_t^T \mathbb{E}[f(u(s,X_s^{0,t,x}))]\,ds \tag{25}$$

*and*

*(ii) we have for all $m,n\in\mathbb{N}$, $t\in[0,T]$, $x\in\mathbb{R}^d$*

$$\left\|U_{n,m}^0(t,x)-u(t,x)\right\|_p \leqslant 2(p-1)^{\frac{n}{2}} e^{5cTn} e^{m^{p/2}/p} m^{-n/2} V^{q_1}(t,x). \tag{26}$$

[1] Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $n\in\mathbb{N}$, and let $(S_k,\mathcal{S}_k)$, $k\in\{1,2,\dots,n\}$, be measurable spaces. Note that for all $X_k\colon\Omega\to S_k$, $k\in\{1,2,\dots,n\}$, it holds that $\mathfrak{S}(X_1,X_2,\dots,X_n)$ is the smallest sigma-algebra on $\Omega$ with respect to which $X_1,X_2,\dots,X_n$ are measurable.

*Proof of Theorem 2.3.* For every random field $H\colon [0,T]\times\mathbb{R}^d\times\Omega\to\mathbb{R}$ and every $s\in[0,T]$ let $|\!|\!|H|\!|\!|_s\in[0,\infty]$ satisfy that

$$|\!|\!|H|\!|\!|_s=\sup_{t\in[s,T],x\in\mathbb{R}^d}\frac{\|H(t,x)\|_p}{(V(t,x))^{q_1}}. \tag{27}$$

Furthermore, for every random variable $\mathfrak{X}\colon\Omega\to\mathbb{R}$ with $\mathbb{E}[|\mathfrak{X}|]<\infty$ let $\mathbb{V}_p(\mathfrak{X})\in[0,\infty]$ satisfy that $\mathbb{V}_p(\mathfrak{X})=\|\mathfrak{X}-\mathbb{E}[\mathfrak{X}]\|_p^2$.

First, measurability and [41, Proposition 2.2] (applied with $d\curvearrowleft d$, $T\curvearrowleft T$, $L\curvearrowleft c$, $\mathcal{O}\curvearrowleft\mathbb{R}^d$, $(X^x_{t,s})_{t\in[0,T],s\in[t,T],x\in\mathbb{R}^d}\curvearrowleft(X^{0,t,x}_s)_{t\in[0,T],s\in[t,T],x\in\mathbb{R}^d}$, $f\curvearrowleft([0,T]\times\mathbb{R}^d\times\mathbb{R}\ni(t,x,w)\mapsto f(w)\in\mathbb{R})$, $g\curvearrowleft g$, $V\curvearrowleft V$ in the notation of [41, Proposition 2.2]) show that there exists a unique measurable $u\colon[0,T]\times\mathbb{R}^d\to\mathbb{R}$ which satisfies for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that $\mathbb{E}[|g(X^{0,t,x}_T)|]+\int_t^T\mathbb{E}[|f(u(s,X^{0,t,x}_s))|]\,ds+\sup_{y\in\mathbb{R}^d,s\in[0,T]}\frac{|u(s,y)|}{V(s,y)}<\infty$ and

$$u(t,x)=\mathbb{E}[g(X^{0,t,x}_T)]+\int_t^T\mathbb{E}[f(u(s,X^{0,t,x}_s))]\,ds \tag{28}$$

and we have for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that $\frac{|u(t,x)|}{V(t,x)}\leqslant 2e^{c(T-t)}$. This, the fact that $q_1\geqslant 1$, and (27) imply for all $s\in[0,T]$ that

$$|\!|\!|u|\!|\!|_s\leqslant 2e^{cT}. \tag{29}$$

This proves (i).

Next, Jensen's inequality, the fact that $p\leqslant p_v$, (22), and the fact that $V\leqslant V^{q_1}$ show for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$\|g(X^{0,t,x}_T)\|_p\leqslant\|V(T,X^{0,t,x}_T)\|_p\leqslant\|V(T,X^{0,t,x}_T)\|_{p_v}\leqslant V(t,x)\leqslant V^{q_1}(t,x). \tag{30}$$

Next, the disintegration theorem, the measurability and independence properties, the fact that $pq_1\leqslant p_v$, and Jensen's inequality prove for all $t\in[0,T]$, $\ell,\nu\in\mathbb{N}_0$, $x\in\mathbb{R}^d$, $H\in\operatorname{span}_{\mathbb{R}}(\{f\circ U^\nu_{\ell,m},f\circ u\})$ that

$$\begin{aligned}\left\|(T-t)H(t+(T-t)\mathfrak{t}^0,X^{0,t,x}_{t+(T-t)\mathfrak{t}^0})\right\|_p&=(T-t)\left\|\left\|\|H(r,y)\|_p\big|_{y=X^{0,t,x}_r}\right\|_p\Big|_{r=t+(T-t)\mathfrak{t}^0}\right\|_p\\&\leqslant(T-t)\left\|\left[|\!|\!|H|\!|\!|_r\left\|V^{q_1}(r,X^{0,t,x}_r)\right\|_p\right]\Big|_{r=t+(T-t)\mathfrak{t}^0}\right\|_p\\&\leqslant(T-t)\left\||\!|\!|H|\!|\!|_{t+(T-t)\mathfrak{t}^0}\right\|_p V^{q_1}(t,x)\end{aligned} \tag{31}$$

Moreover, (27) and (21) show for all $t\in[0,T]$ and all random fields $H,K\colon[0,T]\times\mathbb{R}^d\times\Omega\to\mathbb{R}$ that $|\!|\!|(f\circ H)-(f\circ K)|\!|\!|_t\leqslant c|\!|\!|H-K|\!|\!|_t$. This, (31), and the independence and distributional properties imply for all $t\in[0,T]$, $\nu,\ell\in\mathbb{N}_0$, $m,n\in\mathbb{N}$, $x\in\mathbb{R}^d$ that

$$\begin{aligned}&\left\|(T-t)((f\circ U^\nu_{\ell,m})-(f\circ u))(t+(T-t)\mathfrak{t}^0,X^{0,t,x}_{t+(T-t)\mathfrak{t}^0})\right\|_p\\&\leqslant(T-t)\left\||\!|\!|(f\circ U^\nu_{\ell,m})-(f\circ u)|\!|\!|_{t+(T-t)\mathfrak{t}^0}\right\|_p V^{q_1}(t,x)\\&\leqslant(T-t)c\left\||\!|\!|U^0_{\ell,m}-u|\!|\!|_{t+(T-t)\mathfrak{t}^0}\right\|_p V^{q_1}(t,x).\end{aligned} \tag{32}$$

This and the triangle inequality show for all $t \in [0,T]$, $x \in \mathbb{R}^d$, $m, \ell \in \mathbb{N}$ that

$$\begin{aligned}
&\left\| (T-t)\left[ ((f \circ U^0_{\ell,m}) - (f \circ U^1_{\ell-1,m}))(t+(T-t)\mathfrak{t}^0, X^{0,t,x}_{t+(T-t)\mathfrak{t}^0}) \right] \right\|_p \\
&\leqslant \left\| (T-t)\left[ ((f \circ U^0_{\ell,m}) - (f \circ u))(t+(T-t)\mathfrak{t}^0, X^{0,t,x}_{t+(T-t)\mathfrak{t}^0}) \right] \right\|_p \\
&\quad + \left\| (T-t)\left[ ((f \circ U^1_{\ell-1,m}) - (f \circ u))(t+(T-t)\mathfrak{t}^0, X^{0,t,x}_{t+(T-t)\mathfrak{t}^0}) \right] \right\|_p \\
&\leqslant \sum_{j=\ell-1}^{\ell} \left[ (T-t)c \left\| \left\| U^0_{j,m} - u \right\|_{t+(T-t)\mathfrak{t}^0} \right\|_p \right] V^{q_1}(t,x).
\end{aligned} \tag{33}$$

This, (23), the triangle inequality, the fact that $\forall\, m \in \mathbb{N}\colon U^0_{0,m} = 0$, the independence and distributional properties, (30), (29), (33), and induction prove for all $n \in \mathbb{N}_0$, $m \in \mathbb{N}$, $x \in \mathbb{R}^d$, $t \in [0,T]$, $\theta \in \Theta$ that

$$\left\| U^\theta_{n,m} \right\|_t + \left\| (T-t)(f \circ U^\theta_{n,m})(t+(T-t)\mathfrak{t}^\theta, X^{\theta,t,x}_{t+(T-t)\mathfrak{t}^\theta}) \right\|_p < \infty. \tag{34}$$

Next, linearity, the independence and distributional properties, and a telescoping sum argument prove for all $n, m \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\mathbb{E}[U^0_{n,m}(t,x)] = \frac{1}{m^n}\sum_{i=1}^{m^n} \mathbb{E}[g(X^{(0,0,-i),t,x}_T)] \\
&+ \sum_{\ell=0}^{n-1} \frac{T-t}{m^{n-\ell}} \sum_{i=1}^{m^{n-\ell}} \mathbb{E}\Big[ \Big(f \circ U^{(0,\ell,i)}_{\ell,m} - \mathbb{1}_{\mathbb{N}}(\ell) f \circ U^{(0,-\ell,i)}_{\ell-1,m}\Big)\Big(t+(T-t)\mathfrak{t}^{(0,\ell,i)}, X^{(0,\ell,i),t,x}_{t+(T-t)\mathfrak{t}^{(0,\ell,i)}}\Big)\Big] \\
&= \mathbb{E}[g(X^{0,t,x}_T)] + \sum_{\ell=0}^{n-1}(T-t)\Bigg[ \mathbb{E}\Big[ (f \circ U^0_{\ell,m})(t+(T-t)\mathfrak{t}^0, X^{0,t,x}_{t+(T-t)\mathfrak{t}^0}) \Big] \\
&\qquad\qquad - \mathbb{E}\Big[ (f \circ U^0_{\ell-1,m})(t+(T-t)\mathfrak{t}^0, X^{0,t,x}_{t+(T-t)\mathfrak{t}^0}) \Big] \Bigg] \\
&= \mathbb{E}[g(X^{0,t,x}_T)] + (T-t)\mathbb{E}\Big[ (f \circ U^0_{n-1,m})(t+(T-t)\mathfrak{t}^0, X^{0,t,x}_{t+(T-t)\mathfrak{t}^0}) \Big].
\end{aligned} \tag{35}$$

Moreover, the disintegration theorem and the independence and distributional properties show for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$u(t,x) = \mathbb{E}\big[g(X^{0,t,x}_T)\big] + (T-t)\mathbb{E}\Big[(f \circ u)(t+(T-t)\mathfrak{t}^0, X^{0,t,x}_{t+(T-t)\mathfrak{t}^0})\Big] \tag{36}$$

This, the triangle inequality, (35), Jensen's inequality, and (32) prove for all $n, m \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\frac{\left|\mathbb{E}[U^0_{n,m}(t,x)] - u(t,x)\right|}{V^{q_1}(t,x)} \leqslant (T-t)c \left\| \left\| U^0_{n-1,m} - u \right\|_{t+(T-t)\mathfrak{t}^0} \right\|_p. \tag{37}$$

Moreover, the Marcinkiewicz-Zygmund inequality (see [67, Theorem 2.1]), the fact that $p \in [2,\infty)$, the triangle inequality, and Jensen's inequality show that for all $n \in \mathbb{N}$ and all identically distributed and independent random variables $\mathfrak{X}_k$, $k \in [1,n] \cap \mathbb{Z}$, with $\mathbb{E}[|\mathfrak{X}_1|] < \infty$ it holds that

$$\begin{aligned}
\left( \mathbb{V}_p\left[ \frac{1}{n}\sum_{k=1}^n \mathfrak{X}_k \right] \right)^{1/2} &= \frac{1}{n}\left\| \sum_{k=1}^n (\mathfrak{X}_k - \mathbb{E}[\mathfrak{X}_k]) \right\|_p \leqslant \frac{\sqrt{p-1}}{n}\left( \sum_{k=1}^n \|\mathfrak{X}_k - \mathbb{E}[\mathfrak{X}_k]\|_p^2 \right)^{\frac{1}{2}} \\
&\leqslant \frac{2\sqrt{p-1}\|\mathfrak{X}_1\|_p}{\sqrt{n}}.
\end{aligned} \tag{38}$$

This, (23), the triangle inequality, the independence and distributional properties, (30), and (33) show for all $n, m \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$
\begin{aligned}
&\frac{\left\|U^0_{n,m}(t,x) - \mathbb{E}[U^0_{n,m}(t,x)]\right\|_p}{V^{q_1}(t,x)} = \frac{\left(\mathbb{V}_p(U^0_{n,m}(t,x))\right)^{\frac{1}{2}}}{V^{q_1}(t,x)}\\
&\leqslant \frac{\left(\mathbb{V}_p\left[\frac{1}{m^n}\sum_{i=1}^{m^n} g(X_T^{(0,0,-i),t,x})\right]\right)^{\frac{1}{2}}}{V^{q_1}(t,x)}\\
&\quad + \sum_{\ell=0}^{n-1} \frac{\left(\mathbb{V}_p\left[\frac{T-t}{m^{n-\ell}}\sum_{i=1}^{m^{n-\ell}}\left(f\circ U^{(0,\ell,i)}_{\ell,m} - \mathbb{1}_{\mathbb{N}}(\ell) f\circ U^{(0,-\ell,i)}_{\ell-1,m}\right)\left(t+(T-t)\mathfrak{t}^{(0,\ell,i)}, X^{(0,\ell,i),t,x}_{t+(T-t)\mathfrak{t}^{(0,\ell,i)}}\right)\right]\right)^{\frac{1}{2}}}{V^{q_1}(t,x)}\\
&\leqslant \frac{\frac{2\sqrt{p-1}\left\|g(X_T^{0,t,x})\right\|_p}{\sqrt{m^n}} + \sum_{\ell=1}^{n-1}\frac{2\sqrt{p-1}\left\|(T-t)\left(f\circ U^0_{\ell,m} - \mathbb{1}_{\mathbb{N}}(\ell) f\circ U^1_{\ell-1,m}\right)\left(t+(T-t)\mathfrak{t}^0, X^{0,t,x}_{t+(T-t)\mathfrak{t}^0}\right)\right\|_p}{\sqrt{m^{n-\ell}}}}{V^{q_1}(t,x)}\\
&\leqslant \frac{2\sqrt{p-1}}{\sqrt{m^n}} + \sum_{\ell=1}^{n-1}\sum_{j=\ell-1}^{\ell}\frac{2\sqrt{p-1}}{\sqrt{m^{n-\ell}}}\left[(T-t)c\left\|\left\|\left|U^0_{j,m} - u\right\|\right|_{t+(T-t)\mathfrak{t}^0}\right\|_p\right].
\end{aligned}
\tag{39}
$$

In addition, the fact that $\mathfrak{t}^0$ is uniformly distributed on $[0,1]$ and the substitution rule imply for all $s \in [0,T]$, $t \in [0,T]$, and all measurable $h\colon [0,T] \to \mathbb{R}$ that

$$
(T-t)\left\|h(t+(T-t)\mathfrak{t}^0)\right\|_p = (T-t)^{1-\frac{1}{p}}\left[\int_0^1 (T-t)|h(t+(T-t)\lambda)|^p\,d\lambda\right]^{\frac{1}{p}} \tag{40}
$$

$$
= (T-t)^{1-\frac{1}{p}}\left[\int_t^T |h(\zeta)|^p\,d\zeta\right]^{\frac{1}{p}} \leqslant (T-s)^{1-\frac{1}{p}}\left[\int_s^T |h(\zeta)|^p\,d\zeta\right]^{\frac{1}{p}}. \tag{41}
$$

This, (27), the triangle inequality, (37), (39), and the fact that $\forall\, n, m \in \mathbb{N}, a_0, a_1, \ldots, a_{n-1} \in [0,\infty)\colon (\sum_{\ell=1}^{n-1}\sum_{j=\ell-1}^{\ell}\frac{a_j}{\sqrt{m^{n-\ell}}}) + a_{n-1} \leqslant \sum_{\ell=0}^{n-1}\frac{2a_\ell}{\sqrt{m^{n-\ell-1}}}$ prove for all $n, m \in \mathbb{N}$, $s \in [0,T]$ that

$$
\begin{aligned}
\left\|\left|U^0_{n,m} - u\right\|\right|_s &= \sup_{t\in[s,T]}\frac{\left\|U^0_{n,m}(t,x) - u(t,x)\right\|_p}{V^{q_1}(t,x)}\\
&\leqslant \sup_{t\in[s,T]}\frac{\left\|U^0_{n,m}(t,x) - \mathbb{E}[U^0_{n,m}(t,x)]\right\|_p + \left|\mathbb{E}[U^0_{n,m}(t,x)] - u(t,x)\right|}{V^{q_1}(t,x)}\\
&\leqslant \sup_{t\in[s,T]}\left[\frac{2\sqrt{p-1}}{\sqrt{m^n}} + \sum_{\ell=1}^{n-1}\sum_{j=\ell-1}^{\ell}\left[\frac{2\sqrt{p-1}}{\sqrt{m^{n-\ell}}}(T-t)c\left\|\left\|\left|U^0_{j,m} - u\right\|\right|_{t+(T-t)\mathfrak{t}^0}\right\|_p\right]\right.\\
&\qquad\left. + (T-t)c\left\|\left\|\left|U^0_{n-1,m} - u\right\|\right|_{t+(T-t)\mathfrak{t}^0}\right\|_p\right]\\
&\leqslant \sup_{t\in[s,T]}\left[\frac{2\sqrt{p-1}}{\sqrt{m^n}} + \sum_{\ell=0}^{n-1}\left[\frac{4\sqrt{p-1}}{\sqrt{m^{n-\ell-1}}}(T-t)c\left\|\left\|\left|U^0_{\ell,m} - u\right\|\right|_{t+(T-t)\mathfrak{t}^0}\right\|_p\right]\right]\\
&\leqslant \frac{2\sqrt{p-1}}{\sqrt{m^n}} + \sum_{\ell=0}^{n-1}\left[\frac{4\sqrt{p-1}}{\sqrt{m^{n-\ell-1}}}(T-s)^{1-\frac{1}{p}}c\left[\int_s^T \left\|\left|U^0_{\ell,m} - u\right\|\right|_\zeta^p\,d\zeta\right]^{\frac{1}{p}}\right].
\end{aligned}
\tag{42}
$$

Next, [41, Lemma 3.11] (applied for every $s \in [0,T]$, $n, m \in \mathbb{N}$ with $M \curvearrowleft m$, $N \curvearrowleft n$, $\tau \curvearrowleft s$, $a \curvearrowleft 2\sqrt{p-1}$, $b \curvearrowleft 4(T-s)^{1-\frac{1}{p}}c\sqrt{p-1}$, $(f_j)_{j\in\mathbb{N}_0} \curvearrowleft ([s,T] \ni t \mapsto \left\|\left|U^0_{j,m} - u\right\|\right|_t \in [0,\infty])_{j\in\mathbb{N}_0}$ in the notation of [41, Lemma 3.11]), (29), and the fact that $\forall\, m \in \mathbb{N}\colon U^0_{0,m} = 0$ prove for all $m, n \in \mathbb{N}$,

$s \in [0,T]$ that

$$\begin{aligned}
\left\|\left\| U^0_{n,m} - u \right\|\right\|_s &\leqslant \left( 2\sqrt{p-1} + 4(T-s)^{1-\frac{1}{p}} c\sqrt{p-1} \cdot (T-s)^{\frac{1}{p}} \cdot \sup_{t\in[s,T]} \left\|\left| u \right|\right\|_t \right) \\
&\qquad \cdot e^{m^{p/2}/p} m^{-n/2} \left( 1 + 4(T-s)^{1-\frac{1}{p}} c\sqrt{p-1} \cdot (T-s)^{\frac{1}{p}} \right)^{n-1} \\
&\leqslant \sqrt{p-1}\left(2 + 4cT \cdot 2e^{cT}\right) e^{m^{p/2}/p} m^{-n/2} \left( \sqrt{p-1}(1+4cT) \right)^{n-1} \\
&\leqslant 2(p-1)^{\frac{n}{2}} e^{cT}(1+4cT) e^{m^{p/2}/p} m^{-n/2} (1+4cT)^{n-1} \\
&\leqslant 2(p-1)^{\frac{n}{2}} e^{5cTn} e^{m^{p/2}/p} m^{-n/2}.
\end{aligned} \tag{43}$$

This and (27) imply for all $m,n \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\left\| U^0_{n,m}(t,x) - u(t,x) \right\|_p \leqslant 2(p-1)^{\frac{n}{2}} e^{5cTn} e^{m^{p/2}/p} m^{-n/2} V^{q_1}(t,x). \tag{44}$$

This completes the proof of Theorem 2.3. ☐

### 2.2. Error bounds for MLP approximations involving Euler-Maruyama approximations.

Lemma 2.4 below extends [41, Proposition 4.1] to an $L^{\mathfrak{p}}$-estimate, $\mathfrak{p} \in [2,\infty)$. Its proof can be easily adapted from that of [41, Proposition 4.1]. However, we present it here for convenience of the reader.

**Lemma 2.4.** *Let $d,K \in \mathbb{N}$, $T \in (0,\infty)$, $\mathfrak{p} \in [2,\infty)$, $\beta, b, c \in [1,\infty)$, $p \in [\mathfrak{p}\beta,\infty)$, $\varphi \in C^2(\mathbb{R}^d,[1,\infty))$, $g \in C(\mathbb{R}^d,\mathbb{R})$, $f \in C(\mathbb{R},\mathbb{R})$, $\mu \in C(\mathbb{R}^d,\mathbb{R}^d)$, $\sigma \in C(\mathbb{R}^d,\mathbb{R}^{d\times d})$. Assume for all $x,y \in \mathbb{R}^d$, $z \in \mathbb{R}^d \backslash \{0\}$, $t \in [0,T]$, $v,w \in \mathbb{R}$ that*

$$\max\left\{ \frac{|(\varphi'(x))(z)|}{(\varphi(x))^{\frac{p-1}{p}}\|z\|}, \frac{(\varphi''(x))(z,z)}{(\varphi(x))^{\frac{p-2}{p}}\|z\|^2}, \frac{c\|x\| + \|\mu(0)\|}{(\varphi(x))^{\frac{1}{p}}}, \frac{c\|x\| + \|\sigma(0)\|}{(\varphi(x))^{\frac{1}{p}}} \right\} \leqslant c, \tag{45}$$

$$\max\{|Tf(0)|, |g(x)|\} \leqslant b(\varphi(x))^{\frac{\beta}{p}}, \tag{46}$$

$$|g(x) - g(y)| \leqslant b\frac{(\varphi(x)+\varphi(y))^{\frac{\beta}{p}}}{\sqrt{T}}\|x-y\|, \quad |f(v) - f(w)| \leqslant c|v-w|, \tag{47}$$

$$\max\{\|\mu(x) - \mu(y)\|, \|\sigma(x) - \sigma(y)\|\} \leqslant c\|x-y\|. \tag{48}$$

*Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ be a filtered probability space which satisfies the usual conditions[2]. Let $\Theta = \bigcup_{n\in\mathbb{N}} \mathbb{Z}^n$. Let $\mathfrak{t}^\theta \colon \Omega \to [0,1]$, $\theta \in \Theta$, be identically distributed and independent random variables. Assume for all $t \in (0,1)$ that $\mathbb{P}(\mathfrak{t}^0 \leqslant t) = t$. Let $W^\theta \colon [0,T]\times\Omega \to \mathbb{R}^d$, $\theta \in \Theta$, be independent standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motions. Assume that $(\mathfrak{t}^\theta)_{\theta\in\Theta}$ and $(W^\theta)_{\theta\in\Theta}$ are independent. Let $\lfloor\cdot\rfloor_K \colon \mathbb{R} \to \mathbb{R}$ satisfy for all $t \in \mathbb{R}$ that $\lfloor t\rfloor_K = \max(\{0, \frac{T}{K}, \ldots, \frac{(K-1)T}{T}, T\} \cap ((-\infty,t) \cup \{0\}))$. For every $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ let $Y^{\theta,t,x} = (Y^{\theta,t,x}_s)_{s\in[t,T]} \colon [t,T]\times\Omega \to \mathbb{R}^d$ satisfy for all $s \in [t,T]$ that $Y^{\theta,t,x}_t = x$ and*

$$Y^{\theta,t,x}_s = Y^{\theta,t,x}_{\max\{t,\lfloor s\rfloor_K\}} + \mu(Y^{\theta,t,x}_{\max\{t,\lfloor s\rfloor_K\}})(s - \max\{t,\lfloor s\rfloor_K\}) + \sigma(Y^{\theta,t,x}_{\max\{t,\lfloor s\rfloor_K\}})(W^\theta_s - W^\theta_{\max\{t,\lfloor s\rfloor_K\}}). \tag{49}$$

[2]Let $T \in [0,\infty)$ and let $\mathbf{\Omega} = (\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ be a filtered probability space. Then we say that $\mathbf{\Omega}$ satisfies the usual conditions if and only if it holds that $\{A \in \mathcal{F} : \mathbb{P}(A) = 0\} \subseteq \mathbb{F}_0$ and $\forall\, t \in [0,T) \colon \mathbb{F}_t = \cap_{s\in(t,T]}\mathbb{F}_s$.

