arXiv:2403.09200v1 [math.NA] 14 Mar 2024

RECTIFIED DEEP NEURAL NETWORKS OVERCOME THE CURSE OF DIMENSIONALITY IN THE NUMERICAL APPROXIMATION OF GRADIENT-DEPENDENT SEMILINEAR HEAT EQUATIONS

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

ABSTRACT. Numerical experiments indicate that deep learning algorithms overcome the curse of dimensionality when approximating solutions of semilinear PDEs. For certain linear PDEs and semilinear PDEs with *gradient-independent* nonlinearities this has also been proved mathematically, i.e., it has been shown that the number of parameters of the approximating DNN increases at most polynomially in both the PDE dimension $d \in \mathbb{N}$ and the reciprocal of the prescribed accuracy $\epsilon \in (0,1)$.
The main contribution of this paper is to rigorously prove for the first time that deep neural networks can also overcome the curse dimensionality in the approximation of a certain class of nonlinear PDEs with *gradient-dependent* nonlinearities.

1. INTRODUCTION

Deep learning-based approximation algorithms for certain nonlinear parabolic partial differential equations (PDEs) have been first proposed in [13, 25]. Due to their success in approximately solving nonlinear PDE in higher dimensions where classical methods have failed, there is now a variety of deep learning algorithms for different kinds of PDEs in the scientific literature, see, e.g., [2, 3, 4, 5, 6, 7, 8, 10, 14, 15, 17, 18, 19, 20, 25, 26, 27, 28, 29, 33, 34, 35, 37, 38, 39, 40, 44, 45, 46, 47, 48, 49, 50, 51].

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: [9, 16, 21, 22, 23, 24, 36] prove that deep neural network (DNN) approximations overcome the curse of dimensionality when approximating solutions of linear PDEs, [1, 31] prove that DNN approximations overcome the curse of dimensionality when approximating solutions of semilinear heat equations, and [11, 41] prove that DNN approximations overcome the curse of dimensionality when approximating solutions of general semilinear PDEs.

However, we highlight that all the above results [1, 11, 31, 41] only deal with *gradient-independent* semilinear PDEs, i.e., the case when the nonlinear part of the corresponding semilinear PDE does not depend on the gradient of the solution.

The contribution of our article is to prove for the first time that deep neural networks can also overcome the curse dimensionality in the approximation of nonlinear PDEs with *gradient-dependent* nonlinearities. Our main result, Theorem 1.2, proves that for semilinear heat equations with *gradient-dependent* nonlinear part the number of parameters of the approximating DNN increases at most polynomially in both the PDE dimension $d \in \mathbb{N}$ and the reciprocal of the prescribed accuracy $\epsilon \in (0,1)$, i.e., DNNs do overcome the curse of dimensionality when approximating such PDEs.

1.1. **Notations.**

Throughout our paper we use the following notations. Let $\|\cdot\|, |||\cdot|||\colon (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to [0,\infty)$, $\dim\colon (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to \mathbb{N}$ satisfy for all $d \in \mathbb{N}$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ that $\|x\| = \sqrt{\sum_{i=1}^d (x_i)^2}$, $|||x||| = \max_{i\in[1,d]\cap\mathbb{N}} |x_i|$, and $\dim(x) = d$. Moreover, given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable $\mathfrak{X}\colon \Omega \to \mathbb{R}$, and $s \in [1,\infty)$ let $\|\mathfrak{X}\|_s \in [0,\infty]$ satisfy that $\|\mathfrak{X}\|_s = (\mathbb{E}[|\mathfrak{X}|^s])^{\frac{1}{s}}$.

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

Key words and phrases. curse of dimensionality, high-dimensional PDEs, deep neural networks, information based complexity, tractability of multivariate problems, multilevel Picard approximations, gradient-dependent nonlinearity.

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

1.2. **A mathematical framework for DNNs.** In order to formulate our main result, Theorem 1.2, we need to introduce a mathematical framework for DNNs, see Setting 1.1 below.

Setting 1.1 (A mathematical framework for DNNs)**.** *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*

$$\mathbf{A}_d(x) = (\max\{x_1, 0\}, \max\{x_2, 0\}, \ldots, \max\{x_d, 0\}). \tag{1}$$

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

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

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*

$$\mathcal{P}(\Phi) = \sum_{n=1}^{H+1} k_n(k_{n-1}+1), \quad \mathcal{D}(\Phi) = (k_0, k_1, \ldots, k_H, k_{H+1}), \tag{3}$$

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

$$(\mathcal{R}(\Phi))(x_0) = W_{H+1}x_H + B_{H+1}. \tag{4}$$

Let us comment on the mathematical objects in Setting 1.1. For all $d \in \mathbb{N}$, $\mathbf{A}_d\colon \mathbb{R}^d \to \mathbb{R}^d$ refers to the componentwise rectified linear unit (ReLU) activation function. By $\mathbf{N}$ we denote the set of all (parameters characterizing) artificial feed-forward DNNs, by $\mathcal{R}$ we denote the operator that maps each DNN to its corresponding function, by $\mathcal{P}$ we denote the function that maps a DNN to its number of parameters, and by $\mathcal{D}$ we denote the function that maps a DNN to the vector of its layer dimensions.