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

$$\begin{aligned} U^\theta_{n,m}(t,x) &= \frac{1}{m^n}\sum_{i=1}^{m^n} g(Y_T^{(\theta,0,-i),t,x}) \\ &+ \sum_{\ell=0}^{n-1}\frac{T-t}{m^{n-\ell}}\sum_{i=1}^{m^{n-\ell}}\left(f\circ U^{(\theta,\ell,i)}_{\ell,m}-\mathbb{1}_{\mathbb{N}}(\ell)f\circ U^{(\theta,-\ell,i)}_{\ell-1,m}\right)\left(t+(T-t)\mathfrak{t}^{(\theta,\ell,i)},Y^{(\theta,\ell,i),t,x}_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}}\right). \end{aligned} \tag{50}$$

*Then the following items are true.*

*(i) For every $t\in[0,T]$, $\theta\in\Theta$ there exists an up to indistinguishability unique continuous random field $X^{\theta,t,\cdot}=(X^{\theta,t,x}_s)_{s\in[t,T],x\in\mathbb{R}^d}\colon [t,T]\times\mathbb{R}^d\times\Omega\to\mathbb{R}^d$ which satisfies that for all $x\in\mathbb{R}^d$ it holds that $(X^{\theta,t,x}_s)_{s\in[t,T]}$ is $(\mathbb{F}_s)_{s\in[t,T]}$-adapted and which satisfies that for all $s\in[t,T]$, $x\in\mathbb{R}^d$ it holds $\mathbb{P}$-a.s. that*

$$X^{\theta,t,x}_s = x+\int_t^s \mu(X^{\theta,t,x}_r)\,dr+\int_t^s\sigma(X^{\theta,t,x}_r)\,dW^\theta_r. \tag{51}$$

*(ii) For all $\theta\in\Theta$, $t\in[0,T]$, $s\in[t,T]$, $r\in[s,T]$, $x\in\mathbb{R}^d$ that $\mathbb{P}(X_r^{\theta,s,X_s^{\theta,t,x}}=X_r^{\theta,t,x})=1$.*

*(iii) There exists a unique measurable $u\colon[0,T]\times\mathbb{R}^d\to\mathbb{R}$ which satisfies for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that $\left(\sup_{s\in[0,T],y\in\mathbb{R}^d}[|u(s,y)|(\varphi(y))^{-\beta/p}]\right)+\int_t^T\mathbb{E}\big[|f(u(s,X^{0,t,x}_s))|\big]\,ds+\mathbb{E}\big[|g(X^{0,t,x}_T)|\big]<\infty$ and*

$$u(t,x)=\mathbb{E}\big[g(X^{0,t,x}_T)\big]+\int_t^T\mathbb{E}\big[f(u(s,X^{0,t,x}_s))\big]\,ds. \tag{52}$$

*(iv) For all $t\in[0,T]$, $x\in\mathbb{R}^d$, $n\in\mathbb{N}_0$, $m\in\mathbb{N}$ we have that $U^\theta_{n,m}(t,x)$ is measurable.*

*(v) For all $t\in[0,T]$, $x\in\mathbb{R}^d$, $m,n\in\mathbb{N}$ we have that*

$$\left\|U^0_{n,m}(t,x)-u(t,x)\right\|_{\mathfrak{p}}\leqslant 12bc^2e^{9c^3T}(\varphi(x))^{\frac{\beta+1}{p}}\left[2\mathfrak{p}^{\frac{n}{2}}e^{5cTn}e^{m^{\mathfrak{p}/2}/\mathfrak{p}}m^{-n/2}+\frac{1}{\sqrt{K}}\right]. \tag{53}$$

*Proof of Lemma 2.4.* Observe that (48) prove (i) and (ii). For the rest of the proof let $\Delta=\{(t,s)\in[0,T]^2\colon t\leqslant s\}$ and $\mathfrak{X}^k=(\mathfrak{X}^{k,t,x}_s)_{t\in[0,T],s\in[t,T],x\in\mathbb{R}^d}\colon\Delta\times\mathbb{R}^d\times\Omega\to\mathbb{R}^d$ satisfy for all $t\in[0,T]$, $s\in[t,T]$, $x\in\mathbb{R}^d$ that $\mathfrak{X}^{0,t,x}_s=X^{0,t,x}_s$ and $\mathfrak{X}^{1,t,x}_s=Y^{0,t,x}_s$. For every $x\in\mathbb{R}^d$ let $\mathfrak{Y}^x=(\mathfrak{Y}^x_t)_{t\in[0,T]}\colon[0,T]\times\Omega\to\mathbb{R}^d$ satisfy for all $t\in[0,T]$ that $\mathfrak{Y}^x_t=x+\mu(x)t+\sigma(x)W_t$. For every $x\in\mathbb{R}^d$, $n\in\mathbb{N}$ let $\tau^x_n\colon\Omega\to[0,T]$ satisfy that $\tau^x_n=\inf(\{T\}\cup\{t\in[0,T]\colon[\sup_{s\in[0,t]}\varphi(\mathfrak{Y}^x_s)]+\int_0^t\sum_{i=1}^d|(\varphi'(\mathfrak{Y}^x_s))(\sigma_i(x))|^2\,ds\geqslant n\})$. Next, the triangle inequality, (48), and (45) prove for all $x\in\mathbb{R}^d$ that

$$\begin{aligned}\max\{\|\mu(x)\|,\|\sigma(x)\|\}&\leqslant\max\{\|\mu(x)-\mu(0)\|+\|\mu(0)\|,\|\sigma(x)-\sigma(0)\|+\|\sigma(0)\|\}\\&\leqslant\max\{c\|x\|+\|\mu(0)\|,c\|x\|+\|\sigma(0)\|\}\leqslant c(\varphi(x))^{\frac{1}{p}}\end{aligned} \tag{54}$$

This, (45), and the fact that $\forall\, a,b \in [0,\infty), \lambda \in (0,1) \colon a^{\lambda} b^{1-\lambda} \leqslant \lambda a + (1-\lambda) b$ imply that for all $x,y \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
&|(\varphi'(y))(\mu(x))| + \frac{1}{2}\left|\sum_{k=1}^{d}(\varphi''(y))(\sigma_k(x),\sigma_k(x))\right| \\
&\leqslant c(\varphi(y))^{1-\frac{1}{p}}\|\mu(x)\| + \frac{c}{2}(\varphi(y))^{1-\frac{2}{p}}\sum_{k=1}^{d}\|\sigma_k(x)\|^2 \\
&= c(\varphi(y))^{1-\frac{1}{p}}\|\mu(x)\| + \frac{c}{2}(\varphi(y))^{1-\frac{2}{p}}\|\sigma(x)\|^2 \\
&\leqslant c(\varphi(y))^{1-\frac{1}{p}}c(\varphi(x))^{\frac{1}{p}} + \frac{c}{2}(\varphi(y))^{1-\frac{2}{p}}c^2(\varphi(x))^{\frac{2}{p}} \\
&\leqslant c^2\left[\left(1-\frac{1}{p}\right)\varphi(y) + \frac{1}{p}\varphi(x)\right] + \frac{c^3}{2}\left[\left(1-\frac{2}{p}\right)\varphi(y) + \frac{2}{p}\varphi(x)\right] \\
&\leqslant \left[c^3\left(1-\frac{1}{p}\right) + \frac{c^3}{2}\left(1-\frac{2}{p}\right)\right]\varphi(y) + \left[\frac{c^3}{p} + \frac{2c^3}{2p}\right]\varphi(x) \\
&= \left(\frac{3c^3}{2} - \frac{2c^3}{p}\right)\varphi(y) + \frac{2c^3}{p}\varphi(x).
\end{aligned}
\tag{55}
$$

Combining this and, e.g., [18, Lemma 2.2] (applied for every $t \in [0,T)$, $s \in [t,T]$, $x \in \mathbb{R}^d$, $\theta \in \Theta$ with $T \curvearrowleft T-t$, $O \curvearrowleft \mathbb{R}^d$, $V \curvearrowleft ([0,T-t] \times \mathbb{R}^d \ni (s,x) \mapsto \varphi(x) \in [0,\infty))$, $\alpha \curvearrowleft ([0,T-t] \ni s \mapsto 2c^3 \in [0,\infty))$, $\tau \curvearrowleft s-t$, $X \curvearrowleft (X^{\theta,t,x}_{t+r})_{r\in[0,T-t]}$ in the notation of [18, Lemma 2.2]) demonstrates that for all $\theta \in \Theta$, $x \in \mathbb{R}^d$, $t \in [0,T]$, $s \in [t,T]$ it holds that

$$
\mathbb{E}\big[\varphi(X^{\theta,t,x}_s)\big] \leqslant e^{2c^3(s-t)}\varphi(x). \tag{56}
$$

Itô's formula, (55), and the fact that $\varphi \geqslant 1$ imply that for all $x \in \mathbb{R}^d$, $t \in [0,T]$ it holds that

$$
\begin{aligned}
&\mathbb{E}[\varphi(\mathfrak{Y}^x_{\min\{\tau^x_n,t\}})] \\
&= \varphi(x) + \mathbb{E}\left[\int_0^{\min\{\tau^x_n,t\}}(\varphi'(\mathfrak{Y}^x_s))(\mu(x)) + \frac{1}{2}\sum_{k=1}^{m}(\varphi''(\mathfrak{Y}^x_s))(\sigma_k(x),\sigma_k(x))\,ds\right] \\
&\leqslant \varphi(x) + \mathbb{E}\left[\int_0^{\min\{\tau^x_n,t\}}\left(\frac{3c^3}{2} - \frac{2c^3}{p}\right)\varphi(\mathfrak{Y}^x_s) + \frac{2c^3}{p}\varphi(x)\,ds\right] \\
&\leqslant \varphi(x)\left(1 + \frac{2c^3t}{p}\right) + \left(\frac{3c^3}{2} - \frac{2c^3}{p}\right)\mathbb{E}\left[\int_0^t \varphi(\mathfrak{Y}^x_s)\mathbb{1}_{[0,\tau^x_n]}(s)\,ds\right] \\
&\leqslant \varphi(x)\left(1 + \frac{2c^3t}{p}\right) + \left(\frac{3c^3}{2} - \frac{2c^3}{p}\right)\int_0^t \mathbb{E}[\varphi(\mathfrak{Y}^x_{\min\{\tau^x_n,s\}})]\,ds.
\end{aligned}
\tag{57}
$$

Gronwall's inequality and the fact that for all $a \in \mathbb{R}$ it holds that $1 + a \leqslant e^a$ therefore assure that for all $x \in \mathbb{R}^d$, $t \in [0,T]$ it holds that

$$
\mathbb{E}[\varphi(\mathfrak{Y}^x_{\min\{\tau^x_n,t\}})] \leqslant \exp\!\left(\left[\frac{3c^3}{2} - \frac{2c^3}{p}\right]t\right)\left[1 + \frac{2c^3t}{p}\right]\varphi(x) \leqslant e^{2c^3t}\varphi(x). \tag{58}
$$

Fatou's lemma hence proves that for all $x \in \mathbb{R}^d$, $t \in [0,T]$ it holds that

$$
\mathbb{E}\big[\varphi(x + \mu(x)t + \sigma(x)W_t)\big] = \mathbb{E}[\varphi(\mathfrak{Y}^x_t)] \leqslant e^{2c^3t}\varphi(x). \tag{59}
$$

The tower property for conditional expectations, the fact that for all $t \in [0,T]$, $s \in [t,T]$, $\theta \in \Theta$ it holds that $W^\theta_s - W^\theta_t$ and $\mathbb{F}_t$ are independent, and the fact that for all $t \in [0,T]$, $s \in [t,T]$, $\theta \in \Theta$, $B \in \mathcal{B}(\mathbb{R}^d)$ it holds that $\mathbb{P}((W^\theta_s - W^\theta_t) \in B) = \mathbb{P}(W^\theta_{s-t} \in B)$ hence prove that for all $\theta \in \Theta$,

$x \in \mathbb{R}^d$, $t \in [0,T]$, $s \in [t,T]$ it holds that

$$\begin{aligned}
\mathbb{E}\big[\varphi(Y_s^{\theta,t,x})\big] &= \mathbb{E}\Big[\mathbb{E}\Big[\varphi\big(Y^{\theta,t,x}_{\max\{t,\lfloor s\rfloor_K\}} + \mu(Y^{\theta,t,x}_{\max\{t,\lfloor s\rfloor_K\}})(s-\max\{t,\lfloor s\rfloor_K\}) \\
&\qquad + \sigma(Y^{\theta,t,x}_{\max\{t,\lfloor s\rfloor_K\}})(W_s^\theta - W^\theta_{\max\{t,\lfloor s\rfloor_K\}})\big)\Big|\mathbb{F}_{\lfloor s\rfloor_K}\Big]\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[\varphi\big(z+\mu(z)(s-\max\{t,\lfloor s\rfloor_K\}) + \sigma(z)(W^\theta_{s-\max\{t,\lfloor s\rfloor_K\}})\big)\Big]\Big|_{z=Y^{\theta,x}_{t,\max\{t,\lfloor s\rfloor_K\}}}\Big] \\
&\leqslant e^{2c^3(s-\max\{t,\lfloor s\rfloor_K\})}\mathbb{E}\Big[\varphi\big(Y^{\theta,t,x}_{\max\{t,\lfloor s\rfloor_K\}}\big)\Big].
\end{aligned} \tag{60}$$

Induction and (49) hence show that for all $\theta \in \Theta$, $x \in \mathbb{R}^d$, $t \in [0,T]$, $s \in [t,T]$ it holds that $\mathbb{E}\big[\varphi(Y_s^{\theta,t,x})\big] \leqslant e^{2c^3(s-t)}\varphi(x)$. Jensen's inequality and (56) therefore prove that for all $q \in [0,p]$, $\theta \in \Theta$, $x \in \mathbb{R}^d$, $t \in [0,T]$, $s \in [t,T]$ it holds that

$$\begin{aligned}
&\max\Big\{\mathbb{E}\big[(\varphi(Y_s^{\theta,t,x}))^{\frac{q}{p}}\big], \mathbb{E}\big[(\varphi(X_s^{\theta,t,x}))^{\frac{q}{p}}\big]\Big\} \\
&\leqslant \max\Big\{\big(\mathbb{E}\big[\varphi(Y_s^{\theta,t,x})\big]\big)^{\frac{q}{p}}, \big(\mathbb{E}\big[\varphi(X_s^{\theta,t,x})\big]\big)^{\frac{q}{p}}\Big\} \leqslant e^{2qc^3(s-t)/p}(\varphi(x))^{\frac{q}{p}}.
\end{aligned} \tag{61}$$

Moreover, observe that the fact that $\mu$ is continuous, the fact that $\sigma$ is continuous, the fact that for all $\theta \in \Theta$, $\omega \in \Omega$ it holds that $[0,T] \ni t \mapsto W_t^\theta(\omega) \in \mathbb{R}^d$ is continuous, and Fubini's theorem imply that for all $\theta \in \Theta$ and all measurable $\eta\colon [0,T] \times \mathbb{R}^d \to [0,\infty)$ it holds that

$$\Delta \times \mathbb{R}^d \ni (t,s,x) \mapsto \mathbb{E}\big[\eta\big(s, Y_s^{\theta,t,x}\big)\big] \in [0,\infty] \tag{62}$$

is measurable. Furthermore, note that (45), (48), (55), and, e.g., [7, Lemma 3.7] (applied with $\mathcal{O} \curvearrowleft \mathbb{R}^d$, $V \curvearrowleft ([0,T] \times \mathbb{R}^d \ni (t,x) \mapsto e^{-2c^3t/p}\varphi(x) \in (0,\infty))$ in the notation of [7, Lemma 3.7]) imply that $\Delta \times \mathbb{R}^d \times \mathbb{R}^d \ni (t,s,x,y) \mapsto \big(s, X_s^{\theta,t,x}, X_s^{\theta,t,y}\big) \in \mathcal{L}^0(\Omega; \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d)$ is continuous. This and the dominated convergence theorem prove that for all $\theta \in \Theta$ and all bounded and continuous $\eta\colon [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty)$ it holds that $\Delta \times \mathbb{R}^d \times \mathbb{R}^d \ni (t,s,x,y) \mapsto \mathbb{E}\big[\eta\big(s, X_s^{\theta,t,x}, X_s^{\theta,t,y}\big)\big] \in [0,\infty]$ is continuous. Hence, we obtain that for all $\theta \in \Theta$ and all bounded and continuous $\eta\colon [0,T]\times\mathbb{R}^d\times\mathbb{R}^d \to [0,\infty)$ it holds that $\Delta\times\mathbb{R}^d\times\mathbb{R}^d \ni (t,s,x,y) \mapsto \mathbb{E}\big[\eta\big(s, X_s^{\theta,t,x}, X_s^{\theta,t,y}\big)\big] \in [0,\infty]$ is measurable. This implies that for all $\theta \in \Theta$ and all measurable $\eta\colon [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty)$ it holds that

$$\Delta \times \mathbb{R}^d \times \mathbb{R}^d \ni (t,s,x,y) \mapsto \mathbb{E}\big[\eta\big(s, X_s^{\theta,t,x}, X_s^{\theta,t,y}\big)\big] \in [0,\infty] \tag{63}$$

is measurable. Combining (62), (61), (46), (47), and [41, Proposition 2.2] (applied for every $k \in \{0,1\}$ with $L \curvearrowleft c$, $\mathcal{O} \curvearrowleft \mathbb{R}^d$, $(X^x_{t,s})_{(t,s,x)\in\Delta\times\mathbb{R}^d} \curvearrowleft (\mathfrak{X}_s^{k,t,x})_{(t,s,x)\in\Delta\times\mathbb{R}^d}$, $V \curvearrowleft ([0,T] \times \mathbb{R}^d \ni (s,x) \mapsto e^{2c^3\beta(T-s)/p}(\varphi(x))^{\beta/p} \in (0,\infty))$ in the notation of [41, Proposition 2.2]) hence establishes that

a) there exist unique measurable $u_k\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $k \in \{0,1\}$, which satisfy for all $k \in \{0,1\}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that $\sup_{s\in[0,T]}\sup_{x\in\mathbb{R}^d}\big[|u_k(s,x)|(\varphi(x))^{-\beta/p}\big] + \mathbb{E}\big[\big|g\big(\mathfrak{X}_T^{k,t,x}\big)\big| + \int_t^T \big|f\big(s, \mathfrak{X}_s^{k,t,x}, u_k\big(s, \mathfrak{X}_s^{k,t,x}\big)\big)\big|\,ds\big] < \infty$ and

$$u_k(t,x) = \mathbb{E}\bigg[g\big(\mathfrak{X}_T^{k,t,x}\big) + \int_t^T f\big(s, \mathfrak{X}_s^{k,t,x}, u_k(s, \mathfrak{X}_s^{k,t,x})\big)\,ds\bigg] \tag{64}$$

and
b) it holds for all $k \in \{0,1\}$ that

$$\sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d}\left[\frac{|u_k(t,x)|}{e^{2c^3\beta(T-t)/p}(\varphi(x))^{\beta/p}}\right] \leqslant \sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d}\left[\left[\frac{|g(x)|}{(\varphi(x))^{\beta/p}} + \frac{|Tf(t,x,0)|}{(\varphi(x))^{\beta/p}}\right]e^{cT}\right] \leqslant 2be^{cT}. \tag{65}$$

This proves (iii). Moreover, note that [41, Lemma 3.2] establishes (iv). Next observe that (54) and (61) demonstrate that for all $\theta \in \Theta$, $t \in [0, T]$, $r \in [t, T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\max\Big\{\mathbb{E}\Big[\big\|\mu(Y^{\theta,t,x}_{\max\{t,\lfloor r\rfloor_K\}})\big\|^2\Big], \mathbb{E}\Big[\|\sigma(Y^{\theta,t,x}_{\max\{t,\lfloor r\rfloor_K\}})\|\Big]\Big\} \\ &\leqslant c^2\mathbb{E}\Big[\big(\varphi(Y^{\theta,t,x}_{\max\{t,\lfloor r\rfloor_K\}})\big)^{\frac{2}{p}}\Big] \leqslant c^2 e^{4c^3(r-t)/p}(\varphi(x))^{\frac{2}{p}}. \end{aligned} \tag{66}$$

Furthermore, note that (49) demonstrates that for all $t \in [0, T]$, $r \in [t, T]$, $x \in \mathbb{R}^d$, $\theta \in \Theta$ it holds that $\sigma(\{Y^{\theta,t,x}_{\max\{t,\lfloor r\rfloor_K\}}\}) \subseteq \mathbb{F}_r$. Combining this and (66) with the fact that for all $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that $\mathbb{E}[\|\sigma(x)W_t\|^2] = \|\sigma(x)\|^2 t$ shows that for all $\theta \in \Theta$, $t \in [0, T]$, $r \in [t, T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\mathbb{E}\Big[\big\|\sigma(Y^{\theta,t,x}_{\max\{t,\lfloor r\rfloor_K\}})(W^\theta_r - W^\theta_{\max\{t,\lfloor r\rfloor_K\}})\big\|^2\Big] \\ &= \mathbb{E}\bigg[\mathbb{E}\Big[\big\|\sigma(y)(W^\theta_r - W^\theta_{\max\{t,\lfloor r\rfloor_K\}})\big\|^2\Big]\Big|_{y=Y^{\theta,x}_{t,\max\{t,\lfloor r\rfloor_K\}}}\bigg] \\ &= \mathbb{E}\Big[\|\sigma(Y^{\theta,x}_{t,\max\{t,\lfloor r\rfloor_K\}})\|^2(r - \max\{t, \lfloor r\rfloor_K\})\Big] \\ &\leqslant \mathbb{E}\Big[\|\sigma(Y^{\theta,x}_{t,\max\{t,\lfloor r\rfloor_K\}})\|^2\frac{T}{K}\Big] \leqslant c^2e^{4c^3(r-t)/p}(\varphi(x))^{\frac{2}{p}}\frac{T}{K}. \end{aligned} \tag{67}$$

This, (49), the triangle inequality, and (66) imply that for all $\theta \in \Theta$, $t \in [0, T]$, $r \in [t, T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\Big(\mathbb{E}\Big[\big\|Y^{\theta,t,x}_{\max\{t,\lfloor r\rfloor_K\}} - Y^{\theta,t,x}_r\big\|^2\Big]\Big)^{\frac{1}{2}} \\ &\leqslant \Big(\mathbb{E}\Big[\big\|\mu(Y^{\theta,t,x}_{\max\{t,\lfloor r\rfloor_K\}})\big\|^2\Big]\Big)^{\frac{1}{2}}(r - \max\{t, \lfloor r\rfloor_K\}) \\ &\quad + \Big(\mathbb{E}\Big[\big\|\sigma(Y^{\theta,x}_{t,\max\{t,\lfloor r\rfloor_K\}})(W^\theta_r - W^\theta_{\max\{t,\lfloor r\rfloor_K\}})\big\|^2\Big]\Big)^{\frac{1}{2}} \\ &\leqslant ce^{2c^3(r-t)/p}(\varphi(x))^{\frac{1}{p}}\Big(\frac{T}{K}\Big)^{\frac{1}{2}}|r-t|^{\frac{1}{2}} + ce^{2c^3(r-t)/p}(\varphi(x))^{\frac{1}{p}}\Big(\frac{T}{K}\Big)^{\frac{1}{2}} \\ &= c\big[|r-t|^{\frac{1}{2}} + 1\big]e^{2c^3(r-t)/p}(\varphi(x))^{\frac{1}{p}}\Big(\frac{T}{K}\Big)^{\frac{1}{2}}. \end{aligned} \tag{68}$$

Next, note that (48) and the fact that $c \geqslant 1$ assure that for all $z, y \in \mathbb{R}^d$ with $z \neq y$ it holds that

$$\frac{\langle z-y, \mu(z)-\mu(y)\rangle + \frac{1}{2}\|\sigma(z)-\sigma(y)\|^2}{\|z-y\|^2} + \frac{(\frac{2}{2}-1)\|(\sigma(z)-\sigma(y))^\mathsf{T}(z-y)\|^2}{\|z-y\|^4} \leqslant 2c^2. \tag{69}$$

This, [36, Theorem 1.2] (applied for every $\theta \in \Theta$, $t \in [0, T)$, $s \in (t, T]$, $x \in \mathbb{R}^d$ with $H \curvearrowleft \mathbb{R}^d$, $U \curvearrowleft \mathbb{R}^m$, $D \curvearrowleft \mathbb{R}^d$, $T \curvearrowleft (s-t)$, $(\mathbb{F}_r)_{r\in[0,T]} \curvearrowleft (\mathbb{F}_{r+t})_{r\in[0,s-t]}$, $(W_r)_{r\in[0,T]} \curvearrowleft (W^\theta_{t+r} - W^\theta_t)_{r\in[0,s-t]}$, $(X_r)_{r\in[0,T]} \curvearrowleft (X^{\theta,t,x}_{t+r})_{r\in[0,s-t]}$, $(Y_r)_{r\in[0,T]} \curvearrowleft (Y^{\theta,t,x}_{t+r})_{r\in[0,s-t]}$, $(a_r)_{r\in[0,T]} \curvearrowleft (\mu(Y^{\theta,x}_{t,\max\{t,\lfloor t+r\rfloor_K\}}))_{r\in[0,s-t]}$, $(b_r)_{r\in[0,T]} \curvearrowleft (\sigma(Y^{\theta,t,x}_{\max\{t,\lfloor t+r\rfloor_K\}}))_{r\in[0,s-t]}$, $\epsilon \curvearrowleft 1$, $p \curvearrowleft 2$, $\tau \curvearrowleft (\Omega \ni \omega \mapsto s-t \in [0, s-t])$, $\alpha \curvearrowleft 1$, $\beta \curvearrowleft 1$, $r \curvearrowleft 2$, $q \curvearrowleft \infty$ in the notation of [36, Theorem 1.2]), (48), (68), the fact that for all $t \in [0, \infty)$ it holds that $\sqrt{t}(\sqrt{t}+1) \leqslant e^t$, the fact that $1 \leqslant c$, and the fact that $p \geqslant 2$ imply that for all $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\Big(\mathbb{E}\Big[\big\|X_s^{\theta,t,x} - Y_s^{\theta,t,x}\big\|^2\Big]\Big)^{\frac{1}{2}} \\
&\leqslant \sup_{\substack{z,y\in\mathbb{R}^d,\\ z\neq y}} \exp\left(\int_t^s \left[\frac{\langle z-y,\mu(z)-\mu(y)\rangle + \frac{(2-1)(1+1)}{2}\|\sigma(z)-\sigma(y)\|^2}{\|z-y\|^2} + \frac{1-\frac{1}{2}}{1} + \frac{\frac{1}{2}-\frac{1}{2}}{1}\right]^+ dr\right) \\
&\quad \cdot \left[\left(\int_t^s \mathbb{E}\Big[\big\|\mu(Y^{\theta,t,x}_{\max\{t,\lfloor r\rfloor_K\}}) - \mu(Y_r^{\theta,t,x})\big\|^2\Big]\,dr\right)^{\frac{1}{2}}\right. \\
&\quad \left. + \sqrt{\frac{(2-1)(1+1)}{1}}\left(\int_t^s \mathbb{E}\Big[\big\|\sigma(Y^{\theta,t,x}_{\max\{\lfloor r\rfloor_K\}}) - \sigma(Y_r^{\theta,t,x})\big\|^2\Big]\,dr\right)^{\frac{1}{2}}\right] \\
&\leqslant e^{3c^2(s-t)}3c\left(|s-t| \sup_{r\in[t,s]} \mathbb{E}\Big[\big\|Y^{\theta,t,x}_{\max\{t,\lfloor r\rfloor_K\}} - Y_r^{\theta,t,x}\big\|^2\Big]\right)^{\frac{1}{2}} \\
&\leqslant e^{3c^2(s-t)}3c|s-t|^{\frac{1}{2}}c\big[|s-t|^{\frac{1}{2}}+1\big]e^{2c^3(s-t)/p}(\varphi(x))^{\frac{1}{p}}\left(\frac{T}{K}\right)^{\frac{1}{2}} \\
&\leqslant 3c^2e^{4c^2T}e^{2c^3(s-t)/p}(\varphi(x))^{\frac{1}{p}}\left(\frac{T}{K}\right)^{\frac{1}{2}}.
\end{aligned} \tag{70}$$

Next, observe that (i), (69), and [18, Corollary 2.26] (applied for every $t \in [0,T)$, $s \in (t,T]$ with $T \curvearrowleft s-t$, $O \curvearrowleft \mathbb{R}^d$, $(\mathcal{F}_r)_{r\in[0,T]} \curvearrowleft (\mathbb{F}_{t,t+r})_{r\in[0,s-t]}$, $(W_r)_{r\in[0,T]} \curvearrowleft (W^0_{t+r}-W^0_t)_{r\in[0,s-t]}$, $\alpha_0 \curvearrowleft 0$, $\alpha_1 \curvearrowleft 0$, $\beta_0 \curvearrowleft 0$, $\beta_1 \curvearrowleft 0$, $c \curvearrowleft 2c^2$, $r \curvearrowleft 2$, $p \curvearrowleft 2$, $q_0 \curvearrowleft \infty$, $q_1 \curvearrowleft \infty$, $U_0 \curvearrowleft (\mathbb{R}^d \ni x \mapsto 0 \in \mathbb{R})$, $U_1 \curvearrowleft (\mathbb{R}^d \ni x \mapsto 0 \in [0,\infty))$, $\overline{U} \curvearrowleft (\mathbb{R}^d \ni x \mapsto 0 \in \mathbb{R})$, $(X_r^x)_{r\in[0,T],x\in\mathbb{R}^d} \curvearrowleft (X^{0,t,x}_{t+r})_{r\in[0,s-t],x\in\mathbb{R}^d}$ in the notation of [18, Corollary 2.26]) demonstrate that for all $t \in [0,T)$, $s \in (t,T]$, $x,y \in \mathbb{R}^d$ it holds that $(\mathbb{E}[\|X_s^{0,t,x}-X_s^{0,t,y}\|^2])^{\frac{1}{2}} \leqslant e^{2c^2(s-t)}\|x-y\|$. This and (70) imply that for all $t \in [0,T]$, $s \in [t,T]$, $r \in [s,T]$, $x,y \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\Big[\mathbb{E}\Big[\big\|X_r^{0,s,\mathfrak{x}} - X_r^{0,s,\mathfrak{y}}\big\|^2\Big]\Big|_{(\mathfrak{x},\mathfrak{y})=(X_s^{0,t,x},Y_s^{0,t,x})}\Big]\right)^{\frac{1}{2}} \leqslant \left(\mathbb{E}\Big[\Big[e^{2c^2(r-s)}\big\|X_s^{0,t,x}-Y_s^{0,t,x}\big\|\Big]^2\Big]\right)^{\frac{1}{2}} \\
&\leqslant e^{2c^2(r-s)}3c^2e^{4c^2T}e^{2c^3(s-t)/p}(\varphi(x))^{\frac{1}{p}}\left(\frac{T}{K}\right)^{\frac{1}{2}} \leqslant 3c^2e^{4c^2T}\left(\frac{T}{K}\right)^{\frac{1}{2}}\big[e^{4c^3(T-t)/p}(\varphi(x))^{\frac{2}{p}}\big]^{\frac{1}{2}}.
\end{aligned} \tag{71}$$

Furthermore, note that (i) and Tonelli's theorem ensure that for all $t \in [0,T]$, $s \in [t,T]$, $r \in [s,T]$, $x,y \in \mathbb{R}^d$ and all measurable $h\colon \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty)$ it holds that $\mathbb{R}^d \times \mathbb{R}^d \ni (y_1,y_2) \mapsto \mathbb{E}\big[h\big(X_r^{0,s,y_1},X_r^{0,s,y_2}\big)\big] \in [0,\infty]$ is measurable. Moreover, observe that (i) assures that for all $t \in [0,T]$, $s \in [t,T]$, $r \in [s,T]$, $x,y \in \mathbb{R}^d$ it holds that $X_s^{0,t,x}$ and $X_r^{0,s,y}$ are independent. This and the disintegration theorem show that for all $t \in [0,T]$, $s \in [t,T]$, $r \in [s,T]$, $x,y \in \mathbb{R}^d$ and all measurable $h\colon \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty)$ it holds that $\mathbb{E}\big[\mathbb{E}\big[h\big(X_r^{0,s,\tilde{x}},X_r^{0,s,\tilde{y}}\big)\big]|_{\tilde{x}=X_s^{0,t,x},\tilde{y}=X_s^{0,t,y}}\big] = \mathbb{E}\big[h\big(X_r^{0,t,x},X_r^{0,t,y}\big)\big]$. Combining (i), (49), (61), (63), (47), (71), (64), (65), [41, Lemma 2.3] (applied with $L \curvearrowleft c$, $\rho \curvearrowleft 2c^3$, $\eta \curvearrowleft 1$, $\delta \curvearrowleft 3c^2e^{4c^2T}\left(\frac{T}{K}\right)^{\frac{1}{2}}$, $p \curvearrowleft p/\beta$, $q \curvearrowleft 2$, $(X_{t,s}^{x,1})_{t\in[0,T],s\in[t,T],x\in\mathbb{R}^d} \curvearrowleft (X_s^{0,t,x})_{t\in[0,T],s\in[t,T],x\in\mathbb{R}^d}$, $(X_{t,s}^{x,2})_{t\in[0,T],s\in[t,T],x\in\mathbb{R}^d} \curvearrowleft (Y_s^{0,t,x})_{t\in[0,T],s\in[t,T],x\in\mathbb{R}^d}$, $V \curvearrowleft b^{p/\beta}\varphi$, $\psi \curvearrowleft \big([0,T]\times\mathbb{R}^d \ni (t,x) \mapsto e^{4c^3(T-t)/p}(\varphi(x))^{\frac{2}{p}} \in (0,\infty)\big)$, $u_1 \curvearrowleft u_0$, $u_2 \curvearrowleft u_1$ in the notation of [41, Lemma 2.3]), the fact that $1+cT \leqslant e^{cT}$, the fact that $c \geqslant 1$, the fact that $\varphi \geqslant 1$, the fact that $p \geqslant 2$, and the fact that $p \geqslant 2\beta$ hence implies that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
&|u_0(t,x)-u_1(t,x)| \\
&\leqslant 4(1+cT)T^{-\frac{1}{2}}e^{cT+(2c^3\beta/p+c)T}(b^{p/\beta}\varphi(x))^{\frac{\beta}{p}}\big[e^{4c^3(T-t)/p}(\varphi(x))^{\frac{2}{p}}\big]^{\frac{1}{2}}3c^2e^{4c^2T}\left(\frac{T}{K}\right)^{\frac{1}{2}} \\
&\leqslant 4e^{cT}T^{-\frac{1}{2}}e^{cT+c^3T+cT}b(\varphi(x))^{\frac{\beta}{p}}e^{c^3T}(\varphi(x))^{\frac{1}{p}}3c^2e^{4c^2T}\left(\frac{T}{K}\right)^{\frac{1}{2}} \\
&\leqslant 12bc^2T^{-\frac{1}{2}}e^{9c^3T}(\varphi(x))^{\frac{\beta+1}{p}}\left(\frac{T}{K}\right)^{\frac{1}{2}}.
\end{aligned}
\tag{72}
$$

For the rest of this proof let $V\in C([0,T]\times\mathbb{R}^d,[1,\infty))$ satisfy for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$
V(t,x)=be^{\frac{2c^3\beta(T-t)}{p}}(\varphi(x))^{\frac{\beta}{p}}. \tag{73}
$$

Then (61) and the fact that $p\geqslant\mathfrak{p}\beta$ show for all $t\in[0,T]$, $s\in[t,T]$, $x\in\mathbb{R}^d$ that

$$
\big\|V(s,Y_s^{0,t,x})\big\|_{\mathfrak{p}}=be^{\frac{2c^3\beta(T-s)}{p}}\Big\|\varphi(Y_s^{0,t,x})^{\frac{\beta}{p}}\Big\|_{\mathfrak{p}}\leqslant be^{\frac{2c^3\beta(T-s)}{p}}e^{\frac{2c^3\beta(s-t)}{p}}(\varphi(x))^{\frac{\beta}{p}}=V(t,x). \tag{74}
$$

Then Theorem 2.3 (applied with $d\curvearrowleft d$, $p_v\curvearrowleft\mathfrak{p}$, $c\curvearrowleft c$, $T\curvearrowleft T$, $f\curvearrowleft f$, $g\curvearrowleft g$, $V\curvearrowleft V$, $\Theta\curvearrowleft\Theta$, $(\mathfrak{t}^\theta)_{\theta\in\Theta}\curvearrowleft(\mathfrak{t}^\theta)_{\theta\in\Theta}$, $X\curvearrowleft Y$, $(U^\theta_{n,m})_{\theta\in\Theta,n,m\in\mathbb{Z}}\curvearrowleft(U^\theta_{n,m})_{\theta\in\Theta,n,m\in\mathbb{Z}}$, $p\curvearrowleft\mathfrak{p}$, $q_1\curvearrowleft 1$ in the notation of Theorem 2.3), (50), and the independence and distributional assumptions show for all $m,n\in\mathbb{N}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$
\begin{aligned}
\big\|U^0_{n,m}(t,x)-u_1(t,x)\big\|_{\mathfrak{p}}&\leqslant 2(\mathfrak{p}-1)^{\frac{n}{2}}e^{5cTn}e^{m^{\mathfrak{p}/2}/\mathfrak{p}}m^{-n/2}V(t,x)\\
&\leqslant 2\mathfrak{p}^{\frac{n}{2}}e^{5cTn}e^{m^{\mathfrak{p}/2}/\mathfrak{p}}m^{-n/2}be^{\frac{2c^2\beta T}{p}}(\varphi(x))^{\frac{\beta}{p}}.
\end{aligned}
\tag{75}
$$

This, the triangle inequality, the fact that $p\geqslant 2\beta$, and the fact that $\varphi\geqslant 1$ show for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$
\begin{aligned}
&\big\|U^0_{n,m}(t,x)-u_0(t,x)\big\|_{\mathfrak{p}}\\
&\leqslant\big\|U^0_{n,m}(t,x)-u_1(t,x)\big\|_{\mathfrak{p}}+|u_0(t,x)-u_1(t,x)|\\
&\leqslant 2\mathfrak{p}^{\frac{n}{2}}e^{5cTn}e^{m^{\mathfrak{p}/2}/\mathfrak{p}}m^{-n/2}be^{\frac{2c^2\beta T}{p}}(\varphi(x))^{\frac{\beta}{p}}+12bc^2T^{-\frac{1}{2}}e^{9c^3T}(\varphi(x))^{\frac{\beta+1}{p}}\left(\frac{T}{K}\right)^{\frac{1}{2}}\\
&\leqslant 12bc^2e^{9c^3T}(\varphi(x))^{\frac{\beta+1}{p}}\left[2\mathfrak{p}^{\frac{n}{2}}e^{5cTn}e^{m^{\mathfrak{p}/2}/\mathfrak{p}}m^{-n/2}+\frac{1}{\sqrt{K}}\right]
\end{aligned}
\tag{76}
$$

This completes the proof of Lemma 2.4. □

### 2.3. **Complexity analysis for MLP approximations involving Euler-Maruyama approximations.**

We are now in a position to prove Theorem 1.1.

*Proof of Theorem 1.1.* For every $d\in\mathbb{N}$ let $\varphi_d\in C(\mathbb{R}^d,\mathbb{R})$ satisfy for all $x\in\mathbb{R}^d$ that

$$
\varphi_d(x)=2^{\mathfrak{p}}c^{\mathfrak{p}}d^{c\mathfrak{p}}(d^{2c}+\|x\|^2)^{\frac{\mathfrak{p}}{2}}. \tag{77}
$$

Then (154) shows for all $d\in\mathbb{N}$, $x\in\mathbb{R}^d$ that

$$
\begin{aligned}
&\max\{\|\mu^d(0)\|+c\|x\|,\|\sigma^d(0)\|+c\|x\|\}\\
&\leqslant cd^c+c\|x\|=c(d^c+\|x\|)\leqslant 2c(d^{2c}+\|x\|^2)^{\frac{1}{2}}\leqslant c(\varphi_d(x))^{\frac{1}{\mathfrak{p}}}.
\end{aligned}
\tag{78}
$$

Next, [47, Lemma 2.6] (applied for every $d\in\mathbb{N}$ with $d\curvearrowleft d$, $m\curvearrowleft d$, $a\curvearrowleft d^{2c}$, $c\curvearrowleft 0$, $p\curvearrowleft\mathfrak{p}/2$, $\mu\curvearrowleft 0$, $\sigma\curvearrowleft 0$, $\varphi\curvearrowleft\varphi_d/(2^{\mathfrak{p}}c^{\mathfrak{p}}d^{\mathfrak{p}c})$ in the notation of [47, Lemma 2.6]) and (77) show for all $x,z\in\mathbb{R}^d$ that

$$
\|(\varphi_d'(x))(z)\|\leqslant\mathfrak{p}(\varphi_d(x))^{1-\frac{1}{\mathfrak{p}}}\|z\|,\quad\|(\varphi_d''(x))(z,z)\|\leqslant\mathfrak{p}^2(\varphi_d(x))^{1-\frac{2}{\mathfrak{p}}}\|z\|^2. \tag{79}
$$

This, (78), and the fact that $\mathfrak{p}^2 \leqslant c$ show for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $x, z \in \mathbb{R}^d$ that

$$\max\left\{ \frac{|(\varphi_d'(x))(z)|}{(\varphi_d(x))^{\frac{\mathfrak{p}-1}{\mathfrak{p}}}\|z\|}, \frac{(\varphi_d''(x))(z,z)}{(\varphi_d(x))^{\frac{\mathfrak{p}-2}{\mathfrak{p}}}\|z\|^2}, \frac{c\|x\| + \|\mu^d(0)\|}{(\varphi_d(x))^{\frac{1}{\mathfrak{p}}}}, \frac{c\|x\| + \|\sigma^d(0)\|}{(\varphi_d(x))^{\frac{1}{\mathfrak{p}}}} \right\} \leqslant c. \tag{80}$$

Next, (1) and (77) show for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that

$$\max\{|Tf(0)|, |g^d(x)|\} \leqslant (\varphi_d(x))^{\frac{1}{\mathfrak{p}}}. \tag{81}$$

Furthermore, (2) show for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$ that

$$|g^d(x) - g^d(y)| \leqslant \frac{c}{\sqrt{T}}\|x-y\| \leqslant \frac{(\varphi_d(x))^{\frac{1}{\mathfrak{p}}} + (\varphi_d(y))^{\frac{1}{\mathfrak{p}}}}{2\sqrt{T}}\|x-y\| \leqslant \frac{(\varphi_d(x) + \varphi_d(y))^{\frac{1}{\mathfrak{p}}}}{\sqrt{T}}\|x-y\|. \tag{82}$$

This, Lemma 2.4 (applied for every $d, K \in \mathbb{N}$ with $T \curvearrowleft T$, $\mathfrak{p} \curvearrowleft \mathfrak{p}$, $\beta \curvearrowleft 1$, $b \curvearrowleft 1$, $c \curvearrowleft c$, $p \curvearrowleft \beta$, $\varphi \curvearrowleft \varphi_d$, $g \curvearrowleft g^d$, $f \curvearrowleft f$, $\mu \curvearrowleft \mu^d$, $\sigma \curvearrowleft \sigma^d$, $(\mathfrak{t}^\theta)_{\theta\in\Theta} \curvearrowleft (\mathfrak{t}^\theta)_{\theta\in\Theta}$, $(W^\theta)_{\theta\in\Theta} \curvearrowleft (W^{d,\theta})_{\theta\in\Theta}$, $(Y^{\theta,t,x})_{\theta\in\Theta,t\in[0,T],x\in\mathbb{R}^d} \curvearrowleft (Y^{d,\theta,K,t,x})_{\theta\in\Theta,t\in[0,T],x\in\mathbb{R}^d}$, $(U^\theta_{n,m})_{\theta\in\Theta,n,m\in Z} \curvearrowleft (U^{d,\theta,K}_{n,m})_{\theta\in\Theta,n,m\in Z}$ in the notation of Lemma 2.4) , (80), (81), (3), and (2) show that the following items are true.

(A) For every $t \in [0,T]$, $\theta \in \Theta$, $d \in \mathbb{N}$ there exists an up to indistinguishability unique continuous random field $X^{d,\theta,t,\cdot} = (X^{d,\theta,t,x}_s)_{s\in[t,T],x\in\mathbb{R}^d} \colon [t,T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}^d$ which satisfies that for all $x \in \mathbb{R}^d$ it holds that $(X^{d,\theta,t,x}_s)_{s\in[t,T]}$ is $(\mathbb{F}_s)_{s\in[t,T]}$-adapted and which satisfies that for all $s \in [t,T]$, $x \in \mathbb{R}^d$ it holds $\mathbb{P}$-a.s. that

$$X^{d,\theta,t,x}_s = x + \int_t^s \mu^d(X^{d,\theta,t,x}_r)\,dr + \int_t^s \sigma^d(X^{d,\theta,t,x}_r)\,dW^{d,\theta}_r. \tag{83}$$

(B) For every $d \in \mathbb{N}$ there exists a unique measurable $u^d \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ which satisfies for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\left(\sup_{s\in[0,T],y\in\mathbb{R}^d} [|u^d(s,y)|(\varphi_d(y))^{-1/\mathfrak{p}}]\right) + \int_t^T \mathbb{E}\big[|f(u^d(s,X^{d,0,t,x}_s))|\big]\,ds + \mathbb{E}\big[|g^d(X^{d,0,t,x}_T)|\big] < \infty \tag{84}$$

and

$$u^d(t,x) = \mathbb{E}\Big[g^d(X^{d,0,t,x}_T)\Big] + \int_t^T \mathbb{E}\big[f(u^d(s,X^{0,t,x}_s))\big]\,ds. \tag{85}$$

(C) For all $d, K, m \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$, $n \in \mathbb{N}_0$ we have that $U^{d,\theta,K}_{n,m}(t,x)$ is measurable.

(D) For all $t \in [0,T]$, $x \in \mathbb{R}^d$, $m, n \in \mathbb{N}$ we have that

$$\left\|U^{d,0,K}_{n,m}(t,x) - u^d(t,x)\right\|_{\mathfrak{p}} \leqslant 12c^2e^{9c^3T}(\varphi_d(x))^{\frac{2}{\mathfrak{p}}}\left[2\mathfrak{p}^{\frac{n}{2}}e^{5cTn}e^{m^{\mathfrak{p}/2}/\mathfrak{p}}m^{-n/2} + \frac{1}{\sqrt{K}}\right]. \tag{86}$$

Next, the triangle inequality, (1), (2), and the fact that $\forall\, x \in \mathbb{R}^d \colon (1+x)^2 \leqslant 1(1+x^2)$ show for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} \langle x, \mu^d(x)\rangle &\leqslant \|x\|\left(\|\mu^d(x) - \mu^d(0)\| + \|\mu^d(0)\|\right) \\ &\leqslant \|x\|(c\|x\| + cd^c) \\ &\leqslant (1+\|x\|)^2cd^c \\ &\leqslant 2cd^c(1+\|x\|^2). \end{aligned} \tag{87}$$

Furthermore, the Cauchy-Schwarz inequality implies for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$ that

$$\|\sigma^d(x)y\|^2 = \sum_{i=1}^d \left|\sum_{j=1}^d (\sigma^d)_{ij}(x)y_j\right|^2 \leqslant \sum_{i=1}^d \left(\sum_{j=1}^d |(\sigma^d)_{ij}(x)|^2\right)\left(\sum_{j=1}^d |y_j|^2\right) \leqslant \|\sigma(x)\|^2\|y\|^2. \tag{88}$$

This and (2) show for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$ that

$$\|\sigma^d(x)y\| \leqslant \|\sigma^d(x)\|\|y\| \leqslant (\|\sigma^d(x) - \sigma^d(0)\| + \|\sigma^d(0)\|)\|y\| \leqslant (c\|x\| + cd^c)\|y\| \leqslant cd^c(1+\|x\|)\|y\|. \tag{89}$$

This, (87), [10, Theorem 1.1] (applied for every $d \in \mathbb{N}$ with $d \curvearrowleft d$, $L \curvearrowleft 2cd^c$, $T \curvearrowleft T$, $\mu \curvearrowleft \mu^d$, $\sigma \curvearrowleft \sigma^d$, $f \curvearrowleft (\mathbb{R}^d \times \mathbb{R} \ni (x,w) \mapsto f^d(w) \in \mathbb{R})$, $g \curvearrowleft g^d$, $W \curvearrowleft W^{d,\theta}$ in the notation of [10, Theorem 1.1]), (2), the fact that for every $d \in \mathbb{N}$, $g^d$ is polynomially growing (cf. (1)), and the fact that for every $d \in \mathbb{N}$, $u^d$ is polynomially growing (cf. (77) and (84)) show for every $d \in \mathbb{N}$ that $u^d$ is the unique at most polynomially growing viscosity solution of

$$\frac{\partial u^d}{\partial t}(t,x) + \frac{1}{2}\operatorname{trace}(\sigma^d(\sigma^d(x))^\top(\operatorname{Hess}_x u^d(t,x))) + \langle \mu^d(x), (\nabla_x u^d)(t,x)\rangle + f(u^d(t,x)) = 0 \quad (90)$$

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

Next, (86) show that there exists $\kappa \in (0,\infty)$ such that for all $d,m,n \in \mathbb{N}$ we have that

$$\begin{aligned}\sup_{t\in[0,T],x\in[0,\mathbf{k}]^d} \left\| U^{d,0,m^n}_{n,m}(t,x) - u^d(t,x)\right\|_{\mathfrak{p}} &\leqslant \sup_{x\in[0,\mathbf{k}]^d} \left(12c^2e^{9c^3T}(\varphi_d(x))^{\frac{2}{\mathfrak{p}}}3\mathfrak{p}^{\frac{n}{2}}e^{5cTn}e^{m^{\mathfrak{p}/2}/\mathfrak{p}}m^{-n/2}\right)\\ &\leqslant \kappa d^\kappa \mathfrak{p}^{\frac{n}{2}}e^{5cTn}e^{m^{\mathfrak{p}/2}/\mathfrak{p}}m^{-n/2}.\end{aligned} \quad (91)$$

For every $\varepsilon \in (0,1)$ let

$$N_\varepsilon = \inf\left\{n \in \mathbb{N} \colon \left(\frac{\mathfrak{p}^{\frac{1}{2}}e^{5cT}\exp(\frac{(M_n)^{\mathfrak{p}/2}}{n})}{(M_n)^{\frac{1}{2}}}\right)^n \leqslant \varepsilon\right\}. \quad (92)$$