We are now in a position to state the main result, Theorem 1.2 below.

Theorem 1.2. *Assume Setting 1.1. Let* $T \in (0, T)$*,* $\beta, c \in [2, \infty)$*,* $q \in [1, 2)$*. For every* $d \in \mathbb{N}$ *let* $L_i^d \in \mathbb{R}$*,* $i \in [0, d] \cap \mathbb{Z}$*, satisfy that* $\sum_{i=0}^d L_i^d \leqslant c$*. For every* $d \in \mathbb{N}$ *let* $\Lambda^d = (\Lambda_\nu^d)_{\nu\in[0,d]\cap\mathbb{Z}}\colon [0, T] \to \mathbb{R}^{1+d}$ *satisfy for all* $t \in [0, T]$ *that* $\Lambda^d(t) = (1, \sqrt{t}, \ldots, \sqrt{t})$*. For every* $d \in \mathbb{N}$ *let* $\mathrm{pr}^d = (\mathrm{pr}_\nu^d)_{\nu\in[0,d]\cap\mathbb{Z}}\colon \mathbb{R}^{d+1} \to \mathbb{R}$ *satisfy for all* $w = (w_\nu)_{\nu\in[0,d]\cap\mathbb{Z}}$*,* $i \in [0, d] \cap \mathbb{Z}$ *that* $\mathrm{pr}_i^d(w) = w_i$*. For every* $d \in \mathbb{N}$*,* $\varepsilon \in (0, 1)$ *let* $g^d, g_\varepsilon^d \in C(\mathbb{R}^d, \mathbb{R})$*,* $f^d, f_\varepsilon^d \in C(\mathbb{R}^d, \mathbb{R}^{d+1})$*,* $\Phi_{g_\varepsilon^d}, \Phi_{f_\varepsilon^d} \in \mathbf{N}$ *satisfy that* $\mathcal{R}(\Phi_{g_\varepsilon^d}) = g_\varepsilon^d$ *and* $\mathcal{R}(\Phi_{f_\varepsilon^d}) = f_\varepsilon^d$*. Assume for all* $d \in \mathbb{N}$*,* $\varepsilon \in (0, 1)$*,* $x, y \in \mathbb{R}^d$*,* $w, w_1, w_2 \in \mathbb{R}^{d+1}$ *that*

$$|g_\varepsilon^d(x)| + |T f_\varepsilon^d(x)| \leqslant c(d^c + \|x\|^2)^{\frac{1}{2}}, \tag{5}$$

$$|f_\varepsilon^d(w_1) - f_\varepsilon^d(w_2)| \leqslant \sum_{\nu=0}^d \left[L_\nu^d \Lambda_\nu^d(T) |\mathrm{pr}_\nu^d(w_1 - w_2)|\right], \tag{6}$$

$$|g_\varepsilon^d(x) - g_\varepsilon^d(y)| \leqslant cd^c \frac{\|x-y\|}{\sqrt{T}}, \tag{7}$$

$$\left|f_\varepsilon^d(w) - f^d(w)\right| \leqslant \frac{\varepsilon c d^c}{T}\left(1 + \sum_{\nu=0}^d (\Lambda_\nu^d(T))^q (\mathrm{pr}_\nu^d(w))^q\right), \tag{8}$$

$$\left|g^d(x) - g_\varepsilon^d(x)\right| \leqslant \varepsilon c d^c (d^c + \|x\|^2)^\beta, \tag{9}$$

$$\max\left\{\mathcal{P}(\Phi_{g_\varepsilon^d}), \mathcal{P}(\Phi_{f_\varepsilon^d})\right\} \leqslant c d^c \varepsilon^{-c}. \tag{10}$$

Then the following items hold.

(i) *For all* $d \in \mathbb{N}$ *there exists a unique continuous function* $u^d\colon [0,T) \times \mathbb{R}^d \to \mathbb{R}^{d+1}$ *such that* $v^d := \mathrm{pr}_0^d(u^d)$ *is the unique viscosity solution to the following semilinear PDE of parabolic type:*

$$\frac{\partial v^d}{\partial t}(t,x) + \frac{1}{2}(\Delta v^d)(t,x) + f^d(t,x,v^d(t,x),(\nabla_x v^d)(t,x)) = 0 \quad \forall\, t \in (0,T), x \in \mathbb{R}^d, \tag{11}$$

$$v^d(T,x) = g^d(x) \quad \forall\, x \in \mathbb{R}^d, \tag{12}$$

$\nabla_x v^d = (\mathrm{pr}_1^d(u^d), \mathrm{pr}_2^d(u^d), \ldots, \mathrm{pr}_d^d(u^d))$*, and*

$$\max_{\nu \in [0,d]\cap\mathbb{Z}} \sup_{\tau \in [0,T), \xi \in \mathbb{R}^d} \left[\Lambda_\nu^d(T-\tau) \frac{|\mathrm{pr}_\nu^d(u^d(\tau,\xi))|}{(1+\|\xi\|^2)^{\frac{1}{2}}} \right] < \infty. \tag{13}$$