For every $\epsilon \in (0,1)$, $d \in \mathbb{N}$ let

$$\varepsilon(d,\epsilon) = \frac{\epsilon}{\kappa d^\kappa}, \quad n(d,\epsilon) = N_{\varepsilon(d,\epsilon)}. \quad (93)$$

For every $\delta \in (0,1)$ let

$$C_\delta = \sup_{\varepsilon\in(0,1)} \left[\varepsilon^{4+\delta}(3M_{N_\varepsilon})^{2N_\varepsilon}\right]. \quad (94)$$

Next, [43, Lemma 4.5] and the definition of $(M_n)_{n\in\mathbb{N}}$ show that $\liminf_{j\to\infty} M_j = \infty$, $\limsup_{j\to\infty}\frac{(M_j)^{\mathfrak{p}/2}}{j} < \infty$, and $\sup_{k\in\mathbb{N}}\frac{M_{k+1}}{M_k} < \infty$. Then (94) and [2, Lemma 5.1] (applied with $L \curvearrowleft 1$, $T \curvearrowleft \mathfrak{p}^{\frac{1}{2}}e^{5cT} - 1$, $(m_k)_{k\in\mathbb{N}} \curvearrowleft (M_k)_{k\in\mathbb{N}}$ in the notation of [2, Lemma 5.1]) show for all $\delta, \varepsilon \in (0,1)$ that $N_\varepsilon < \infty$ and $C_\delta < \infty$. Next, (91) and (92) show for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ that

$$\begin{aligned}&\sup_{t\in[0,T],x\in[0,\mathbf{k}]^d}\left\|U^{d,0,(M_{n(d,\varepsilon)})^{n(d,\epsilon)}}_{n(d,\varepsilon),M_{n(d,\varepsilon)}}(t,x) - u^d(t,x)\right\|_{\mathfrak{p}} \leqslant \kappa d^\kappa\mathfrak{p}^{\frac{n}{2}}e^{5cTN_{\varepsilon(d,\epsilon)}}e^{(M_{N_{\varepsilon(d,\epsilon)}})^{\mathfrak{p}/2}/\mathfrak{p}}(M_{N_{\varepsilon(d,\epsilon)}})^{-N_{\varepsilon(d,\epsilon)}/2}\\ &\leqslant \kappa d^\kappa \varepsilon(d,\epsilon) = \epsilon.\end{aligned} \quad (95)$$

Next, (6) show for all $d,K,n,m \in \mathbb{N}$ that

$$C^{d,K}_{0,m} = 0, \quad C^{d,K}_{n,m} \leqslant 2cd^cKm^n + \sum_{\ell=0}^{n-1} m^{n-\ell}\left(3cd^cK + C^{d,K}_{\ell,m} + C^{d,K}_{\ell-1,m}\right). \quad (96)$$

This and [8, Lemma 3.14] show for all $d,K,m,n \in \mathbb{N}$ that

$$C^{d,K}_{n,m} \leqslant 3cd^cK(3m)^n. \quad (97)$$

This, (93), and (94) show that for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ that

$$\begin{aligned}C^{d,(M_{n(d,\epsilon)})^{n(d,\epsilon)}}_{n(d,\epsilon),M_{n(d,\epsilon)}} &= C^{d,(M_{N_{\varepsilon(d,\epsilon)}})^{N_{\varepsilon(d,\epsilon)}}}_{N_{\varepsilon(d,\epsilon)},M_{N_{\varepsilon(d,\epsilon)}}} \leqslant cd^c(3M_{N_{\varepsilon(d,\epsilon)}})^{2N_{\varepsilon(d,\epsilon)}}\\ &\leqslant cd^cC_\delta(\varepsilon(d,\varepsilon))^{-(4+\delta)} = cd^cC_\delta\left(\frac{\epsilon}{\kappa d^\kappa}\right)^{-(4+\delta)} = cd^c(\kappa d^\kappa)^{4+\delta}\epsilon^{-(4+\delta)}.\end{aligned} \quad (98)$$

This, (95), (93), the fact that $\forall\, \varepsilon \in (0,1) \colon N_\varepsilon < \infty$, and the fact that $\forall\, \delta \in (0,1) \colon C_\delta < \infty$ complete the proof of Theorem 1.1. □

## 3. DNNs

Our main goal in this section is to prove Lemma 3.15, which states that the MLP approximations defined by (120) can be represented by DNNs. Furthermore, in Lemma 3.15 we also bound the length and the supremum norm of the vectors of their layer dimensions. Note that in this paper we consider different types of activation functions than ReLU.

### 3.1. DNN representation of the one-dimensional identity.

In Lemma 3.1 and 3.2 we prove that the identity in $\mathbb{R}$ can be represented by a DNN. Here, we consider ReLU, leaky ReLU, and softplus activation function. Later in Setting 3.13 as well as in the setting of Theorem 4.1 we consider this as an assumption (see (103) and (151)).

**Lemma 3.1.** *Assume Setting 1.2. Let $\alpha \in [0,\infty)$ satisfy for all $x \in \mathbb{R}^d$ that $a(x) = \max\{x, \alpha x\}$. Then $\mathrm{Id}_{\mathbb{R}} \in \mathcal{R}(\{\Phi \in \mathbf{N} \colon \mathcal{D}(\Phi) = (1,2,1)\})$.*

*Proof of Lemma 3.1.* See [1, Lemma 3.5]. □

**Lemma 3.2.** *Assume Setting 1.2 and assume for all $x \in \mathbb{R}$ that $a(x) = \ln(1+e^x)$. Then $\mathrm{Id}_{\mathbb{R}} \in \mathcal{R}(\{\Phi \in \mathbf{N} \colon \mathcal{D}(\Phi) = (1,2,1)\})$.*

*Proof of Lemma 3.2.* See [1, Lemma 3.8]. □

### 3.2. DNN representation of the $d$-dimensional identity.

In Lemma 3.3 below we prove that if the identity in $\mathbb{R}$ can be represented by a DNN then the identity in $\mathbb{R}^d$ can also be represented by a DNN.

**Lemma 3.3.** *Assume Setting 1.2. Let $d, \mathfrak{d} \in \mathbb{N}$, $\phi \in \mathbf{N}$ satisfy for all $x \in \mathbb{R}$ that $\mathcal{D}(\phi) = (1, \mathfrak{d}, 1)$ and $(\mathcal{R}(\phi))(x) = x$. Then there exists $\Phi \in \mathbf{N}$ which satisfies for all $x \in \mathbb{R}^d$ that $\mathcal{D}(\Phi) = (d, \mathfrak{d}d, d) \in \mathbb{R}^3$ and $(\mathcal{R}(\Phi))(x) = x$.*

*Proof of Lemma 3.3.* Let $W_1 \in \mathbb{R}^{\mathfrak{d}\times 1}$, $B_1 \in \mathbb{R}^{\mathfrak{d}}$, $W_2 \in \mathbb{R}^{1\times\mathfrak{d}}$, $B_2 \in \mathbb{R}$ satisfy $\phi = ((W_1, B_1), (W_2, B_2))$. Then by definition for all $x^0 \in \mathbb{R}$, $x^1 \in \mathbb{R}^{\mathfrak{d}}$ with $x^1 = \mathbf{A}_{\mathfrak{d}}(W_1 x^0 + B_1)$ we have that $(\mathcal{R}(\phi))(x^0) = W_2 x^1 + B_2$, i.e.,

$$(\mathcal{R}(\phi))(x^0) = W_2 \mathbf{A}_{\mathfrak{d}}(W_1 x^0 + B_1) + B_2. \tag{99}$$

Now, let $\Phi \in \mathbf{N}$, $\widehat{W}_1 \in \mathbb{R}^{\mathfrak{d}d\times d}$, $\widehat{B}_1 \in \mathbb{R}^{\mathfrak{d}d}$, $\widehat{W}_2 \in \mathbb{R}^{d\times\mathfrak{d}d}$, $\widehat{B}_2 \in \mathbb{R}^d$ satisfy for all $n \in [1,H] \cap \mathbb{Z}$ that $\Phi = ((\widehat{W}_1, \widehat{B}_1), (\widehat{W}_2, \widehat{B}_2))$,

$$\widehat{W}_1 = \begin{pmatrix} W_1 & & \\ & \ddots & \\ & & W_1 \end{pmatrix}, \quad \widehat{B}_1 = \begin{pmatrix} B_1 \\ \vdots \\ B_1 \end{pmatrix}, \quad \widehat{W}_2 = \begin{pmatrix} W_2 & & \\ & \ddots & \\ & & W_2 \end{pmatrix}, \quad \widehat{B}_2 = \begin{pmatrix} B_2 \\ \vdots \\ B_2 \end{pmatrix}. \tag{100}$$

Then $\mathcal{D}(\Phi) = (d, \mathfrak{d}d, d) \in \mathbb{R}^3$. Furthermore, (99) shows that for all $x^0 = (x^0_1, \ldots, x^0_d)^\top \in \mathbb{R}^d$, $x^1 \in \mathbb{R}^{\mathfrak{d}d}$ satisfying that $x^1 = \mathbf{A}_{\mathfrak{d}d}(\widehat{W}_n x^0 + \widehat{B}_n)$ we have that

$$x^1 = \mathbf{A}_{\mathfrak{d}d}(\widehat{W}^1 x^0 + \widehat{B}_1) = \mathbf{A}_{\mathfrak{d}d}\begin{pmatrix} W_1 x^0_1 + B_1 \\ \vdots \\ W_1 x^0_d + B_1 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{\mathfrak{d}}(W_1 x^0_1 + B_1) \\ \vdots \\ \mathbf{A}_{\mathfrak{d}}(W_1 x^0_d + B_1) \end{pmatrix} \tag{101}$$

and

$$(\mathcal{R}(\Phi))(x^0) = W_2 x^1 + B_2 = \begin{pmatrix} W_2 \mathbf{A}_{\mathfrak{d}}(W_1 x^0_1 + B_1) + B_2 \\ \vdots \\ W_2 \mathbf{A}_{\mathfrak{d}}(W_1 x^0_d + B_1) + B_2 \end{pmatrix} = \begin{pmatrix} x^0_1 \\ \vdots \\ x^0_d \end{pmatrix} = x^0. \tag{102}$$

This completes the proof of Lemma 3.3. □

3.3. **Approximation of one-dimensional Lipschitz functions by DNNs.** In Lemmas 3.4 and 3.5 we prove that one dimensional Lipschitz functions can be well approximated by DNNs. Later in Theorem 4.1 we consider this fact as an assumption.

**Lemma 3.4.** *Assume Setting 1.2. Let $\alpha \in [0,\infty)\backslash\{1\}$ and assume for all $x \in \mathbb{R}$ that $a(x) = \max\{x, \alpha x\}$. Let $f \in C(\mathbb{R}, \mathbb{R})$, $L \in \mathbb{R}$, $q \in (1,\infty)$ satisfy for all $x, y \in \mathbb{R}$ that $|f(x)-f(y)| \leqslant L|x-y|$. Then there exist $c \in (0,\infty)$, $(f_\varepsilon)_{\varepsilon\in(0,1)} \subseteq C(\mathbb{R}, \mathbb{R})$ such that for all $\varepsilon \in (0,1)$, $x, y \in \mathbb{R}$ we have that $|f_\varepsilon(x) - f_\varepsilon(y)| \leqslant L|x - y|$, $|f_\varepsilon(x) - f(x)| \leqslant \varepsilon(1 + |x|^q)$, and $f_\varepsilon \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon \dim(\mathcal{D}(\Phi)) = 3, \|\!|\mathcal{D}(\Phi)|\!\| \leqslant c\varepsilon^{-c}\})$.*

*Proof of Lemma 3.4.* See [1, Corollary 4.13]. □

**Lemma 3.5.** *Assume Setting 1.2. Assume for all $x \in \mathbb{R}$ that $a(x) = \ln(1 + e^x)$. Let $f \in C(\mathbb{R}, \mathbb{R})$, $L \in \mathbb{R}$, $q \in (1,\infty)$ satisfy for all $x, y \in \mathbb{R}$ that $|f(x) - f(y)| \leqslant L|x - y|$. Then there exist $c \in (0,\infty)$, $(f_\varepsilon)_{\varepsilon\in(0,1)} \subseteq C(\mathbb{R}, \mathbb{R})$ such that for all $\varepsilon \in (0,1)$, $x, y \in \mathbb{R}$ we have that $|f_\varepsilon(x) - f_\varepsilon(y)| \leqslant L|x - y|$, $|f_\varepsilon(x) - f(x)| \leqslant \varepsilon(1 + |x|^q)$, and $f_\varepsilon \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon \dim(\mathcal{D}(\Phi)) = 3, \|\!|\mathcal{D}(\Phi)|\!\| \leqslant c\varepsilon^{-c}\})$.*

*Proof of Lemma 3.5.* See [1, Corollary 4.14]. □

3.4. **Properties of operations associated to DNNs.**

**Setting 3.6.** *Assume Setting 1.2. Let $\odot\colon \mathbf{D} \times \mathbf{D} \to \mathbf{D}$ satisfy for all $H_1, H_2 \in \mathbb{N}$, $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_{H_1}, \alpha_{H_1+1}) \in \mathbb{N}^{H_1+2}$, $\beta = (\beta_0, \beta_1, \ldots, \beta_{H_2}, \beta_{H_2+1}) \in \mathbb{N}^{H_2+2}$ that $\alpha \odot \beta = (\alpha_0, \ldots, \alpha_{H_1}, \beta_1, \ldots, \beta_{H_2+1})$. Let $\boxplus\colon \mathbf{D}\times\mathbf{D} \to \mathbf{D}$ satisfy for all $H \in \mathbb{N}$, $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_H, \alpha_{H+1}) \in \mathbb{N}^{H+2}$, $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_H, \beta_{H+1}) \in \mathbb{N}^{H+2}$ that $\alpha\boxplus\beta = (\alpha_0, \alpha_1+\beta_1, \ldots, \alpha_H+\beta_H, \beta_{H+1}) \in \mathbb{N}^{H+2}$.*

**Lemma 3.7.** *Assume Setting 3.6 and let $\alpha, \beta, \gamma \in \mathbf{D}$. Then $(\alpha \odot \beta) \odot \gamma = \alpha \odot (\beta \odot \gamma)$.*

*Proof of Lemma 3.7.* Straightforward. □

**Lemma 3.8.** *Assume Setting 3.6, let $H, k, l \in \mathbb{N}$, and let $\alpha, \beta, \gamma \in \big(\{k\} \times \mathbb{N}^H \times \{l\}\big)$. Then*

*(i) we have that $\alpha \boxplus \beta \in \big(\{k\} \times \mathbb{N}^H \times \{l\}\big)$,*
*(ii) we have that $\beta \boxplus \gamma \in \big(\{k\} \times \mathbb{N}^H \times \{l\}\big)$, and*
*(iii) we have that $(\alpha \boxplus \beta) \boxplus \gamma = \alpha \boxplus (\beta \boxplus \gamma)$.*

*Proof of Lemma 3.8.* Straightforward. We could use the proof of [42, Lemma 3.4]. □

Lemma 3.9 below is later important to estimate the maximum norm of the vector of layer dimensions of DNNs.

**Lemma 3.9** (Triangle inequality)**.** *Assume Setting 3.6, let $H, k, l \in \mathbb{N}$, and let $\alpha, \beta \in \{k\}\times\mathbb{N}^H\times\{l\}$. Then we have that $\|\!|\alpha \boxplus \beta|\!\| \leqslant \|\!|\alpha|\!\| + \|\!|\beta|\!\|$.*

*Proof of Lemma 3.9.* We can use the proof of [42, Lemma 3.5]. □

Lemma 3.10 below show that affine transformations of DNNs can be represented by DNNs with the same vector of layer dimensions.

**Lemma 3.10** (DNNs for affine transformations)**.** *Assume Setting 1.2 and let $d, m \in \mathbb{N}$, $\lambda \in \mathbb{R}$, $b \in \mathbb{R}^d$, $a \in \mathbb{R}^m$, $\Psi \in \mathbf{N}$ satisfy that $\mathcal{R}(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^m)$. Then we have that $\lambda\left((\mathcal{R}(\Psi))(\cdot + b) + a\right) \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = \mathcal{D}(\Psi)\})$.*

*Proof of Lemma 3.10.* We can use the proof of [42, Lemma 3.7], which also works for other activation functions than ReLU. □

Lemma 3.11 below shows that compositions of DNN functions can be represented by DNNs.

**Lemma 3.11** (Composition of functions generated by DNNs)**.** *Assume Setting 3.6 and let $d_1, d_2, d_3 \in \mathbb{N}$, $f \in C(\mathbb{R}^{d_2}, \mathbb{R}^{d_3})$, $g \in C(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$, $\alpha, \beta \in \mathbf{D}$ satisfy that $f \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = \alpha\})$ and $g \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = \beta\})$. Then we have that $(f \circ g) \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = \alpha \odot \beta\})$.*

*Proof of Lemma 3.11.* See [51, Proposition 2.1.2], which especially works for general activation functions. □

Lemma 3.12 below shows that sums of DNNs of the same length can be represented by DNNs. In order to represent sums of DNNs with different lengths we note that the identity function can be represented as DNNs. We then take the composition of a DNN function with the identity to change the its length. This is one of the main techniques in the proof of Lemmas 3.14 and 3.15.

**Lemma 3.12** (Sum of DNNs of the same length)**.** *Assume Setting 3.6 and let* $p, q, M, H \in \mathbb{N}$, $\alpha_1, \alpha_2, \ldots, \alpha_M \in \mathbb{R}$, $k_i \in \mathbf{D}$, $g_i \in C(\mathbb{R}^p, \mathbb{R}^q)$, $i \in [1, M] \cap \mathbb{N}$, *satisfy for all* $i \in [1, M] \cap \mathbb{N}$ *that* $\dim(k_i) = H + 2$ *and* $g_i \in \mathcal{R}(\{\Phi \in \mathbf{N} \colon \mathcal{D}(\Phi) = k_i\})$. *Then we have that* $\sum_{i=1}^M \alpha_i g_i \in \mathcal{R}\left(\left\{\Phi \in \mathbf{N} \colon \mathcal{D}(\Phi) = \boxplus_{i=1}^M k_i\right\}\right)$.

*Proof of Lemma 3.12.* We can use the proof of [42, Lemma 3.9], which can be extended to other activation functions than ReLU. See also [51, Lemma 2.4.11]. □

3.5. **DNN representation of our Euler-Maruyama approximations.** In Lemma 3.14 below we prove that Euler-Maruyama approximations can be represented by DNNs if their coefficients are represented by DNNs and if the identity in $\mathbb{R}$ can be represented by a DNN (see (103)).

**Setting 3.13.** *Assume Setting 1.2. Let* $\mathfrak{d} \in \mathbb{N}$, $\mathfrak{n}_{1,\mathfrak{d}} = (1, \mathfrak{d}, 1) \in \mathbf{D}$ *satisfy that*

$$\mathrm{Id}_{\mathbb{R}} \in \mathcal{R}(\{\Phi \in \mathbb{N} \colon \mathcal{D}(\Phi) = \mathfrak{n}_{1,\mathfrak{d}}\}). \tag{103}$$

*Let* $T \in (0, \infty)$, $K \in \mathbb{N}$. *Let* $\lfloor \cdot \rfloor_K \colon \mathbb{R} \to \mathbb{R}$ *satisfy for all* $t \in \mathbb{R}$ *that* $\lfloor t \rfloor_K = \max(\{0, \frac{T}{K}, \frac{2T}{K}, \ldots, T\} \cap ((-\infty, t) \cup \{0\}))$. *For every* $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $v \in \mathbb{R}^d$ *let* $\mu_\varepsilon^d \in C(\mathbb{R}^d, \mathbb{R}^d)$, $\sigma_\varepsilon^d \in C(\mathbb{R}^d, \mathbb{R}^{d \times d})$, $\Phi_{\mu_\varepsilon^d}, \Phi_{\sigma_\varepsilon^d, v} \in \mathbf{N}$ *satisfy that* $\mu_\varepsilon^d = \mathcal{R}(\Phi_{\mu_\varepsilon^d})$, $\sigma_\varepsilon^d(\cdot)v = \mathcal{R}(\Phi_{\sigma_\varepsilon^d, v})$. *Assume for all* $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $v \in \mathbb{R}^d$ *that* $\mathcal{D}(\Phi_{\sigma_\varepsilon^d, v}) = \mathcal{D}(\Phi_{\sigma_\varepsilon^d, 0})$. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. For every* $d \in \mathbb{N}$ *let* $W^{d,\theta} = (W_t^{d,\theta})_{t \in [0,T]} \colon [0,T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, *be independent standard Brownian motions. For every* $d \in \mathbb{N}$, $\theta \in \Theta$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1)$, $t \in [0,T)$ *let* $(X_s^{d,\theta,K,\varepsilon,t,x})_{s \in [t,T]}$ *satisfy that* $X_t^{d,\theta,K,\varepsilon,t,x} = x$ *and*

$$X_s^{d,\theta,K,\varepsilon,t,x} = x + \int_t^s \mu_\varepsilon^d(X^{d,\theta,K,\varepsilon,t,x}_{\max\{t,\lfloor u \rfloor_K\}})du + \int_t^s \sigma_\varepsilon^d(X^{d,\theta,K,\varepsilon,t,x}_{\max\{t,\lfloor u \rfloor_K\}})dW_u^{d,\theta}. \tag{104}$$

**Lemma 3.14.** *Assume Setting 3.13. Let* $\omega \in \Omega$. *Then there exists* $(\mathcal{X}_s^{d,\theta,K,\varepsilon,t})_{d \in \mathbb{N}, \theta \in \Theta, \varepsilon \in (0,1), t \in [0,T), s \in (t,T]} \subseteq \mathbf{N}$ *such that the following items are true.*

*(i) For all* $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t \in [0,T)$, $s \in (t,T]$, $x \in \mathbb{R}^d$ *we have that* $\mathcal{R}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t}) \in C(\mathbb{R}^d, \mathbb{R}^d)$ *and* $(\mathcal{R}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t}))(x) = X_s^{d,\theta,K,\varepsilon,t,x}(\omega)$.

*(ii) For all* $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t_1 \in [0,T)$, $s_1 \in (t_1,T]$, $t_2 \in [0,T)$, $s_2 \in (t_2,T]$, $x \in \mathbb{R}^d$ *we have that* $\mathcal{D}(\mathcal{X}_{s_1}^{d,\theta_1,K,\varepsilon,t_1}) = \mathcal{D}(\mathcal{X}_{s_2}^{d,\theta_2,K,\varepsilon,t_2})$.

*(iii) For all* $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t \in [0,T)$, $s \in (t,T]$ *we have that* $\dim(\mathcal{D}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t})) = K(\max\{\dim(\mathcal{D}(\Phi_{\mu_\varepsilon^d})), \dim(\mathcal{D}(\Phi_{\sigma_\varepsilon^d,0}))\} - 2) + 2$.

*(iv) For all* $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t \in [0,T)$, $s \in (t,T]$ *we have that* $\left\|\!\left|\mathcal{D}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t})\right|\!\right\| \leq 3\max\{d\mathfrak{d}, \left\|\!\left|\mathcal{D}(\Phi_{\mu_\varepsilon^d})\right|\!\right\|, \left\|\!\left|\mathcal{D}(\Phi_{\sigma_\varepsilon^d,0})\right|\!\right\|\}$.

*Proof of Lemma 3.14.* Throughout this proof let the notation in Setting 3.6 be given. Moreover, for every $d, n \in \mathbb{N}$ let $\mathfrak{n}_{d,\mathfrak{d}} = (d, \mathfrak{d}d, d) \in \mathbf{D}$ and $\mathfrak{n}_{d,\mathfrak{d}}^{\odot n} = \mathfrak{n}_{d,\mathfrak{d}} \odot \ldots \odot \mathfrak{n}_{d,\mathfrak{d}}$ ($n$ times). Lemmas 3.3, 3.11, and a simple induction argument show for all $d, n \in \mathbb{N}$ that

$$\mathrm{Id}_{\mathbb{R}^d} \in \mathcal{R}(\{\Phi \in \mathbb{N} \colon \mathcal{D}(\Phi) = \mathfrak{n}_{d,\mathfrak{d}}^{\odot n}\}), \quad \mathfrak{n}_{d,\mathfrak{d}}^{\odot n} = (d, \mathfrak{d}d, \ldots, \mathfrak{d}d, d) \in \mathbb{R}^{n+2}. \tag{105}$$

This, Lemma 3.11, and the definition of $\odot$ show for all $d, n \in \mathbb{N}$, $\varepsilon \in (0,1)$ that

$$\mu_\varepsilon^d \in \mathcal{R}(\{\Phi \in \mathbb{N} \colon \mathcal{D}(\Phi) = \mathcal{D}(\Phi_{\mu_\varepsilon^d}) \odot \mathfrak{n}_{d,\mathfrak{d}}^{\odot n}\}) \tag{106}$$

and

$$\begin{aligned}\dim(\mathcal{D}(\Phi_{\mu_\varepsilon^d}) \odot \mathfrak{n}_{d,\mathfrak{d}}^{\odot n}) &= \dim(\mathcal{D}(\Phi_{\mu_\varepsilon^d})) + \dim(\mathfrak{n}_{d,\mathfrak{d}}^{\odot n}) - 2 \\ &= \dim(\mathcal{D}(\Phi_{\mu_\varepsilon^d})) + n + 2 - 2 \\ &= \dim(\mathcal{D}(\Phi_{\mu_\varepsilon^d})) + n.\end{aligned} \tag{107}$$

Similarly, for all $d, n \in \mathbb{N}$, $\varepsilon \in (0,1)$, $v \in \mathbb{R}^d$ we have that

$$\sigma_\varepsilon^d(\cdot)v \in \mathcal{R}(\{\Phi \in \mathbb{N}\colon \mathcal{D}(\Phi) = \mathcal{D}(\Phi_{\sigma_\varepsilon^d,0}) \odot \mathfrak{n}_{d,\mathfrak{d}}^{\odot n}\}) \tag{108}$$

and

$$\dim(\mathcal{D}(\Phi_{\sigma_\varepsilon^d,0}) \odot \mathfrak{n}_{d,\mathfrak{d}}^{\odot n}) = \dim(\mathcal{D}(\Phi_{\sigma_\varepsilon^d,0})) + n. \tag{109}$$

This and (105)–(108) prove that we can assume without lost of generality that

$$\dim(\mathcal{D}(\Phi_{\mu_\varepsilon^d})) = \dim(\mathcal{D}(\Phi_{\sigma_\varepsilon^d,0})). \tag{110}$$

Next, observe that for all $d \in \mathbb{N}$, $\theta \in \Theta$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1)$, $k \in [1,K] \cap \mathbb{Z}$, $t \in [0,T)$, $s \in [\frac{kT}{K}, \frac{(k+1)T}{K}]$ we have that

$$\begin{aligned} X_s^{d,\theta,K,\varepsilon,t,x}(\omega) &= X_{\max\{t,\frac{kT}{K}\}}^{d,\theta,K,\varepsilon,t,x}(\omega) + \mu_\varepsilon^d\left(X_{\max\{t,\frac{kT}{K}\}}^{d,\theta,K,\varepsilon,t,x}(\omega)\right)\left(s - \max\{t, \frac{kT}{K}\}\right) \\ &\quad + \sigma_\varepsilon^d\left(X_{\max\{t,\frac{kT}{K}\}}^{d,\theta,K,\varepsilon,t,x}(\omega)\right)\left(W_s^{d,\theta}(\omega) - W_{\max\{t,\frac{kT}{K}\}}^{d,\theta}(\omega)\right). \end{aligned} \tag{111}$$

Next, for every $d \in \mathbb{N}$, $\theta \in \Theta$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1)$, $k \in [1,K] \cap \mathbb{Z}$, $t \in [0,T)$, $s \in (t,T]$ let $J_k(s) \in \mathbb{R}$, $\phi_{t,s,k}^{d,\theta,K,\varepsilon}(x) \in \mathbb{R}^d$ satisfy that

$$\begin{aligned} J_k(s) &= \max\{t, \frac{(k-1)T}{K}\}\mathbb{1}_{[0,\max\{t,\frac{(k-1)T}{K}\}]}(s) \\ &\quad + s\mathbb{1}_{(\max\{t,\frac{(k-1)T}{K}\},\max\{t,\frac{kT}{K}\}]}(s) + \max\{t, \frac{kT}{K}\}\mathbb{1}_{(\max\{t,\frac{kT}{K}\},T]}(s) \end{aligned} \tag{112}$$

and

$$\phi_{t,s,k}^{d,\theta,K,\varepsilon}(x) = x + \mu_\varepsilon^d(x)\left(J_k(s) - \max\{t, \frac{(k-1)T}{K}\}\right) + \sigma_\varepsilon^d(x)\left(W_{J_k(s)}^{d,\theta}(\omega) - W_{\max\{t,\frac{(k-1)T}{K}\}}^{d,\theta}(\omega)\right). \tag{113}$$

Next, for every $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $k \in [1,K] \cap \mathbb{Z}$, $t \in [0,T)$, $s \in (t,T]$ let

$$\psi_{t,s,k}^{d,\theta,K,\varepsilon} = \phi_{t,s,k}^{d,\theta,K,\varepsilon} \circ \phi_{t,s,k-1}^{d,\theta,K,\varepsilon} \circ \ldots \circ \phi_{t,s,1}^{d,\theta,K,\varepsilon}. \tag{114}$$

Note that for all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $k \in [1,K-1] \cap \mathbb{Z}$, $s \in [0, \max\{t, \frac{(k-1)T}{K}\}]$ we have that $\phi_{t,s,k}^{d,\theta,K,\varepsilon} = \mathrm{Id}_{\mathbb{R}^d}$. This ensures for all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $k \in [1,K-1] \cap \mathbb{Z}$, $n \in [k+1,K] \cap \mathbb{Z}$, $s \in [0, \max\{t, \frac{kT}{K}\}]$ that $\psi_{t,s,k}^{d,\theta,K,\varepsilon} = \psi_{t,s,n}^{d,\theta,K,\varepsilon}$ and in particular $\psi_{t,s,k}^{d,\theta,K,\varepsilon} = \psi_{t,s,K}^{d,\theta,K,\varepsilon}$. Observe that for all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $k \in [1,K] \cap \mathbb{Z}$, $s \in [0, \max\{t, \frac{kT}{K}\}]$, $x \in \mathbb{R}^d$ that $\psi_{t,s,k}^{d,\theta,K,\varepsilon}(x) = X_s^{d,\theta,K,\varepsilon,t,x}(\omega)$. Therefore, for all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $s \in [0,T]$, $x \in \mathbb{R}^d$ we have that $\psi_{t,s,K}^{d,\theta,K,\varepsilon}(x) = X_s^{d,\theta,K,\varepsilon,t,x}$, i.e.,

$$X_s^{d,\theta,K,\varepsilon,t,x}(\omega) = \phi_{t,s,K}^{d,\theta,K,\varepsilon} \circ \phi_{t,s,K-1}^{d,\theta,K,\varepsilon} \circ \ldots \circ \phi_{t,s,1}^{d,\theta,K,\varepsilon}(x). \tag{115}$$

Next, (105), (113), (110), and Lemma 3.12 show for all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $k \in [1,K] \cap \mathbb{Z}$, $t \in [0,T)$, $s \in (t,T]$ that

$$\phi_{t,s,k}^{d,\theta,K,\varepsilon}(\cdot) \in \mathcal{R}\left(\left\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = \mathfrak{n}_{d,\mathfrak{d}}^{\odot \dim(\Phi_{\mu_\varepsilon^d})-2} \boxplus \mathcal{D}(\Phi_{\mu_\varepsilon^d}) \boxplus \mathcal{D}(\Phi_{\sigma_\varepsilon^d,0})\right\}\right). \tag{116}$$

This, (115), and Lemma 3.11 show that there exists $(\mathcal{X}_s^{d,\theta,K,\varepsilon,t})_{d\in\mathbb{N},\theta\in\Theta,\varepsilon\in(0,1),t\in[0,T),s\in(t,T]} \subseteq \mathbf{N}$ such that for all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t \in [0,T)$, $s \in (t,T]$, $x \in \mathbb{R}^d$ we have that

$$\begin{aligned} &\mathcal{D}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t}) = \bigodot_{k=1}^{K}\left[\mathfrak{n}_{d,\mathfrak{d}}^{\odot \dim(\Phi_{\mu_\varepsilon^d})-2} \boxplus \mathcal{D}(\Phi_{\mu_\varepsilon^d}) \boxplus \mathcal{D}(\Phi_{\sigma_\varepsilon^d,0})\right], \\ &(\mathcal{R}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t}))(x) = X_s^{d,\theta,K,\varepsilon,t,x}(\omega). \end{aligned} \tag{117}$$

This, the definition of $\odot$, and an induction argument show that for all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t \in [0,T)$, $s \in (t,T]$, $x \in \mathbb{R}^d$ we have that

$$\dim(\mathcal{D}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t})) = K(\dim(\mathcal{D}(\Phi_{\mu_\varepsilon^d})) - 2) + 2. \tag{118}$$

Next, (117), the definition of $\odot$, the triangle inequality (cf. Lemma 3.9), and (105) show that for all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t \in [0,T)$, $s \in (t,T]$, $x \in \mathbb{R}^d$ we have that

$$\begin{aligned} \left\| \mathcal{D}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t}) \right\| &= \left\| \bigodot_{k=1}^{K} \left[ \mathfrak{n}_{d,\mathfrak{d}}^{\odot \dim(\Phi_{\mu_\varepsilon^d})-2} \boxplus \mathcal{D}(\Phi_{\mu_\varepsilon^d}) \boxplus \mathcal{D}(\Phi_{\sigma_\varepsilon^d,0}) \right] \right\| \\ &\leqslant \left\| \mathfrak{n}_{d,\mathfrak{d}}^{\odot \dim(\Phi_{\mu_\varepsilon^d})-2} \boxplus \mathcal{D}(\Phi_{\mu_\varepsilon^d}) \boxplus \mathcal{D}(\Phi_{\sigma_\varepsilon^d,0}) \right\| \\ &\leqslant 3 \max\left\{ d\mathfrak{d}, \left\| \mathcal{D}(\Phi_{\mu_\varepsilon^d}) \right\|, \left\| \mathcal{D}(\Phi_{\sigma_\varepsilon^d,0}) \right\| \right\}. \end{aligned} \tag{119}$$

The proof of Lemma 3.14 is thus completed. □

**Lemma 3.15.** *Assume Setting 3.13. For every $d \in \mathbb{N}$, $\varepsilon \in (0,1)$ let $f_\varepsilon \in C(\mathbb{R},\mathbb{R})$, $g_\varepsilon^d \in C(\mathbb{R}^d,\mathbb{R})$, $\Phi_{f_\varepsilon}, \Phi_{g_\varepsilon^d} \in \mathbf{N}$ satisfy that $\mathcal{R}(\Phi_{f_\varepsilon}) = f_\varepsilon$ and $\mathcal{R}(\Phi_{g_\varepsilon^d}) = g_\varepsilon^d$. Let $\mathfrak{t}^\theta\colon [0,1] \to \mathbb{R}^d$, $\theta \in \Theta$, be independent random variables which satisfy for all $t \in [0,1]$ that $\mathbb{P}(\mathfrak{t}^0 \leqslant t) = t$. Assume that $(W^{d,\theta})_{d\in\mathbb{N},\theta\in\Theta}$ and $(\mathfrak{t}^\theta)_{\theta\in\Theta}$ are independent. For every $d \in \mathbb{N}$, $\varepsilon \in (0,1)$ let $U_{n,m}^{d,\theta,K,\varepsilon}\colon [0,T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$, $\theta \in \Theta$, $n,m \in \mathbb{Z}$, satisfy for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $m \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that*

$$\begin{aligned} U_{n,m}^{d,\theta,K,\varepsilon}(t,x) &= \frac{\mathbb{1}_{\mathbb{N}}(n)}{m^n} \sum_{i=1}^{m^n} g_\varepsilon^d\Big( X_T^{d,(\theta,0,-i),K,\varepsilon,t,x} \Big) \\ &+ \sum_{\ell=0}^{n-1} \frac{(T-t)}{m^{n-\ell}} \sum_{i=1}^{m^{n-\ell}} \Big( f_\varepsilon \circ U_{\ell,m}^{d,(\theta,\ell,i),K,\varepsilon} - \mathbb{1}_{\mathbb{N}}(\ell) f_\varepsilon \circ U_{\ell-1,m}^{d,(\theta,-\ell,i),K,\varepsilon} \Big) \Big( t + (T-t)\mathfrak{t}^{(\theta,\ell,i)}, X_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}}^{d,(\theta,\ell,i),K,\varepsilon,t,x} \Big). \end{aligned} \tag{120}$$

*For every $d \in \mathbb{N}$, $\varepsilon \in (0,1)$ let*

$$L_{d,\varepsilon} = K(\max\{\dim(\mathcal{D}(\Phi_{\mu_\varepsilon^d})), \dim(\mathcal{D}(\Phi_{\sigma_\varepsilon^d,0}))\} - 2) + 2. \tag{121}$$

*Let $(c_{d,\varepsilon})_{d\in\mathbb{N},\varepsilon\in(0,1)} \subseteq \mathbb{R}$ satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$ that*

$$c_{d,\varepsilon} \geqslant 3\max\{d\mathfrak{d}, \|\mathcal{D}(\Phi_{f_\varepsilon})\|, \|\mathcal{D}(\Phi_{g_\varepsilon^d})\|, \|\mathcal{D}(\Phi_{\mu_\varepsilon^d})\|, \|\mathcal{D}(\Phi_{\sigma_\varepsilon^d,0})\|\}. \tag{122}$$

*Let $\omega \in \Omega$. Then for all $m \in \mathbb{N}$, $d \in \mathbb{N}$, $n \in \mathbb{N}_0$, $\varepsilon \in (0,1)$ there exists $(\Phi_{n,m,t}^{d,\theta,K,\varepsilon})_{t\in[0,T],\theta\in\Theta} \subseteq \mathbf{N}$ such that the following items are true.*

*(i) We have for all $t_1, t_2 \in [0,T]$, $\theta_1, \theta_2 \in \Theta$ that $\mathcal{D}(\Phi_{n,m,t_1}^{d,\theta_1,K,\varepsilon}) = \mathcal{D}(\Phi_{n,m,t_2}^{d,\theta_2,K,\varepsilon})$.*

*(ii) We have for all $t \in [0,T]$, $\theta \in \Theta$ that*

$$\dim(\mathcal{D}(\Phi_{n,m,t}^{d,\theta,K,\varepsilon})) = n\,(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4) + \dim\big(\mathcal{D}(\Phi_{g_\varepsilon^d})\big) + L_{d,\varepsilon} - 2. \tag{123}$$

*(iii) We have for all $t \in [0,T]$, $\theta \in \Theta$ that $\left\| \mathcal{D}(\Phi_{n,m,t}^{d,\theta,K,\varepsilon}) \right\| \leqslant c_{d,\varepsilon}(3m)^n$.*

*(iv) We have for all $t \in [0,T]$, $\theta \in \Theta$, $x \in \mathbb{R}^d$ that $U_{n,m}^{d,\theta,K,\varepsilon}(t,x,\omega) = (\mathcal{R}(\Phi_{n,m,t}^{d,\theta,K,\varepsilon}))(x)$.*

*Proof of Lemma 3.15.* Throughout this proof let the notation in Setting 3.6 be given and let $d,m \in \mathbb{N}$, $\varepsilon \in (0,1)$ be fixed. Moreover, for every $n \in \mathbb{N}$ let $\mathfrak{n}_{1,\mathfrak{d}}^{\odot n} = \mathfrak{n}_{1,\mathfrak{d}} \odot \ldots \odot \mathfrak{n}_{1,\mathfrak{d}}$ ($n$ times). Lemmas 3.3 and 3.11 and a simple induction argument show for all $d,n \in \mathbb{N}$ that

$$\mathrm{Id}_{\mathbb{R}} \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = \mathfrak{n}_{1,\mathfrak{d}}^{\odot n}\}), \quad \mathfrak{n}_{1,\mathfrak{d}}^{\odot n} = (1,\mathfrak{d},\ldots,\mathfrak{d},1) \in \mathbb{R}^{n+2}. \tag{124}$$

Furthermore, Lemma 3.14 shows that there exists $(\mathcal{X}_s^{d,\theta,K,\varepsilon,t})_{d\in\mathbb{N},\theta\in\Theta,\varepsilon\in(0,1),t\in[0,T),s\in(t,T]} \subseteq \mathbf{N}$ such that the following items are true.

(A) For all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t \in [0,T)$, $s \in (t,T]$, $x \in \mathbb{R}^d$ we have that $\mathcal{R}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t}) \in C(\mathbb{R}^d,\mathbb{R}^d)$ and

$$(\mathcal{R}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t}))(x) = X_s^{d,\theta,K,\varepsilon,t,x}(\omega). \tag{125}$$

(B) For all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t_1 \in [0,T)$, $s_1 \in (t_1,T]$, $t_2 \in [0,T)$, $s_2 \in (t_2,T]$, $x \in \mathbb{R}^d$ we have that

$$\mathcal{D}(\mathcal{X}_{s_1}^{d,\theta_1,K,\varepsilon,t_1}) = \mathcal{D}(\mathcal{X}_{s_2}^{d,\theta_2,K,\varepsilon,t_2}). \tag{126}$$

(C) For all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t \in [0,T)$, $s \in (t,T]$ we have that

$$\dim(\mathcal{D}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t})) = K(\max\{\dim(\mathcal{D}(\Phi_{\mu_\varepsilon^d})), \dim(\mathcal{D}(\Phi_{\sigma_\varepsilon^d,0}))\} - 2) + 2. \tag{127}$$

(D) For all $d \in \mathbb{N}$, $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t \in [0,T)$, $s \in (t,T]$ we have that

$$\left\|\mathcal{D}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t})\right\| \leqslant 3\max\{d\mathfrak{d}, \|\mathcal{D}(\Phi_{\mu_\varepsilon^d})\|, \|\mathcal{D}(\Phi_{\sigma_\varepsilon^d,0})\|\}. \tag{128}$$

By (121) and (127) for all $\theta \in \Theta$, $\varepsilon \in (0,1)$, $t \in [0,T)$, $s \in (t,T]$ we have that

$$\dim(\mathcal{D}(\mathcal{X}_s^{d,\theta,K,\varepsilon,t})) = L_{d,\varepsilon}. \tag{129}$$

We will prove the result by induction. First, the base case is true since the zero function can be represented by DNN with arbitrary number of hidden layer. For the induction step $\mathbb{N}_0 \ni n \mapsto n+1 \in \mathbb{N}$ let $n \in \mathbb{N}_0$ and assume that there exists $(\Phi_{\ell,m,t}^{d,\theta,K,\varepsilon})_{t\in[0,T],\theta\in\Theta} \subseteq \mathbf{N}$, $\ell \in [0,n] \cap \mathbb{Z}$, such that the following items are true.

(A) We have for all $t_1, t_2 \in [0,T]$, $\theta_1, \theta_2 \in \Theta$, $\ell \in [0,n] \cap \mathbb{Z}$ that

$$\mathcal{D}(\Phi_{\ell,m,t_1}^{d,\theta_1,K,\varepsilon}) = \mathcal{D}(\Phi_{\ell,m,t_2}^{d,\theta_2,K,\varepsilon}). \tag{130}$$

(B) We have for all $t \in [0,T]$, $\theta \in \Theta$, $\ell \in [0,n] \cap \mathbb{Z}$ that

$$\dim(\mathcal{D}(\Phi_{\ell,m,t}^{d,\theta,K,\varepsilon})) = \ell\,(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4) + \dim\big(\mathcal{D}(\Phi_{g_\varepsilon^d})\big) + L_{d,\varepsilon} - 2. \tag{131}$$

(C) We have for all $t \in [0,T]$, $\theta \in \Theta$, $\ell \in [0,n] \cap \mathbb{Z}$ that

$$\left\|\mathcal{D}(\Phi_{\ell,m,t}^{d,\theta,K,\varepsilon})\right\| \leqslant c_{d,\varepsilon}(3m)^\ell. \tag{132}$$

(D) We have for all $t \in [0,T]$, $\theta \in \Theta$, $x \in \mathbb{R}^d$, $\ell \in [0,n] \cap \mathbb{Z}$ that

$$U_{\ell,m}^{d,\theta,K,\varepsilon}(t,x,\omega) = (\mathcal{R}(\Phi_{\ell,m,t}^{d,\theta,K,\varepsilon}))(x). \tag{133}$$

Next, Lemma 3.11, (124), the fact that $g_\varepsilon^d = \mathcal{R}(\Phi_{g_\varepsilon^d})$ prove for all $\theta \in \Theta$, $i \in [1, m^{n+1}] \cap \mathbb{Z}$, $t \in [0,T]$ that

$$\begin{aligned} &g_\varepsilon^d\Big(X_T^{d,(\theta,0,-i),K,\varepsilon,t,\cdot}\Big) = \mathrm{Id}_\mathbb{R}\Big(g_\varepsilon^d\Big(X_T^{d,(\theta,0,-i),K,\varepsilon,t,\cdot}\Big)\Big)\\ &\in \mathcal{R}\Big(\Big\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = \mathfrak{n}_{1,\mathfrak{d}}^{\odot(n+1)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L-4)} \odot \mathcal{D}(\Phi_{g_\varepsilon^d}) \odot \mathcal{D}(\mathcal{X}_T^{d,0,K,\varepsilon,0})\Big\}\Big) \end{aligned} \tag{134}$$

In addition, the definition of $\odot$, (124), and (129) imply that

$$\begin{aligned} &\dim\Big(\mathfrak{n}_{1,\mathfrak{d}}^{\odot(n+1)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L_{d,\varepsilon}-4)} \odot \mathcal{D}(\Phi_{g_\varepsilon^d}) \odot \mathcal{D}(\mathcal{X}_T^{d,0,K,\varepsilon,0})\Big)\\ &= \dim\Big(\mathfrak{n}_{1,\mathfrak{d}}^{\odot(n+1)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L_{d,\varepsilon}-4)}\Big) + \dim\big(\mathcal{D}(\Phi_{g_\varepsilon^d})\big) + \dim\Big(\mathcal{D}(\mathcal{X}_T^{d,0,K,\varepsilon,0})\Big) - 4\\ &= (n+1)(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4) + 2 + \dim\big(\mathcal{D}(\Phi_{g_\varepsilon^d})\big) + L_{d,\varepsilon} - 4\\ &= (n+1)(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4) + \dim\big(\mathcal{D}(\Phi_{g_\varepsilon^d})\big) + L_{d,\varepsilon} - 2. \end{aligned} \tag{135}$$

Furthermore, Lemma 3.11, the fact that $f_\varepsilon^d = \mathcal{R}(\Phi_{f_\varepsilon^d})$, (130), (133), (126), and (125) show for all $i \in [1,m]$, $\theta \in \Theta$, $t \in [0,T]$ that

$$\begin{aligned} &\Big(f_\varepsilon \circ U_{n,m}^{d,(\theta,n,i),K,\varepsilon}\Big)\Big(t + (T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega), X_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega)}^{d,(\theta,\ell,i),K,\varepsilon,t,\cdot}(\omega)\Big)\\ &\in \mathcal{R}\left(\Big\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = \mathcal{D}(\Phi_{f_\varepsilon}) \odot \mathcal{D}(\Phi_{n,m,0}^{d,0,K,\varepsilon}) \odot \mathcal{D}(\mathcal{X}_T^{d,0,K,\varepsilon,0})\Big\}\right). \end{aligned} \tag{136}$$

Moreover, the definition of $\odot$, (131), and (129) show that

$$\begin{aligned} &\dim\Big(\mathcal{D}(\Phi_{f_\varepsilon}) \odot \mathcal{D}(\Phi_{n,m,0}^{d,0,K,\varepsilon}) \odot \mathcal{D}(\mathcal{X}_T^{d,0,K,\varepsilon,0})\Big)\\ &= \dim(\mathcal{D}(\Phi_{f_\varepsilon})) + \dim\Big(\mathcal{D}(\Phi_{n,m,0}^{d,0,K,\varepsilon})\Big) + \dim\Big(\mathcal{D}(\mathcal{X}_T^{d,0,K,\varepsilon,0})\Big) - 4\\ &= \dim(\mathcal{D}(\Phi_{f_\varepsilon})) + n\,(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4) + \dim\big(\mathcal{D}(\Phi_{g_\varepsilon^d})\big) + L_{d,\varepsilon} - 2 + L_{d,\varepsilon} - 4\\ &= (n+1)\,(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4) + \dim\big(\mathcal{D}(\Phi_{g_\varepsilon^d})\big) + L_{d,\varepsilon} - 2. \end{aligned} \tag{137}$$

Furthermore, Lemma 3.11, the fact that $f_\varepsilon = \mathcal{R}(\Phi_{f_\varepsilon})$, (124), (133), (130), (125), and (126) show for all $\ell \in [0, n-1] \cap \mathbb{Z}$, $\theta \in \Theta$, $i \in [1, m^{n+1-\ell}] \cap \mathbb{Z}$, $t \in [0, T]$ that

$$\begin{aligned}
&\left(f_\varepsilon \circ U^{d,(\theta,\ell,i),K,\varepsilon}_{\ell,m}\right)\left(t + (T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega), X^{d,(\theta,\ell,i),K,\varepsilon,t,\cdot}_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega)}, \omega\right) \\
&= \left(f_\varepsilon \circ \mathrm{Id}_{\mathbb{R}} \circ U^{d,(\theta,\ell,i),K,\varepsilon}_{\ell,m}\right)\left(t + (T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega), X^{d,(\theta,\ell,i),K,\varepsilon,t,\cdot}_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega)}, \omega\right) \\
&\in \mathcal{R}\left(\left\{\Phi \in \mathbf{N} \colon \mathcal{D}(\Phi) = \mathcal{D}(\Phi_{f_\varepsilon}) \odot \mathfrak{n}_{1,\mathfrak{d}}^{\odot(n-\ell)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L_{d,\varepsilon}-4)} \odot \mathcal{D}(\Phi^{d,0,K,\varepsilon}_{\ell,m,0}) \odot \mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right\}\right).
\end{aligned} \tag{138}$$

Next, the definition of $\odot$, (124), (131), and (129) show for all $\ell \in [0, n-1] \cap \mathbb{Z}$ that

$$\begin{aligned}
&\dim\left(\mathcal{D}(\Phi_{f_\varepsilon}) \odot \mathfrak{n}_{1,\mathfrak{d}}^{\odot(n-\ell)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L_{d,\varepsilon}-4)} \odot \mathcal{D}(\Phi^{d,0,K,\varepsilon}_{\ell,m,0}) \odot \mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right) \\
&= \dim(\mathcal{D}(\Phi_{f_\varepsilon})) + \dim\left(\mathfrak{n}_{1,\mathfrak{d}}^{\odot(n-\ell)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L_{d,\varepsilon}-4)}\right) + \dim\left(\mathcal{D}(\Phi^{d,0,K,\varepsilon}_{\ell,m,0})\right) \\
&\quad + \dim\left(\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right) - 6 \\
&= \dim(\mathcal{D}(\Phi_{f_\varepsilon})) + (n-\ell)(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4) + 2 \\
&\quad + \ell\,(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4) + \dim\left(\mathcal{D}(\Phi_{g^d_\varepsilon})\right) + L_{d,\varepsilon} - 2 + L_{d,\varepsilon} - 6 \\
&= \dim(\mathcal{D}(\Phi_{f_\varepsilon})) + n(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4) \\
&\quad + \dim\left(\mathcal{D}(\Phi_{g^d_\varepsilon})\right) + L_{d,\varepsilon} - 2 + L_{d,\varepsilon} - 4 \\
&= (n+1)\,(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4) + \dim\left(\mathcal{D}(\Phi_{g^d_\varepsilon})\right) + L_{d,\varepsilon} - 2.
\end{aligned} \tag{139}$$

Similarly, for all $\ell \in [1, n] \cap \mathbb{Z}$, $\theta \in \Theta$, $i \in m^{n+1-\ell}$, $t \in [0, T]$ we have that

$$\begin{aligned}
&\left(f_\varepsilon \circ U^{d,(\theta,-\ell,i),K,\varepsilon}_{\ell-1,m}\right)\left(t + (T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega), X^{d,(\theta,\ell,i),K,\varepsilon,t,\cdot}_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega)}(\omega), \omega\right) \\
&= \left(f_\varepsilon \circ \mathrm{Id}_{\mathbb{R}} \circ U^{d,(\theta,-\ell,i),K,\varepsilon}_{\ell-1,m}\right)\left(t + (T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega), X^{d,(\theta,\ell,i),K,\varepsilon,t,\cdot}_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega)}(\omega), \omega\right) \\
&\in \mathcal{R}\left(\left\{\Phi \in \mathbf{N} \colon \mathcal{D}(\Phi) = \mathcal{D}(\Phi_{f_\varepsilon}) \odot \mathfrak{n}_{1,\mathfrak{d}}^{\odot(n-\ell+1)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L_{d,\varepsilon}-4)} \odot \mathcal{D}(\Phi^{d,0,K,\varepsilon}_{\ell-1,m,0}) \odot \mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right\}\right).
\end{aligned} \tag{140}$$

and

$$\begin{aligned}
&\dim\left(\mathcal{D}(\Phi_{f_\varepsilon}) \odot \mathfrak{n}_{1,\mathfrak{d}}^{\odot(n-\ell+1)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L_{d,\varepsilon}-4)} \odot \mathcal{D}(\Phi^{d,0,K,\varepsilon}_{\ell-1,m,0}) \odot \mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right) \\
&= (n+1)\,(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4) + \dim\left(\mathcal{D}(\Phi_{g^d_\varepsilon})\right) + L_{d,\varepsilon} - 2.
\end{aligned} \tag{141}$$

Now, (134)–(141) and Lemma 3.12 show that there exists $(\Phi^{d,\theta,K,\varepsilon}_{n+1,m,t})_{t\in[0,T],\theta\in\Theta}$ such that $t \in [0, T]$, $\theta \in \Theta$, $x \in \mathbb{R}^d$ we have that

$$\begin{aligned}
&(\mathcal{R}(\Phi^{d,\theta,K,\varepsilon}_{n+1,m,t}))(x) \\
&= \frac{1}{m^{n+1}} \sum_{i=1}^{m^{n+1}} g^d_\varepsilon\left(X^{d,(\theta,0,-i),K,\varepsilon,t,x}_T(\omega)\right) \\
&\quad + \frac{1}{m} \sum_{i=1}^{m} \left(f_\varepsilon \circ U^{d,(\theta,n,i),K,\varepsilon}_{n,m}\right)\left(\mathfrak{T}^{(\theta,\ell,i)}_t(\omega), X^{d,(\theta,\ell,i),K,\varepsilon,t,x}_{\mathfrak{T}^{(\theta,\ell,i)}_t(\omega)}(\omega), \omega\right) \\
&\quad + \sum_{\ell=0}^{n-1} \frac{(T-t)}{m^{n+1-\ell}} \sum_{i=1}^{m^{n+1-\ell}} \left(f_\varepsilon \circ U^{d,(\theta,\ell,i),K,\varepsilon}_{\ell,m}\right)\left(t + (T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega), X^{d,(\theta,\ell,i),K,\varepsilon,t,x}_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega)}(\omega), \omega\right) \\
&\quad - \sum_{\ell=1}^{n} \frac{(T-t)}{m^{n+1-\ell}} \sum_{i=1}^{m^{n+1-\ell}} \left(f_\varepsilon \circ U^{d,(\theta,-\ell,i),K,\varepsilon}_{\ell-1,m}\right)\left(t + (T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega), X^{d,(\theta,\ell,i),K,\varepsilon,t,x}_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}(\omega)}(\omega), \omega\right) \\
&= U^{d,\theta,K,\varepsilon}_{n+1,m}(t, x),
\end{aligned} \tag{142}$$

$$\dim\Big(\mathcal{D}(\Phi^{d,\theta,K,\varepsilon}_{n+1,m,t})\Big)\ (n+1)\left(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L_{d,\varepsilon}-4\right)+\dim\big(\mathcal{D}(\Phi_{g^d_\varepsilon})\big)+L_{d,\varepsilon}-2, \tag{143}$$

and

$$\begin{aligned}
&\mathcal{D}(\Phi^{d,\theta,K,\varepsilon}_{n+1,m,t})\\
&=\left[\mathop{\boxplus}_{i=1}^{m^{n+1}}\left[\mathfrak{n}_{1,\mathfrak{d}}^{\odot(n+1)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L-4)}\odot\mathcal{D}(\Phi_{g^d_\varepsilon})\odot\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right]\right]\\
&\quad\boxplus\left[\mathop{\boxplus}_{i=1}^{m}\left[\mathcal{D}(\Phi_{f_\varepsilon})\odot\mathcal{D}(\Phi^{d,0,K,\varepsilon}_{n,m,0})\odot\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right]\right]\\
&\quad\boxplus\left[\mathop{\boxplus}_{\ell=0}^{n-1}\mathop{\boxplus}_{i=1}^{m^{n+1-\ell}}\left[\mathcal{D}(\Phi_{f_\varepsilon})\odot\mathfrak{n}_{1,\mathfrak{d}}^{\odot(n-\ell)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L-4)}\odot\mathcal{D}(\Phi^{d,0,K,\varepsilon}_{\ell,m,0})\odot\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right]\right]\\
&\quad\boxplus\left[\mathop{\boxplus}_{\ell=1}^{n}\mathop{\boxplus}_{i=1}^{m^{n+1-\ell}}\left[\mathcal{D}(\Phi_{f_\varepsilon})\odot\mathfrak{n}_{1,\mathfrak{d}}^{\odot(n-\ell+1)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L-4)}\odot\mathcal{D}(\Phi^{d,0,K,\varepsilon}_{\ell-1,m,0})\odot\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right]\right].
\end{aligned}\tag{144}$$

This shows for all $t_1,t_2\in[0,T]$, $\theta_1,\theta_2\in\Theta$ that

$$\mathcal{D}(\Phi^{d,\theta_1,K,\varepsilon}_{n+1,m,t_1})=\mathcal{D}(\Phi^{d,\theta_2,K,\varepsilon}_{n+1,m,t_2}).\tag{145}$$

Next, the definition of $\odot$, (124), (128), and (122) prove that

$$\begin{aligned}
&\left|\!\left|\!\left|\mathfrak{n}_{1,\mathfrak{d}}^{\odot(n+1)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L-4)}\odot\mathcal{D}(\Phi_{g^d_\varepsilon})\odot\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right|\!\right|\!\right|\\
&\leqslant\max\Big\{d\mathfrak{d},|||\mathcal{D}(\Phi_{g^d_\varepsilon})|||,\Big|\!\Big|\!\Big|\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\Big|\!\Big|\!\Big|\Big\}\\
&\leqslant\max\big\{d\mathfrak{d},|||\mathcal{D}(\Phi_{g^d_\varepsilon})|||,3\max\{d\mathfrak{d},|||\mathcal{D}(\Phi_{\mu^d_\varepsilon})|||,|||\mathcal{D}(\Phi_{\sigma^d_\varepsilon,0})|||\}\big\}\\
&\leqslant3\max\big\{d\mathfrak{d},|||\mathcal{D}(\Phi_{g^d_\varepsilon})|||,|||\mathcal{D}(\Phi_{\mu^d_\varepsilon})|||,|||\mathcal{D}(\Phi_{\sigma^d_\varepsilon,0})|||\big\}\\
&\leqslant c_{d,\varepsilon}.
\end{aligned}\tag{146}$$

Furthermore, the definition of $\odot$, (132), (128), and (122) prove that

$$\begin{aligned}
&\left|\!\left|\!\left|\mathcal{D}(\Phi_{f_\varepsilon})\odot\mathcal{D}(\Phi^{d,0,K,\varepsilon}_{n,m,0})\odot\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right|\!\right|\!\right|\\
&\leqslant\max\Big\{|||\mathcal{D}(\Phi_{f_\varepsilon})|||,\Big|\!\Big|\!\Big|\mathcal{D}(\Phi^{d,0,K,\varepsilon}_{n,m,0})\Big|\!\Big|\!\Big|,\Big|\!\Big|\!\Big|\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\Big|\!\Big|\!\Big|\Big\}\\
&\leqslant\max\big\{|||\mathcal{D}(\Phi_{f_\varepsilon})|||,c_{d,\varepsilon}(3m)^n,3\max\{d\mathfrak{d},|||\mathcal{D}(\Phi_{\mu^d_\varepsilon})|||,|||\mathcal{D}(\Phi_{\sigma^d_\varepsilon,0})|||\}\big\}\\
&\leqslant c_{d,\varepsilon}(3m)^n.
\end{aligned}\tag{147}$$

In addition, the definition of $\odot$, (124), (132), (128), and (122) show for all $\ell\in[0,n-1]\cap\mathbb{Z}$ that

$$\begin{aligned}
&\left|\!\left|\!\left|\mathcal{D}(\Phi_{f_\varepsilon})\odot\mathfrak{n}_{1,\mathfrak{d}}^{\odot(n-\ell)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L-4)}\odot\mathcal{D}(\Phi^{d,0,K,\varepsilon}_{\ell,m,0})\odot\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right|\!\right|\!\right|\\
&\leqslant\max\Big\{|||\mathcal{D}(\Phi_{f_\varepsilon})|||,d\mathfrak{d},\Big|\!\Big|\!\Big|\mathcal{D}(\Phi^{d,0,K,\varepsilon}_{\ell,m,0})\Big|\!\Big|\!\Big|,\Big|\!\Big|\!\Big|\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\Big|\!\Big|\!\Big|\Big\}\\
&\leqslant\max\big\{|||\mathcal{D}(\Phi_{f_\varepsilon})|||,d\mathfrak{d},c_{d,\varepsilon}(3m)^\ell,3\max\{d\mathfrak{d},|||\mathcal{D}(\Phi_{\mu^d_\varepsilon})|||,|||\mathcal{D}(\Phi_{\sigma^d_\varepsilon,0})|||\}\big\}\\
&\leqslant c_{d,\varepsilon}(3m)^\ell.
\end{aligned}\tag{148}$$

Similarly, we have for all $\ell\in[1,n]\cap\mathbb{Z}$ that

$$\begin{aligned}
&\left|\!\left|\!\left|\mathcal{D}(\Phi_{f_\varepsilon})\odot\mathfrak{n}_{1,\mathfrak{d}}^{\odot(n-\ell+1)(\dim(\mathcal{D}(\Phi_{f_\varepsilon}))+L-4)}\odot\mathcal{D}(\Phi^{d,0,K,\varepsilon}_{\ell-1,m,0})\odot\mathcal{D}(\mathcal{X}^{d,0,K,\varepsilon,0}_T)\right|\!\right|\!\right|\\
&\leqslant c_{d,\varepsilon}(3m)^{\ell-1}.
\end{aligned}\tag{149}$$

This, (144), (146)–(148), the triangle inequality (cf. Lemma 3.9) show that for all $\theta \in \Theta$, $t \in [0,T]$ that

$$\begin{aligned}
&\Big|\!\Big|\!\Big|\mathcal{D}(\Phi^{d,\theta,K,\varepsilon}_{n+1,m,t})\Big|\!\Big|\!\Big| \\
&\leqslant \left[\sum_{i=1}^{m^{n+1}} c_{d,\varepsilon}\right] + \left[\sum_{i=1}^{m} c_{d,\varepsilon}(3m)^n\right] + \left[\sum_{\ell=0}^{n-1}\sum_{i=1}^{m^{n+1-\ell}} c_{d,\varepsilon}(3m)^\ell\right] + \left[\sum_{\ell=1}^{n}\sum_{i=1}^{m^{n+1-\ell}} c_{d,\varepsilon}(3m)^{\ell-1}\right] \\
&= m^{n+1}c_{d,\varepsilon} + mc_{d,\varepsilon}(3m)^n + \left[\sum_{\ell=0}^{n-1} m^{n+1-\ell}c_{d,\varepsilon}(3m)^\ell\right] + \left[\sum_{\ell=1}^{n} m^{n+1-\ell}c_{d,\varepsilon}(3m)^{\ell-1}\right] \\
&= m^{n+1}c_{d,\varepsilon}\left[1 + 3^n + \sum_{\ell=0}^{n-1} 3^\ell + \sum_{\ell=1}^{n} 3^{\ell-1}\right] = m^{n+1}c_{d,\varepsilon}\left[1 + \sum_{\ell=0}^{n} 3^\ell + \sum_{\ell=1}^{n} 3^{\ell-1}\right] \\
&\leqslant cm^{n+1}\left[1 + 2\sum_{\ell=0}^{n} 3^\ell\right] = cm^{n+1}\left[1 + 2\frac{3^{n+1}-1}{3-1}\right] = c_{d,\varepsilon}(3m)^{n+1}.
\end{aligned} \tag{150}$$

This, (145), (142), and (143) complete the induction step. The proof of Lemma 3.15 is thus completed. □

## 4. DNN approximations for PDEs

**Theorem 4.1.** *Assume Setting 1.2. Let $\mathfrak{d} \in \mathbb{N}$, $\mathfrak{n}_{1,\mathfrak{d}} = (1,\mathfrak{d},1) \in \mathbf{D}$ satisfy that*

$$\mathrm{Id}_{\mathbb{R}} \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = \mathfrak{n}_{1,\mathfrak{d}}\}). \tag{151}$$

*Let $\beta, \mathfrak{p} \in [2,\infty)$, $c \in [\max\{3\mathfrak{d}, \beta^2\mathfrak{p}^2\},\infty)$. For every $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $v \in \mathbb{R}^d$ let $\Phi_{f_\varepsilon}$, $\Phi_{\mu^d_\varepsilon}, \Phi_{\sigma^d_\varepsilon,v}, \Phi_{g^d_\varepsilon} \in \mathbf{N}$, $f, f_\varepsilon \in C(\mathbb{R},\mathbb{R})$, $g^d, g^d_\varepsilon \in C(\mathbb{R}^d,\mathbb{R})$, $\mu^d, \mu^d_\varepsilon \in C(\mathbb{R}^d,\mathbb{R}^d)$, $\sigma^d, \sigma^d_\varepsilon \in C(\mathbb{R}^{d\times d},\mathbb{R}^d)$ satisfy for all $v \in \mathbb{R}^d$ that $f_\varepsilon = \Phi_{f_\varepsilon}$, $\mu^d_\varepsilon = \mathcal{R}(\Phi_{\mu^d_\varepsilon})$, $\sigma^d_\varepsilon(\cdot)v = \mathcal{R}(\Phi_{\sigma^d_\varepsilon,v})$, $g^d_\varepsilon = \mathcal{R}(\Phi_{g^d_\varepsilon})$. Assume for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $v \in \mathbb{R}^d$ that $\mathcal{D}(\Phi_{\sigma^d_\varepsilon,v}) = \mathcal{D}(\Phi_{\sigma^d_\varepsilon,0})$. Assume for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $v,w \in \mathbb{R}$, $x,y \in \mathbb{R}^d$ that*

$$\max\{\|\mu^d_\varepsilon(x) - \mu^d_\varepsilon(y)\|, \|\sigma^d_\varepsilon(x) - \sigma^d_\varepsilon(y)\|\} \leqslant c\|x-y\|, \tag{152}$$

$$|f_\varepsilon(w) - f_\varepsilon(v)| \leqslant c|w-v|, \quad |g^d_\varepsilon(x) - g^d_\varepsilon(y)| \leqslant c\frac{(d^c + \|x\|)^\beta + (d^c + \|y\|)^\beta}{2\sqrt{T}}\|x-y\|, \tag{153}$$

$$|g^d_\varepsilon(x)| \leqslant c(d^c + \|x\|)^\beta, \quad \max\big\{\|\mu^d_\varepsilon(0)\|, \|\sigma^d_\varepsilon(0)\|, |Tf_\varepsilon(0)|, |g^d_\varepsilon(0)|\big\} \leqslant cd^c, \tag{154}$$

$$\max\{\|\mu^d_\varepsilon(x) - \mu^d(x)\|, \|\sigma^d_\varepsilon(x) - \sigma^d(x)\|, \|g^d_\varepsilon(x) - g^d(x)\|\} \leqslant \varepsilon cd^c(d^c + \|x\|)^\beta, \tag{155}$$

$$|f_\varepsilon(w) - f(w)| \leqslant \varepsilon(1 + |w|^\beta), \tag{156}$$

$$\max\big\{|\!|\!|\mathcal{D}(\Phi_{f_\varepsilon})|\!|\!|, |\!|\!|\mathcal{D}(\Phi_{g^d_\varepsilon})|\!|\!|, |\!|\!|\mathcal{D}(\Phi_{\mu^d_\varepsilon})|\!|\!|, |\!|\!|\mathcal{D}(\Phi_{\sigma^d_\varepsilon,0})|\!|\!|\big\} \leqslant cd^c\varepsilon^{-c}, \tag{157}$$

$$\max\big\{\dim(\mathcal{D}(\Phi_{f_\varepsilon})), \dim(\mathcal{D}(\Phi_{g^d_\varepsilon})), \dim(\mathcal{D}(\Phi_{\mu^d_\varepsilon})), \dim(\mathcal{D}(\Phi_{\sigma^d_\varepsilon,0}))\big\} \leqslant cd^c\varepsilon^{-c}. \tag{158}$$

*Then the following items are true.*

*(i) For every $d \in \mathbb{N}$ there exists a unique at most polynomially growing viscosity solution $u^d$ of*

$$\frac{\partial u^d}{\partial t}(t,x) + \frac{1}{2}\mathrm{trace}(\sigma^d(x)(\sigma^d(x))^\top(\mathrm{Hess}_x u^d(t,x))) + \langle \mu^d(x), (\nabla_x u^d)(t,x)\rangle + f(u^d(t,x)) = 0 \tag{159}$$

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

*(ii) There exists* $(C_\delta)_{\delta\in(0,1)} \subseteq (0,\infty)$, $\eta \in (0,\infty)$, $(\Psi_{d,\epsilon})_{d\in\mathbb{N},\epsilon\in(0,1)} \subseteq \mathbf{N}$ *such that for all* $d \in \mathbb{N}$, $\epsilon \in (0,1)$ *we have that* $\mathcal{R}(\Psi_{d,\epsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, $\mathcal{P}(\Psi_{d,\epsilon}) \leqslant C_\delta \eta d^\eta \epsilon^{-(4+\delta)-6c}$, *and*

$$\left(\int_{[0,1]^d} \left|(\mathcal{R}(\Psi_{d,\epsilon}))(x) - u^d(0,x)\right|^{\mathfrak{p}} dx\right)^{\frac{1}{\mathfrak{p}}} < \epsilon. \tag{160}$$

*Proof of Theorem 4.1.* Let $p \in [3,\infty)$ satisfy that $p = \beta\mathfrak{p}$. For every $d \in \mathbb{N}$ let $\varphi_d \in C(\mathbb{R}^d,[1,\infty))$ satisfy for all $x \in \mathbb{R}^d$ that

$$\varphi_d(x) = 2^p c^p d^{pc}(d^{2c} + \|x\|^2)^{\frac{p}{2}}. \tag{161}$$

Then (154) shows for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} &\max\{\|\mu_\varepsilon^d(0)\| + c\|x\|, \|\sigma_\varepsilon^d(0)\| + c\|x\|\} \\ &\leqslant cd^c + c\|x\| = c(d^c + \|x\|) \leqslant 2c(d^{2c} + \|x\|^2)^{\frac{1}{2}} \leqslant c(\varphi_d(x))^{\frac{1}{p}}. \end{aligned} \tag{162}$$

Next, [47, Lemma 2.6] (applied for every $d \in \mathbb{N}$ with $d \curvearrowleft d$, $m \curvearrowleft d$, $a \curvearrowleft d^{2c}$, $c \curvearrowleft 0$, $p \curvearrowleft p/2$, $\mu \curvearrowleft 0$, $\sigma \curvearrowleft 0$, $\varphi \curvearrowleft \varphi_d/(2^p c^p d^{pc})$ in the notation of [47, Lemma 2.6]) and (161) show for all $x, z \in \mathbb{R}^d$ that

$$\|(\varphi_d'(x))(z)\| \leqslant p(\varphi_d(x))^{1-\frac{1}{p}}\|z\|, \quad \|(\varphi_d''(x))(z,z)\| \leqslant p^2(\varphi(x))^{1-\frac{2}{p}}\|z\|^2. \tag{163}$$

This, (162), and the fact that $p^2 = \beta^2\mathfrak{p}^2 \leqslant c$ show for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $x, z \in \mathbb{R}^d$ that

$$\max\left\{\frac{|(\varphi_d'(x))(z)|}{(\varphi_d(x))^{\frac{p-1}{p}}\|z\|}, \frac{(\varphi_d''(x))(z,z)}{(\varphi_d(x))^{\frac{p-2}{p}}\|z\|^2}, \frac{c\|x\| + \|\mu_\varepsilon^d(0)\|}{(\varphi_d(x))^{\frac{1}{p}}}, \frac{c\|x\| + \|\sigma_\varepsilon^d(0)\|}{(\varphi_d(x))^{\frac{1}{p}}}\right\} \leqslant c, \tag{164}$$

This, (152), and (155) we have for all $d \in \mathbb{N}$, $x, z \in \mathbb{R}^d$ that

$$\max\left\{\frac{|(\varphi_d'(x))(z)|}{(\varphi_d(x))^{\frac{p-1}{p}}\|z\|}, \frac{(\varphi_d''(x))(z,z)}{(\varphi_d(x))^{\frac{p-2}{p}}\|z\|^2}, \frac{c\|x\| + \|\mu^d(0)\|}{(\varphi_d(x))^{\frac{1}{p}}}, \frac{c\|x\| + \|\sigma^d(0)\|}{(\varphi_d(x))^{\frac{1}{p}}}\right\} \leqslant c. \tag{165}$$

and

$$\max\{\|\mu^d(x) - \mu^d(y)\|, \|\sigma^d(x) - \sigma^d(y)\|\} \leqslant c\|x - y\|. \tag{166}$$

Furthermore, (153), (154), and (155) prove for all $d \in \mathbb{N}$, $w, v \in \mathbb{R}$, $x, y \in \mathbb{R}^d$ that

$$|f(w) - f(v)| \leqslant c|w - v|, \quad |g^d(x) - g^d(y)| \leqslant c\frac{(d^c + \|x\|)^\beta + (d^c + \|y\|)^\beta}{2\sqrt{T}}\|x - y\|, \tag{167}$$

$$\max\left\{\|\mu^d(0)\|, \|\sigma^d(0)\|, |f(0)|, |g^d(0)|\right\} \leqslant cd^c. \tag{168}$$

Let $(\Omega,\mathcal{F},\mathbb{P},(\mathbb{F}_t)_{t\in[0,T]})$ be a filtered probability space which satisfies the usual conditions. Let $\Theta = \bigcup_{n\in\mathbb{N}}\mathbb{Z}^n$. Let $\mathfrak{t}^\theta\colon \Omega \to [0,1]$, $\theta \in \Theta$, be identically distributed and independent random variables. Assume for all $t \in (0,1)$ that $\mathbb{P}(\mathfrak{t}^0 \leqslant t) = t$. For every $d \in \mathbb{N}$ let $W^{d,\theta}\colon [0,T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, be independent standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motions. Assume that $(\mathfrak{t}^\theta)_{\theta\in\Theta}$ and $(W^{d,\theta})_{\theta\in\Theta,d\in\mathbb{N}}$ are independent. For every $K \in \mathbb{N}$ let $\lfloor\cdot\rfloor_K\colon \mathbb{R} \to \mathbb{R}$ satisfy for all $t \in \mathbb{R}$ that $\lfloor t\rfloor_K = \max(\{0, \frac{T}{K}, \ldots, \frac{(K-1)T}{T}, T\} \cap ((-\infty,t) \cup \{0\}))$. For every $\theta \in \Theta$, $d, K \in \mathbb{N}$, $\varepsilon \in (0,1)$, $t \in [0,T]$, $x \in \mathbb{R}^d$ let $X^{d,\theta,K,\varepsilon,t,x} = (X_s^{d,\theta,K,\varepsilon,t,x})_{s\in[t,T]}\colon [t,T] \times \Omega \to \mathbb{R}^d$ satisfy for all $s \in [t,T]$ that $X_t^{d,\theta,K,\varepsilon,t,x} = x$ and

$$\begin{aligned} X_s^{d,\theta,K,\varepsilon,t,x} = X_{\max\{t,\lfloor s\rfloor_K\}}^{d,\theta,K,\varepsilon,t,x} + \mu_\varepsilon^d(X_{\max\{t,\lfloor s\rfloor_K\}}^{d,\theta,K,\varepsilon,t,x})(s - \max\{t,\lfloor s\rfloor_K\}) \\ + \sigma_\varepsilon^d(X_{\max\{t,\lfloor s\rfloor_K\}}^{d,\theta,K,\varepsilon,t,x})(W_s^\theta - W_{\max\{t,\lfloor s\rfloor_K\}}^\theta). \end{aligned} \tag{169}$$

Let $U^{d,\theta,K,\varepsilon}_{n,m}\colon [0,T]\times\mathbb{R}^d\times\Omega\to\mathbb{R}$, $n\in\mathbb{Z}$, $K,d,m\in\mathbb{N}$, $\theta\in\Theta$, $\varepsilon\in(0,1)$, satisfy for all $\theta\in\Theta$, $K,d,m\in\mathbb{N}$, $n\in\mathbb{N}_0$, $t\in[0,T]$, $x\in\mathbb{R}^d$, $\varepsilon\in(0,1)$ that $U^{d,\theta,K,\varepsilon}_{-1,m}(t,x)=U^{d,\theta,K,\varepsilon}_{0,m}(t,x)=0$ and

$$
\begin{aligned}
&U^{d,\theta,K,\varepsilon}_{n,m}(t,x)=\frac{1}{m^n}\sum_{i=1}^{m^n} g^d_\varepsilon(X^{d,(\theta,0,-i),K,\varepsilon,t,x}_T)\\
&+\sum_{\ell=0}^{n-1}\frac{T-t}{m^{n-\ell}}\sum_{i=1}^{m^{n-\ell}}\bigl(f_\varepsilon\circ U^{d,(\theta,\ell,i),K,\varepsilon}_{\ell,m}-\mathbb{1}_{\mathbb{N}}(\ell)f_\varepsilon\circ U^{d,(\theta,-\ell,i),K,\varepsilon}_{\ell-1,m}\bigr)\Bigl(t+(T-t)\mathfrak{t}^{(\theta,\ell,i)},X^{d,(\theta,\ell,i),K,\varepsilon,t,x}_{t+(T-t)\mathfrak{t}^{(\theta,\ell,i)}}\Bigr).
\end{aligned}
\tag{170}
$$

Next, (166) and (152) prove for all $d\in\mathbb{N}$, $\varepsilon\in(0,1)$, $t\in[0,T]$, $\theta\in\Theta$ that there exist up to indistinguishability unique continuous random fields $X^{d,\theta,\varepsilon,t,\cdot}=(X^{d,\theta,t,x}_s)_{s\in[t,T],x\in\mathbb{R}^d}, X^{d,\theta,t,\cdot}=(X^{d,\theta,t,x}_s)_{s\in[t,T],x\in\mathbb{R}^d}\colon[t,T]\times\mathbb{R}^d\times\Omega\to\mathbb{R}^d$ which satisfy that for all $x\in\mathbb{R}^d$ it holds that $(X^{d,\theta,\varepsilon,t,x}_s)_{s\in[t,T]}$, $(X^{d,\theta,t,x}_s)_{s\in[t,T]}$ are $(\mathbb{F}_s)_{s\in[t,T]}$-adapted and which satisfy that for all $s\in[t,T]$, $x\in\mathbb{R}^d$ it holds $\mathbb{P}$-a.s. that

$$
X^{d,\theta,\varepsilon,t,x}_s=x+\int_t^s\mu^d_\varepsilon(X^{d,\theta,\varepsilon,t,x}_r)\,dr+\int_t^s\sigma^d_\varepsilon(X^{d,\theta,\varepsilon,t,x}_r)\,dW^{d,\theta}_r,
\tag{171}
$$

$$
X^{d,\theta,t,x}_s=x+\int_t^s\mu^d(X^{d,\theta,t,x}_r)\,dr+\int_t^s\sigma^d(X^{d,\theta,t,x}_r)\,dW^{d,\theta}_r.
\tag{172}
$$

Hence, [17, Lemma 2.1] (applied for every $\theta\in\Theta$, $d\in\mathbb{N}$, $\varepsilon\in(0,1)$ with $d\curvearrowleft d$, $m\curvearrowleft m$, $c\curvearrowleft c$, $\kappa\curvearrowleft 1$, $p\curvearrowleft p$, $\varphi\curvearrowleft\varphi_d$, $\mu\curvearrowleft\mu^d_\varepsilon$, $\sigma\curvearrowleft\sigma^d_\varepsilon$ and applied for every $\theta\in\Theta$, $d\in\mathbb{N}$ with $d\curvearrowleft d$, $m\curvearrowleft m$, $c\curvearrowleft c$, $\kappa\curvearrowleft 1$, $p\curvearrowleft p$, $\varphi\curvearrowleft\varphi_d$, $\mu\curvearrowleft\mu^d$, $\sigma\curvearrowleft\sigma^d$ in the notation of [17, Lemma 2.1]), (152), and (166) prove for all $d\in\mathbb{N}$, $\varepsilon\in(0,1)$, $x\in\mathbb{R}^d$ that

$$
\langle(\nabla\varphi_d)(x),\mu^d_\varepsilon(x)\rangle+\frac{1}{2}\mathrm{trace}(\sigma^d_\varepsilon(\sigma^d_\varepsilon)^\top\mathrm{Hess}\varphi_d(x))\leqslant 1.5c^3\varphi_d(x),
\tag{173}
$$

$$
\langle(\nabla\varphi_d)(x),\mu^d(x)\rangle+\frac{1}{2}\mathrm{trace}(\sigma^d(\sigma^d)^\top(\mathrm{Hess}\varphi_d)(x))\leqslant 1.5c^3\varphi_d(x),
\tag{174}
$$

$$
\max\bigl\{\mathbb{E}\bigl[\varphi_d(X^{d,\theta,\varepsilon,t,x}_s)\bigr],\mathbb{E}\bigl[\varphi_d(X^{d,\theta,t,x}_s)\bigr]\bigr\}\leqslant e^{1.5c^3(s-t)}\varphi(x),
\tag{175}
$$

This, [41, Proposition 2.2] (applied for every $d\in\mathbb{N}$ with $d\curvearrowleft d$, $L\curvearrowleft c$, $T\curvearrowleft T$, $\mathcal{O}\curvearrowleft\mathbb{R}^d$, $\|\cdot\|\curvearrowleft\|\cdot\|$, $f\curvearrowleft([0,T]\times\mathbb{R}^d\times\mathbb{R}\ni(t,x,w)\mapsto f^d(w)\in\mathbb{R})$, $g\curvearrowleft g^d$, $(X^x_{t,s})_{t\in[0,T],s\in[t,T],x\in\mathbb{R}^d}\curvearrowleft(X^{d,0,t,x}_s)_{t\in[0,T],s\in[t,T],x\in\mathbb{R}^d}$, $V\curvearrowleft([0,T]\times\mathbb{R}^d\ni(s,x)\mapsto e^{1.5c(T-s)}\varphi_d(s,x)\in(0,\infty))$ and applied for every $d\in\mathbb{N}$, $\varepsilon\in(0,1)$ with $d\curvearrowleft d$, $L\curvearrowleft c$, $T\curvearrowleft T$, $\mathcal{O}\curvearrowleft\mathbb{R}^d$, $\|\cdot\|\curvearrowleft\|\cdot\|$, $f\curvearrowleft([0,T]\times\mathbb{R}^d\times\mathbb{R}\ni(t,x,w)\mapsto f^d_\varepsilon(w)\in\mathbb{R})$, $g\curvearrowleft g^d_\varepsilon$, $(X^x_{t,s})_{t\in[0,T],s\in[t,T],x\in\mathbb{R}^d}\curvearrowleft(X^{d,0,\varepsilon,t,x}_s)_{t\in[0,T],s\in[t,T],x\in\mathbb{R}^d}$, $V\curvearrowleft([0,T]\times\mathbb{R}^d\ni(s,x)\mapsto e^{1.5c(T-s)}\varphi_d(s,x)\in(0,\infty))$ in the notation of [41, Proposition 2.2]), (153), (154), (161), (167), and (168) show for all $d\in\mathbb{N}$, $\varepsilon\in(0,1)$ that there exist unique measurable $u^d,u^{d,\varepsilon}\colon[0,T]\times\mathbb{R}^d\to\mathbb{R}$ such that for all $t\in[0,T]$, $x\in\mathbb{R}^d$ it holds that $\sup_{s\in[0,T],y\in\mathbb{R}^d}\frac{|u^d(s,y)|+|u^{d,\varepsilon}(s,y)|}{\varphi_d(s,y)}<\infty$, $\mathbb{E}\Bigl[|g^d(X^{d,0,t,x}_T)|\Bigr]+\int_t^T\mathbb{E}\bigl[|f(u^d(s,X^{d,0,t,x}_s))|\bigr]\,ds+\mathbb{E}\Bigl[|g^d(X^{d,0,\varepsilon,t,x}_T)|\Bigr]+\int_t^T\mathbb{E}\bigl[|f(u^d(s,X^{d,0,\varepsilon,t,x}_s))|\bigr]\,ds<\infty$,

$$
u^d(t,x)=\mathbb{E}\Bigl[g^d(X^{d,0,t,x}_T)\Bigr]+\int_t^T\mathbb{E}\bigl[f(u^d(s,X^{d,0,t,x}_s))\bigr]\,ds
\tag{176}
$$

and

$$
u^{d,\varepsilon}(t,x)=\mathbb{E}\Bigl[g^d_\varepsilon(X^{d,0,\varepsilon,t,x}_T)\Bigr]+\int_t^T\mathbb{E}\bigl[f_\varepsilon(u^d(s,X^{d,0,\varepsilon,t,x}_s))\bigr]\,ds.
\tag{177}
$$

Next, the triangle inequality, (152), (154), and the fact that $\forall\, x \in \mathbb{R}^d\colon (1+x)^2 \leqslant 1(1+x^2)$ show for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1)$ that

$$\begin{aligned} \langle x, \mu_\varepsilon^d(x)\rangle &\leqslant \|x\| \left(\|\mu_\varepsilon^d(x) - \mu_\varepsilon^d(0)\| + \|\mu_\varepsilon^d(0)\|\right) \\ &\leqslant \|x\|(c\|x\| + cd^c) \\ &\leqslant (1+\|x\|)^2 cd^c \\ &\leqslant 2cd^c(1+\|x\|^2). \end{aligned} \tag{178}$$

Furthermore, the Cauchy-Schwarz inequality implies for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $\varepsilon \in (0,1)$ that

$$\|\sigma_\varepsilon^d(x)y\|^2 = \sum_{i=1}^d \left|\sum_{j=1}^d (\sigma_\varepsilon^d)_{ij}(x) y_j\right|^2 \leqslant \sum_{i=1}^d \left(\sum_{j=1}^d |(\sigma_\varepsilon^d)_{ij}(x)|^2\right)\left(\sum_{j=1}^d |y_j|^2\right) \leqslant \|\sigma(x)\|^2\|y\|^2. \tag{179}$$

This and (152) show for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $\varepsilon \in (0,1)$ that

$$\|\sigma_\varepsilon^d(x)y\| \leqslant \|\sigma_\varepsilon^d(x)\|\|y\| \leqslant (\|\sigma_\varepsilon^d(x) - \sigma_\varepsilon^d(0)\| + \|\sigma_\varepsilon^d(0)\|)\|y\| \leqslant (c\|x\| + cd^c)\|y\| \leqslant cd^c(1+\|x\|)\|y\|. \tag{180}$$

This, (178), and (155) prove for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $\varepsilon \in (0,1)$ that

$$\langle x, \mu^d(x)\rangle \leqslant 2cd^c(1+\|x\|^2), \quad \|\sigma^d(x)y\| \leqslant cd^c(1+\|x\|)\|y\|. \tag{181}$$

This, [10, Theorem 1.1] (applied with $d \curvearrowleft d$, $L \curvearrowleft 2cd^c$, $T \curvearrowleft T$, $\mu \curvearrowleft \mu^d$, $\sigma \curvearrowleft \sigma^d$, $f \curvearrowleft (\mathbb{R}^d \times \mathbb{R} \ni (x,w) \mapsto f^d(w) \in \mathbb{R})$, $g \curvearrowleft g^d$, $W \curvearrowleft W^{d,\theta}$ in the notation of [10, Theorem 1.1]), (166), the fact that $g$ is polynomially growing (cf. (153)–(154)), and the fact that $u^d$ is polynomially growing show that $u^d$ is the unique at most polynomially growing viscosity solution of

$$\frac{\partial u^d}{\partial t}(t,x) + \frac{1}{2}\operatorname{trace}(\sigma^d(\sigma^d(x))^\top(\operatorname{Hess}_x u^d(t,x))) + \langle \mu(x), (\nabla_x u^d)(t,x)\rangle + f(u^d(t,x)) = 0 \tag{182}$$

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

Next, (152)–(156) show for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $x, y \in \mathbb{R}^d$ that

$$|Tf_\varepsilon(0)| \leqslant (\varphi_d(x))^{\frac{\beta}{p}}, \quad |g_\varepsilon^d(x) - g_\varepsilon^d(y)| \leqslant \frac{(\varphi_d(x) + \varphi_d(y))^{\frac{\beta}{p}}}{\sqrt{T}}\|x-y\|, \quad |g_\varepsilon^d(x)| \leqslant (\varphi_d(x))^{\frac{\beta}{p}}, \tag{183}$$

$$|Tf(0)| \leqslant (\varphi_d(x))^{\frac{\beta}{p}}, \quad |g^d(x) - g^d(y)| \leqslant \frac{(\varphi_d(x) + \varphi_d(y))^{\frac{\beta}{p}}}{\sqrt{T}}\|x-y\|, \quad |g^d(x)| \leqslant (\varphi_d(x))^{\frac{\beta}{p}}, \tag{184}$$

$$\max\{|f_\varepsilon(v) - f(v)|, \|\mu_\varepsilon^d(x) - \mu^d(x)\|, \|\sigma_\varepsilon^d(x) - \sigma^d(x)\|, \|g_\varepsilon^d(x) - g^d(x)\|\} \leqslant \varepsilon((\varphi_d(x))^\beta + |v|^\beta). \tag{185}$$

This, (164), (165), (152), (166), and [17, Lemma 2.3] (applied for every $d \in \mathbb{N}$, $\varepsilon \in (0,1)$ with $d \curvearrowleft d$, $m \curvearrowleft d$, $\delta \curvearrowleft \varepsilon$, $\beta \curvearrowleft \beta$, $b \curvearrowleft 1$, $c \curvearrowleft c$, $q \curvearrowleft \beta$, $p \curvearrowleft p$, $\varphi \curvearrowleft \varphi_d$, $g_1 \curvearrowleft g_\varepsilon^d$, $\mu_1 \curvearrowleft \mu_\varepsilon^d$, $\sigma_1 \curvearrowleft \sigma_\varepsilon^d$, $f_1 \curvearrowleft ([0,T] \times \mathbb{R}^d \times \mathbb{R} \ni (t,x,w) \mapsto f_\varepsilon(w) \in \mathbb{R})$, $W \curvearrowleft W^{d,0}$, $(X_{t,s}^{x,1})_{x\in\mathbb{R}^d, t\in[0,T], s\in[t,T]} \curvearrowleft (X_s^{d,0,\varepsilon,t,x})_{x\in\mathbb{R}^d, t\in[0,T], s\in[t,T]}$, $g_2 \curvearrowleft g^d$, $\mu_2 \curvearrowleft \mu^d$, $\sigma_2 \curvearrowleft \sigma^d$, $f_2 \curvearrowleft ([0,T] \times \mathbb{R}^d \times \mathbb{R} \ni (t,x,w) \mapsto f(w) \in \mathbb{R})$, $(X_{t,s}^{x,2})_{x\in\mathbb{R}^d, t\in[0,T], s\in[t,T]} \curvearrowleft (X_s^{d,0,t,x})_{x\in\mathbb{R}^d, t\in[0,T], s\in[t,T]}$ in the notation of [17, Lemma 2.3]) show for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$|u^{d,\varepsilon}(t,x) - u^d(t,x)| \leqslant \varepsilon 2^{\beta+2} e^T e^{5\beta^2 c^4 + 2^\beta c^\beta T^\beta + 4c^2} (\varphi(x))^{\beta+0.5}. \tag{186}$$

Next, Lemma 2.4 (applied for all $d, K \in \mathbb{N}$, $\varepsilon \in (0,1)$ with $d \curvearrowleft d$, $K \curvearrowleft K$, $T \curvearrowleft T$, $\mathfrak{p} \curvearrowleft \mathfrak{p}$, $\beta \curvearrowleft \beta$, $b \curvearrowleft 1$, $c \curvearrowleft c$, $p \curvearrowleft p$, $\varphi \curvearrowleft \varphi_d$, $g \curvearrowleft g_\varepsilon^d$, $f \curvearrowleft f_\varepsilon$, $\mu \curvearrowleft \mu_\varepsilon^d$, $\sigma \curvearrowleft \sigma_\varepsilon^d$, $(\mathfrak{t}^\theta)_{\theta\in\Theta} \curvearrowleft (\mathfrak{t}^\theta)_{\theta\in\Theta}$, $(W^\theta)_{\theta\in\Theta} \curvearrowleft (W^{d,\theta})_{\theta\in\Theta}$, $(Y_s^{\theta,t,x})_{\theta\in\Theta, t\in[0,T], s\in[t,T], x\in\mathbb{R}^d} \curvearrowleft (X_s^{d,\theta,K,\varepsilon,t,x})_{\theta\in\Theta, t\in[0,T], s\in[t,T], x\in\mathbb{R}^d}$ in the notation of Lemma 2.4), (164), (183), (153), (155), (169), (170), and the independence and distributional properties imply that for all $d, K, m, n \in \mathbb{N}$, $\varepsilon \in (0,1)$, $t \in [0,T]$, $x \in \mathbb{R}^d$ we have that

$$\left\| U^{d,0,K,\varepsilon}_{n,m}(t,x) - u^{d,\varepsilon}(t,x) \right\|_{\mathfrak{p}} \leqslant 12c^2 e^{9c^3T} (\varphi_d(x))^{\frac{\beta+1}{p}} \left[ 2\mathfrak{p}^{\frac{n}{2}} e^{5cTn} e^{m^{\mathfrak{p}/2}/\mathfrak{p}} m^{-n/2} + \frac{1}{\sqrt{K}} \right]. \tag{187}$$

This, the triangle inequality, and (186) show for all $d, K, m, n \in \mathbb{N}$, $\varepsilon \in (0,1)$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} &\left\| U^{d,0,K,\varepsilon}_{n,m}(t,x) - u^{d}(t,x) \right\|_{\mathfrak{p}} \leqslant \left\| U^{d,0,K,\varepsilon}_{n,m}(t,x) - u^{d,\varepsilon}(t,x) \right\|_{\mathfrak{p}} + |u^{d,\varepsilon}(t,x) - u^d(t,x)| \\ &\leqslant 12^{\beta+2} c^2 e^{9c^3T+T+5\beta^2c^4+2^\beta c^\beta T^\beta+4c^2} (\varphi_d(x))^{\beta+0.5} \left[ 2\mathfrak{p}^{\frac{n}{2}} e^{5cTn} e^{m^{\mathfrak{p}/2}/\mathfrak{p}} m^{-n/2} + \frac{1}{\sqrt{K}} + \varepsilon \right]. \end{aligned} \tag{188}$$

Hence, (161) shows that there exists $\kappa \in (0,\infty)$ such that for all $d, K, n, m \in \mathbb{N}$, $\varepsilon \in (0,1)$ it holds that

$$\begin{aligned} &\left( \int_{[0,1]^d} \left\| U^{d,0,K,\varepsilon}_{n,m}(0,x) - u^d(0,x) \right\|^{\mathfrak{p}}_{\mathfrak{p}} \, dx \right)^{\frac{1}{\mathfrak{p}}} \\ &\leqslant 12^{\beta+2} c^2 e^{9c^3T+T+5\beta^2c^4+2^\beta c^\beta T^\beta+4c^2} \left( \int_{[0,1]^d} (\varphi_d(x))^{\mathfrak{p}(\beta+0.5)} \, dx \right)^{\frac{1}{\mathfrak{p}}} \left[ 2\mathfrak{p}^{\frac{n}{2}} e^{5cTn} e^{m^{\mathfrak{p}/2}/\mathfrak{p}} m^{-n/2} + \frac{1}{\sqrt{K}} + \varepsilon \right] \\ &\leqslant \kappa d^\kappa \left[ 2\mathfrak{p}^{\frac{n}{2}} e^{5cTn} e^{m^{\mathfrak{p}/2}/\mathfrak{p}} m^{-n/2} + \frac{1}{\sqrt{K}} + \varepsilon \right] \\ &\leqslant \kappa d^\kappa \left[ \left( \frac{2\mathfrak{p}^{\frac{1}{2}} e^{5cT} \exp(\frac{m^{\mathfrak{p}/2}}{n})}{m^{\frac{1}{2}}} \right)^n + \frac{1}{\sqrt{K}} + \varepsilon \right]. \end{aligned} \tag{189}$$

For the next step let $(M_n)_{n\in\mathbb{N}} \colon \mathbb{N} \to \mathbb{N}$ satisfy that $\liminf_{j\to\infty} M_j = \infty$, $\limsup_{j\to\infty} \frac{(M_j)^{\mathfrak{p}/2}}{j} < \infty$, and $\sup_{k\in\mathbb{N}} \frac{M_{k+1}}{M_k} < \infty$ (see, e.g., [43, Lemma 4.5] for an example). For every $\varepsilon \in (0,1)$, $d \in \mathbb{N}$ let

$$K_\varepsilon = \inf\{k \in \mathbb{N} \colon 1/\sqrt{k} \leqslant \varepsilon\}, \tag{190}$$

$$N_\varepsilon = \inf\left\{ n \in \mathbb{N} \colon \left( \frac{2\mathfrak{p}^{\frac{1}{2}} e^{5cT} \exp(\frac{(M_n)^{\mathfrak{p}/2}}{n})}{(M_n)^{\frac{1}{2}}} \right)^n \leqslant \varepsilon \right\}, \tag{191}$$

$$L_{d,\varepsilon} = K_\varepsilon(\max\{\dim(\mathcal{D}(\Phi_{\mu^d_\varepsilon})), \dim(\mathcal{D}(\Phi_{\sigma^d_\varepsilon,0}))\} - 2) + 2, \tag{192}$$

$$c_{d,\varepsilon} = 3\max\left\{ d\mathfrak{d}, |||\mathcal{D}(\Phi_{f_\varepsilon})|||, |||\mathcal{D}(\Phi_{g^d_\varepsilon})|||, |||\mathcal{D}(\Phi_{\mu^d_\varepsilon})|||, |||\mathcal{D}(\Phi_{\sigma^d_\varepsilon,0})||| \right\}. \tag{193}$$

For every $\epsilon \in (0,1)$, $d \in \mathbb{N}$ let

$$\varepsilon(d,\epsilon) = \frac{\epsilon}{3\kappa d^\kappa} \tag{194}$$

For every $\delta \in (0,1)$ let

$$C_\delta = \sup_{\varepsilon\in(0,1)} \left[ \varepsilon^{4+\delta} N_\varepsilon (3M_{N_\varepsilon})^{2N_\varepsilon} \right]. \tag{195}$$

Then [2, Lemma 5.1] (applied with $L \curvearrowleft 1$, $T \curvearrowleft 2\mathfrak{p}^{\frac{1}{2}} e^{5cT} - 1$, $(m_k)_{k\in\mathbb{N}} \curvearrowleft (M_k)_{k\in\mathbb{N}}$ in the notation of [2, Lemma 5.1]) show for all $\delta \in (0,1)$ that $C_\delta < \infty$. Furthermore, (190) show for all $\varepsilon \in (0,1)$ that $\frac{1}{\sqrt{K_\varepsilon - 1}} > \varepsilon$, i.e.,

$$K_\varepsilon = K_\varepsilon - 1 + 1 < \varepsilon^{-2} + 1 < 2\varepsilon^{-2}. \tag{196}$$

Next, Tonelli's theorem, (189)–(191), and (194) show for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ that

$$\begin{aligned}
&\mathbb{E}\biggl[\int_{[0,1]^d} \Bigl|U^{d,0,K_{\varepsilon(d,\epsilon)},\varepsilon(d,\epsilon)}_{N_{\varepsilon(d,\epsilon)},M_{N_{\varepsilon(d,\epsilon)}}}(t,x) - u^d(t,x)\Bigr|^{\mathfrak{p}}\,dx\biggr] \\
&= \int_{[0,1]^d} \mathbb{E}\biggl[\Bigl|U^{d,0,K_{\varepsilon(d,\epsilon)},\varepsilon(d,\epsilon)}_{N_{\varepsilon(d,\epsilon)},M_{N_{\varepsilon(d,\epsilon)}}}(t,x) - u^d(t,x)\Bigr|^{\mathfrak{p}}\biggr]\,dx \\
&\leqslant \left(\kappa d^\kappa \left[\left(\frac{2\mathfrak{p}^{\frac{1}{2}} e^{5cT} \exp\left(\frac{(M_{N_{\varepsilon(d,\epsilon)}})^{\mathfrak{p}/2}}{N_{\varepsilon(d,\epsilon)}}\right)}{(M_{N_{\varepsilon(d,\epsilon)}})^{\frac{1}{2}}}\right)^n + \frac{1}{\sqrt{K_{\varepsilon(d,\epsilon)}}} + \varepsilon(d,\epsilon)\right]\right)^{\mathfrak{p}} \\
&\leqslant \left(\kappa d^\kappa \left[\varepsilon(d,\epsilon) + \varepsilon(d,\epsilon) + \varepsilon(d,\epsilon)\right]\right)^{\mathfrak{p}} = \epsilon^{\mathfrak{p}}.
\end{aligned} \tag{197}$$

Therefore, for every $d \in \mathbb{N}$, $\epsilon \in (0,1)$ there exists $\omega(d,\epsilon) \in \Omega$ such that

$$\int_{[0,1]^d} \Bigl|U^{d,0,K_{\varepsilon(d,\epsilon)},\varepsilon(d,\epsilon)}_{N_{\varepsilon(d,\epsilon)},M_{N_{\varepsilon(d,\epsilon)}}}(t,x,\omega(d,\epsilon)) - u^d(t,x)\Bigr|^{\mathfrak{p}}\,dx < \epsilon^{\mathfrak{p}}. \tag{198}$$

Furthermore, Lemma 3.15 and (192) show that there exist $(\Phi^\omega_{d,\varepsilon})_{d\in\mathbb{N},\varepsilon\in(0,1),\omega\in\Omega} \subseteq \mathbf{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $\omega \in \Omega$ it holds that

$$\dim(\mathcal{D}(\Phi_{d,\varepsilon})) = N_\varepsilon\left(\dim(\mathcal{D}(\Phi_{f_\varepsilon})) + L_{d,\varepsilon} - 4\right) + \dim\bigl(\mathcal{D}(\Phi_{g^d_\varepsilon})\bigr) + L_{d,\varepsilon} - 2, \tag{199}$$

$$\left|\!\left|\!\left|\mathcal{D}(\Phi^\omega_{d,\varepsilon})\right|\!\right|\!\right| \leqslant c_{d,\varepsilon}(3M_{N_\varepsilon})^{N_\varepsilon}, \quad U^{d,0,K_\varepsilon,\varepsilon}_{N_\varepsilon,M_{N_\varepsilon}}(0,x,\omega) = (\mathcal{R}(\Phi^\omega_{d,\varepsilon}))(x). \tag{200}$$

Next, (192) and (158) show for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$ that $L_{d,\varepsilon} \leqslant K_\varepsilon c d^c \varepsilon^{-c}$. This, (199), and (158) show for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $\omega \in \Omega$ that

$$\begin{aligned}
\dim(\mathcal{D}(\Phi^\omega_{d,\varepsilon})) &\leqslant 4N_\varepsilon L_{d,\varepsilon}\max\{\dim(\mathcal{D}(\Phi_{f_\varepsilon})), \dim(\mathcal{D}(\Phi_{g^d_\varepsilon}))\} \\
&\leqslant 4N_\varepsilon K_\varepsilon c d^c \varepsilon^{-c} c d^c \varepsilon^{-c} = 4N_\varepsilon K_\varepsilon c^2 d^{2c}\varepsilon^{-2c}
\end{aligned} \tag{201}$$

Next, (193) and the fact that $c \geqslant 3\mathfrak{d}$ show for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$ that

$$c_{d,\varepsilon} = 3\max\bigl\{d\mathfrak{d}, |\!|\!|\mathcal{D}(\Phi_{f_\varepsilon})|\!|\!|, |\!|\!|\mathcal{D}(\Phi_{g^d_\varepsilon})|\!|\!|, |\!|\!|\mathcal{D}(\Phi_{\mu^d_\varepsilon})|\!|\!|, |\!|\!|\mathcal{D}(\Phi_{\sigma^d_\varepsilon,0})|\!|\!|\bigr\} \leqslant c d^c \varepsilon^{-c}. \tag{202}$$

This and (200) show for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $\omega \in \Omega$ that

$$\left|\!\left|\!\left|\mathcal{D}(\Phi^\omega_{d,\varepsilon})\right|\!\right|\!\right| \leqslant c_{d,\varepsilon}(3M_{N_\varepsilon})^{N_\varepsilon} \leqslant c d^c \varepsilon^{-c}(3M_{N_\varepsilon})^{N_\varepsilon}. \tag{203}$$

This, the fact that $\forall\, \Phi \in \mathbf{N}\colon \mathcal{P}(\Phi) \leqslant 2\dim(\mathcal{D}(\Phi))|\!|\!|\mathcal{D}(\Phi)|\!|\!|^2$, (201), (196), (195), and the fact that $\forall\, \delta \in (0,1)\colon C_\delta < \infty$ show for all $d \in \mathbb{N}$, $\varepsilon, \delta \in (0,1)$ that

$$\begin{aligned}
\mathcal{P}(\Phi^\omega_{d,\varepsilon}) &\leqslant 2\cdot 4N_\varepsilon K_\varepsilon c^2 d^{2c}\varepsilon^{-2c}\left(c d^c \varepsilon^{-c}(3M_{N_\varepsilon})^{N_\varepsilon}\right)^2 \\
&= 8K_\varepsilon c^4 d^{4c}\varepsilon^{-4c} N_\varepsilon (3M_{N_\varepsilon})^{2N_\varepsilon} \\
&= 8K_\varepsilon c^4 d^{4c}\varepsilon^{-4c}\varepsilon^{4+\delta} N_\varepsilon (3M_{N_\varepsilon})^{2N_\varepsilon}\varepsilon^{-(4+\delta)} \\
&\leqslant 8\cdot 2\varepsilon^{-2} c^4 d^{4c}\varepsilon^{-4c} C_\delta \varepsilon^{-(4+\delta)} \\
&= 16 C_\delta c^4 d^{4c}\varepsilon^{-(6+\delta)-4c} < \infty.
\end{aligned} \tag{204}$$

Next, for every $d \in \mathbb{N}$, $\epsilon \in (0,1)$ let

$$\Psi_{d,\epsilon} = \Phi^{\omega(d,\epsilon)}_{d,\varepsilon(d,\epsilon)}. \tag{205}$$

Then (204) and (194) show for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ that

$$\begin{aligned}
\mathcal{P}(\Psi_{d,\epsilon}) = \mathcal{P}(\Phi^{\omega(d,\epsilon)}_{d,\varepsilon(d,\epsilon)}) &\leqslant 16 C_\delta c^4 d^{4c}(\varepsilon(d,\epsilon))^{-(6+\delta)-4c} \\
&= 16 C_\delta c^4 d^{4c}\left(\frac{\epsilon}{3\kappa d^\kappa}\right)^{-(6+\delta)-4c} \\
&= 16 C_\delta (3\kappa)^{(6+\delta)+4c} c^4 d^{4c+\kappa((6+\delta)+4c)}\epsilon^{-(6+\delta)-4c}.
\end{aligned} \tag{206}$$

Furthermore, (200), (205), and (198) show for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ that $U^{d,0,K_{\varepsilon(d,\epsilon)},\varepsilon(d,\epsilon)}_{N_{\varepsilon(d,\epsilon)},M_{N_{\varepsilon(d,\epsilon)}}}(0,x,\omega(d,\epsilon)) = (\mathcal{R}(\Phi^{\omega(d,\epsilon)}_{d,\varepsilon(d,\epsilon)}))(x) = (\mathcal{R}(\Psi_{d,\epsilon}))(x)$ and

$$\int_{[0,1]^d} \left|(\mathcal{R}(\Psi_{d,\epsilon}))(x) - u^d(0,x)\right|^{\mathfrak{p}} dx < \epsilon^{\mathfrak{p}}. \tag{207}$$

This, (206), and the fact that $\forall\, \delta \in (0,1)\colon C_\delta < \infty$ complete the proof of Theorem 4.1. □

*Proof of Theorem 1.3.* The definitions of $\mathfrak{a}_0, \mathfrak{a}_1$, the fact that $a \in \{\mathfrak{a}_0, \mathfrak{a}_1\}$, and Lemmas 3.1 and 3.2 show that there exists $\mathfrak{d} \in \mathbb{N}$, $\mathfrak{n}_{1,\mathfrak{d}} = (1,\mathfrak{d},1) \in \mathbf{D}$ such that

$$\mathrm{Id}_{\mathbb{R}} \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = \mathfrak{n}_{1,\mathfrak{d}}\}). \tag{208}$$

Next, the definitions of $\mathfrak{a}_0, \mathfrak{a}_1$, the fact that $a \in \{\mathfrak{a}_0, \mathfrak{a}_1\}$, and Lemmas 3.4 and 3.5 show that there exists $\tilde{c} \in (0,\infty)$, $(f_\varepsilon)_{\varepsilon\in(0,1)} \subseteq C(\mathbb{R},\mathbb{R})$ such that for all $\varepsilon \in (0,1)$, $x, y \in \mathbb{R}$ we have that

$$|f_\varepsilon(x) - f_\varepsilon(y)| \leqslant c|x-y|, \quad |f_\varepsilon(x) - f(x)| \leqslant \varepsilon(1+|x|^\beta), \tag{209}$$

$$f_\varepsilon \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon\ \dim(\mathcal{D}(\Phi)) = 3, \|\!|\mathcal{D}(\Phi)|\!\| \leqslant \tilde{c}\varepsilon^{-\tilde{c}}\}). \tag{210}$$

Then (15) shows for all $\varepsilon \in (0,1)$ that

$$T|f_\varepsilon(0)| \leqslant T(|f(0)| + \varepsilon) \leqslant c + T. \tag{211}$$

Furthermore, (17) shows that there exists $\bar{c} \in (0,\infty)$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$ that

$$\max\left\{\|\!|\mathcal{D}(\Phi_{f_\varepsilon})|\!\|, \|\!|\mathcal{D}(\Phi_{g^d_\varepsilon})|\!\|, \|\!|\mathcal{D}(\Phi_{\mu^d_\varepsilon})|\!\|, \|\!|\mathcal{D}(\Phi_{\sigma^d_\varepsilon,0})|\!\|\right\} \leqslant \bar{c}d^{\bar{c}}\varepsilon^{-\bar{c}}, \tag{212}$$

$$\max\left\{\dim(\mathcal{D}(\Phi_{f_\varepsilon})), \dim(\mathcal{D}(\Phi_{g^d_\varepsilon})), \dim(\mathcal{D}(\Phi_{\mu^d_\varepsilon})), \dim(\mathcal{D}(\Phi_{\sigma^d_\varepsilon,0}))\right\} \leqslant \bar{c}d^{\bar{c}}\varepsilon^{-\bar{c}}. \tag{213}$$

Combining (208)–(213), the assumptions of Theorem 1.3, and Theorem 4.1 (applied with $c$ replaced by a suitable large constant) we complete the proof of Theorem 1.3. □

## References


[1] Ackermann, J., Jentzen, A., Kruse, T., Kuckuck, B., and Padgett, J. L. Deep neural networks with ReLU, leaky ReLU, and softplus activation provably overcome the curse of dimensionality for Kolmogorov partial differential equations with Lipschitz nonlinearities in the $L^p$-sense. *arXiv:2309.13722* (2023).

[2] Ackermann, J., Jentzen, A., Kuckuck, B., and Padgett, J. L. Deep neural networks with ReLU, leaky ReLU, and softplus activation provably overcome the curse of dimensionality for space-time solutions of semilinear partial differential equations. *arXiv:2406.10876* (2024).

[3] Al-Aradi, A., Correia, A., Jardim, G., de Freitas Naiff, D., and Saporito, Y. Extensions of the deep Galerkin method. *Applied Mathematics and Computation 430* (2022), 127287.

[4] Beck, C., Becker, S., Cheridito, P., Jentzen, A., and Neufeld, A. Deep learning based numerical approximation algorithms for stochastic partial differential equations and high-dimensional nonlinear filtering problems. *arXiv preprint arXiv:2012.01194* (2020).

[5] Beck, C., Becker, S., Cheridito, P., Jentzen, A., and Neufeld, A. Deep splitting method for parabolic PDEs. *SIAM Journal on Scientific Computing 43*, 5 (2021), A3135–A3154.

[6] Beck, C., E, W., and Jentzen, A. Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *Journal of Nonlinear Science 29* (2019), 1563–1619.

[7] Beck, C., Gonon, L., Hutzenthaler, M., and Jentzen, A. On existence and uniqueness properties for solutions of stochastic fixed point equations. *Discrete and Continuous Dynamical Systems - Series B 26*, 9 (2019), 4927–4962. Discrete Contin. Dyn. Syst. Ser. B. (in press).

[8] Beck, C., Gonon, L., and Jentzen, A. Overcoming the curse of dimensionality in the numerical approximation of high-dimensional semilinear elliptic partial differential equations. *arXiv preprint arXiv:2003.00596* (2020).

[9] Beck, C., Hornung, F., Hutzenthaler, M., Jentzen, A., and Kruse, T. Overcoming the curse of dimensionality in the numerical approximation of Allen–Cahn partial differential equations via truncated full-history recursive multilevel Picard approximations. *Journal of Numerical Mathematics 28*, 4 (2020), 197–222.

[10] Beck, C., Hutzenthaler, M., and Jentzen, A. On nonlinear Feynman–Kac formulas for viscosity solutions of semilinear parabolic partial differential equations. *Stochastics and Dynamics 21*, 08 (2021), 2150048.

[11] Beck, C., Hutzenthaler, M., Jentzen, A., and Kuckuck, B. An overview on deep learning-based approximation methods for partial differential equations. *arXiv preprint arXiv:2012.12348* (2020).

[12] Becker, S., Braunwarth, R., Hutzenthaler, M., Jentzen, A., and von Wurstemberger, P. Numerical simulations for full history recursive multilevel Picard approximations for systems of high-dimensional partial differential equations. *arXiv preprint arXiv:2005.10206* (2020).

[13] Berner, J., Dablander, M., and Grohs, P. Numerically solving parametric families of high-dimensional Kolmogorov partial differential equations via deep learning. *Advances in Neural Information Processing Systems 33* (2020), 16615–16627.

[14] Berner, J., Grohs, P., and Jentzen, A. Analysis of the Generalization Error: Empirical Risk Minimization over Deep Artificial Neural Networks Overcomes the Curse of Dimensionality in the Numerical Approximation of Black–Scholes Partial Differential Equations. *SIAM Journal on Mathematics of Data Science 2*, 3 (2020), 631–657.

[15] Castro, J. Deep learning schemes for parabolic nonlocal integro-differential equations. *Partial Differential Equations and Applications 3*, 6 (2022), 77.

[16] Cheridito, P., and Rossmannek, F. Efficient Sobolev approximation of linear parabolic PDEs in high dimensions. *arXiv:2306.16811* (2023).

[17] Cioica-Licht, P. A., Hutzenthaler, M., and Werner, P. T. Deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear partial differential equations. *arXiv:2205.14398v1* (2022).

[18] Cox, S. G., Hutzenthaler, M., and Jentzen, A. Local Lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations. *arXiv:1309.5595* (2014). Revision requested from Mem. Amer. Math. Soc.

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

[20] E, W., Han, J., and Jentzen, A. Algorithms for solving high dimensional PDEs: from nonlinear Monte Carlo to machine learning. *Nonlinearity 35*, 1 (2021), 278.

[21] E, W., and Yu, B. The deep Ritz method: A deep learning-based numerical algorithm for solving variational problems. *Commun Math Stat 6*, 1 (2018), 1–12.

[22] Elbrächter, D., Grohs, P., Jentzen, A., and Schwab, C. DNN Expression Rate Analysis of High-dimensional PDEs: Application to Option Pricing. *Constructive Approximation 55* (2022), 3–71.

[23] Frey, R., and Köck, V. Convergence analysis of the deep splitting scheme: the case of partial integro-differential equations and the associated FBSDEs with jumps. *arXiv preprint arXiv:2206.01597* (2022).

[24] Frey, R., and Köck, V. Deep neural network algorithms for parabolic PIDEs and applications in insurance mathematics. In *Methods and Applications in Fluorescence*. Springer, 2022, pp. 272–277.

[25] Germain, M., Pham, H., and Warin, X. Approximation error analysis of some deep backward schemes for nonlinear PDEs. *SIAM Journal on Scientific Computing 44*, 1 (2022), A28–A56.

[26] Giles, M. B., Jentzen, A., and Welti, T. Generalised multilevel picard approximations. *arXiv preprint arXiv:1911.03188* (2019).

[27] Gnoatto, A., Patacca, M., and Picarelli, A. A deep solver for BSDEs with jumps. *arXiv preprint arXiv:2211.04349* (2022).

[28] Gonon, L. Random feature neural networks learn Black-Scholes type PDEs without curse of dimensionality. *Journal of Machine Learning Research 24*, 189 (2023), 1–51.

[29] Gonon, L., and Schwab, C. Deep ReLU network expression rates for option prices in high-dimensional, exponential Lévy models. *Finance and Stochastics 25*, 4 (2021), 615–657.

[30] Gonon, L., and Schwab, C. Deep ReLU neural networks overcome the curse of dimensionality for partial integrodifferential equations. *Analysis and Applications 21*, 01 (2023), 1–47.

[31] Grohs, P., Hornung, F., Jentzen, A., and Von Wurstemberger, P. A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black–Scholes partial differential equations. *Memoirs of the American Mathematical Society 284* (2023).

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

[33] Han, J., and Long, J. Convergence of the deep BSDE method for coupled FBSDEs. *Probability, Uncertainty and Quantitative Risk 5* (2020), 1–33.

[34] Han, J., Zhang, L., and E, W. Solving many-electron Schrödinger equation using deep neural networks. *Journal of Computational Physics 399* (2019), 108929.

[35] Huré, C., Pham, H., and Warin, X. Deep backward schemes for high-dimensional nonlinear PDEs. *Mathematics of Computation 89*, 324 (2020), 1547–1579.

[36] Hutzenthaler, M., and Jentzen, A. On a perturbation theory and on strong convergence rates for stochastic ordinary and partial differential equations with non-globally monotone coefficients. *Annals of Probability 48*, 1 (2020), 53–93.

[37] Hutzenthaler, M., Jentzen, A., and Kruse, T. On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations. *Journal of Scientific Computing 79*, 3 (2019), 1534–1571.

[38] Hutzenthaler, M., Jentzen, A., and Kruse, T. Multilevel Picard iterations for solving smooth semilinear parabolic heat equations. *Partial Differential Equations and Applications 2*, 6 (2021), 1–31.

[39] Hutzenthaler, M., Jentzen, A., and Kruse, T. Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities. *Foundations of Computational Mathematics 22* (2022), 905–966.

[40] Hutzenthaler, M., Jentzen, A., Kruse, T., Nguyen, T., and von Wurstemberger, P. Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations. *Proceedings of the Royal Society A 476*, 2244 (2020), 20190630.

[41] Hutzenthaler, M., Jentzen, A., Kruse, T., and Nguyen, T. A. Multilevel Picard approximations for high-dimensional semilinear second-order PDEs with Lipschitz nonlinearities. *arXiv:2009.02484* (2020).

[42] Hutzenthaler, M., Jentzen, A., Kruse, T., and Nguyen, T. A. A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations. *SN partial differential equations and applications 1* (2020), 1–34.

[43] Hutzenthaler, M., Jentzen, A., Kuckuck, B., and Padgett, J. L. Strong $L^p$-error analysis of nonlinear Monte Carlo approximations for high-dimensional semilinear partial differential equations. *arXiv:2110.08297* (2021).

[44] Hutzenthaler, M., Jentzen, A., and von Wurstemberger, P. Overcoming the curse of dimensionality in the approximative pricing of financial derivatives with default risks. *Electronic Journal of Probability 25* (2020), 1–73.

[45] Hutzenthaler, M., and Kruse, T. Multilevel Picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities. *SIAM Journal on Numerical Analysis 58*, 2 (2020), 929–961.

[46] Hutzenthaler, M., Kruse, T., and Nguyen, T. A. Multilevel Picard approximations for McKean-Vlasov stochastic differential equations. *Journal of Mathematical Analysis and Applications 507*, 1 (2022), 125761.

[47] Hutzenthaler, M., and Nguyen, T. A. Multilevel Picard approximations of high-dimensional semilinear partial differential equations with locally monotone coefficient functions. *Applied Numerical Mathematics 181* (2022), 151–175.

[48] Ito, K., Reisinger, C., and Zhang, Y. A neural network-based policy iteration algorithm with global $H^2$-superlinear convergence for stochastic games on domains. *Foundations of Computational Mathematics 21*, 2 (2021), 331–374.

[49] Jacquier, A., and Oumgari, M. Deep curve-dependent PDEs for affine rough volatility. *SIAM Journal on Financial Mathematics 14*, 2 (2023), 353–382.

[50] Jacquier, A., and Zuric, Z. Random neural networks for rough volatility. *arXiv preprint arXiv:2305.01035* (2023).

[51] Jentzen, A., Kuckuck, B., and von Wurstemberger, P. Mathematical introduction to deep learning: Methods, implementations, and theory. *arXiv:2310.20360v1* (2023).

[52] Jentzen, A., Salimova, D., and Welti, T. A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. *Communications in Mathematical Sciences 19*, 5 (2021), 1167–1205.

[53] Lu, L., Meng, X., Mao, Z., and Karniadakis, G. E. DeepXDE: A deep learning library for solving differential equations. *SIAM review 63*, 1 (2021), 208–228.

[54] Neufeld, A., and Nguyen, T. A. Rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of gradient-dependent semilinear heat equations. *arXiv preprint arXiv:2403.09200* (2024).

[55] Neufeld, A., and Nguyen, T. A. Rectified deep neural networks overcome the curse of dimensionality when approximating solutions of McKean–Vlasov stochastic differential equations. *Journal of Mathematical Analysis and Applications 541*, 1 (2025), 128661.

[56] Neufeld, A., Nguyen, T. A., and Wu, S. Deep ReLU neural networks overcome the curse of dimensionality in the numerical approximation of semilinear partial integro-differential equations. *arXiv preprint arXiv:2310.15581* (2023).

[57] Neufeld, A., Nguyen, T. A., and Wu, S. Multilevel Picard approximations overcome the curse of dimensionality in the numerical approximation of general semilinear PDEs with gradient-dependent nonlinearities. *arXiv:2311.11579* (2023).

[58] Neufeld, A., and Schmocker, P. Universal approximation property of random neural networks. *arXiv:2312.08410* (2023).

[59] Neufeld, A., Schmocker, P., and Wu, S. Full error analysis of the random deep splitting method for nonlinear parabolic PDEs and PIDEs. *arXiv:2405.05192* (2024).

[60] Neufeld, A., and Wu, S. Multilevel Picard approximation algorithm for semilinear partial integro-differential equations and its complexity analysis. *arXiv:2205.09639v3* (2022).

[61] Neufeld, A., and Wu, S. Multilevel Picard algorithm for general semilinear parabolic PDEs with gradient-dependent nonlinearities. *arXiv:2310.12545* (2023).

[62] Nguwi, J. Y., Penent, G., and Privault, N. A deep branching solver for fully nonlinear partial differential equations. *arXiv preprint arXiv:2203.03234* (2022).

[63] Nguwi, J. Y., Penent, G., and Privault, N. Numerical solution of the incompressible Navier-Stokes equation by a deep branching algorithm. *arXiv preprint arXiv:2212.13010* (2022).

[64] Nguwi, J. Y., and Privault, N. A deep learning approach to the probabilistic numerical solution of path-dependent partial differential equations. *Partial Differential Equations and Applications 4*, 4 (2023), 37.

[65] Raissi, M., Perdikaris, P., and Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics 378* (2019), 686–707.

[66] Reisinger, C., and Zhang, Y. Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems. *Analysis and Applications 18*, 06 (2020), 951–999.

[67] Rio, E. Moment inequalities for sums of dependent random variables under projective conditions. *Journal of Theoretical Probability 22* (2009), 146–163.

[68] Sirignano, J., and Spiliopoulos, K. DGM: A deep learning algorithm for solving partial differential equations. *Journal of computational physics 375* (2018), 1339–1364.

[69] Zhang, D., Guo, L., and Karniadakis, G. E. Learning in modal space: Solving time-dependent stochastic PDEs using physics-informed neural networks. *SIAM Journal on Scientific Computing 42*, 2 (2020), A639–A665.

[1] Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
*Email address*: `ariel.neufeld@ntu.edu.sg`

[2] Faculty of Mathematics, Bielefeld University, Germany
*Email address*: `tnguyen@math.uni-bielefeld.de`