(ii) *There exists* $\eta \in (0,\infty)$ *such that for all* $d \in \mathbb{N}$, $\epsilon \in (0,1)$ *there exists* $\Psi_{d,\epsilon} \in \mathbf{N}$ *satisfying* $\mathcal{R}(\Psi_{d,\epsilon}) \in C(\mathbb{R}^d, \mathbb{R}^{d+1})$, $\mathcal{P}(\Psi_{d,\epsilon}) \leqslant \eta d^\eta \epsilon^{-\eta}$*, and*

$$\left(\int_{[0,1]^d} \sum_{\nu=0}^{d} \left| \Lambda_\nu^d(T) \mathrm{pr}_\nu^d((\mathcal{R}(\Psi_{d,\epsilon}))(x) - u^d(0,x)) \right|^2 dx \right)^{\frac{1}{2}} \leqslant \epsilon. \tag{14}$$

Let us make some comments on the mathematical objects in Theorem 1.2 above. First, (5)–(7) are growth and Lipschitz conditions. Condition (8)–(9) ensure that the input functions f^d, g^d can be approximated by the functions f_ε^d, g_ε^d. The bound $cd^c\varepsilon^{-c}$ in condition (10), which is a polynomial of d and ε^{-1}, ensures that the functions f_ε^d, g_ε^d can be represented by DNNs without curse of dimensionality. Under these assumptions Theorem 1.2 states that, roughly speaking, if DNNs can approximate the initial condition, the linear part, and the nonlinear part in (11)–(12) without curse of dimensionality, then they can also approximate its solution without curse of dimensionality.

The proof of Theorem 1.2 above relies on full history recursive multilevel Picard (MLP) approximations which have been proved to overcome the curse of dimensionality when approximating solutions of semilinear heat equations in the gradient dependent case, see [30, 42]. Our proof essentially consists of two main lemmas: Proposition 2.12 shows that realizations of certain MLP approximations can be represented by DNNs and Proposition 3.1 is a perturbation lemma.

The remaining part of our paper is organized as follows. In Section 2 we prove that multilevel Picard approximations can be represented by DNNs and we provide bounds for the number of parameters of the representing DNN. In Section 3 we provide the perturbation lemma, Proposition 3.1, demonstrating that the stochastic fixed point equation corresponding to the PDE (11)–(12) is stable against perturbations in the nonlinearity f and the terminal condition g. The main theorem is proven in Section 4.

2. Deep neural networks

2.1. Properties of operations associated to DNNs.

In Setting 2.1 below we introduce operations which are important for constructing the random DNN that represents the MLP approximations in the proof of Proposition 2.12.

Setting 2.1. *Assume Setting 1.1. 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* $\alpha \odot \beta = (\beta_0, \beta_1, \ldots, \beta_{H_2}, \beta_{H_2+1} + \alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{H_1+1}) \in \mathbb{N}^{H_1+H_2+3}$*. 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* $\alpha \boxplus \beta = (\alpha_0, \alpha_1 + \beta_1, \ldots, \alpha_H + \beta_H, \beta_{H+1}) \in \mathbb{N}^{H+2}$*. Let* $\mathfrak{n}_n \in \mathbf{D}$, $n \in [3,\infty) \cap \mathbb{Z}$*, satisfy for all* $n \in [3,\infty) \cap \mathbb{N}$ *that*

$$\mathfrak{n}_n = (1, \underbrace{2, \ldots, 2}_{(n-2)\text{ times}}, 1) \in \mathbb{N}^n. \tag{15}$$

For every $n \in \mathbb{N}$ let $R_n\colon \mathbf{D} \to \mathbf{D}$ satisfy for all $H \in \mathbb{N}$, $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_H, \alpha_{H+1}) \in \mathbb{N}^{H+2}$ that

$$R_n(\alpha) = (\alpha_0, \alpha_1, \ldots, \alpha_H, n). \tag{16}$$

For the proof of our main result in this section, Proposition 2.12, we need several auxiliary results, Lemmas 2.2–2.11, which are basic facts on DNNs. The proof of Lemma 2.2–2.11 can be found in [11, 31] and therefore omitted.

Lemma 2.2 ($\odot$ is associative– [31, Lemma 3.3])**.** *Assume Setting 2.1 and let $\alpha, \beta, \gamma \in \mathbf{D}$. Then we have that $(\alpha \odot \beta) \odot \gamma = \alpha \odot (\beta \odot \gamma)$.*

Lemma 2.3 ($\boxplus$ and associativity– [31, Lemma 3.4])**.** *Assume Setting 2.1, 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 $(\alpha \boxplus \beta) \boxplus \gamma = \alpha \boxplus (\beta \boxplus \gamma)$.

Lemma 2.4 (Triangle inequality– [31, Lemma 3.5])**.** *Assume Setting 2.1, let $H, k, l \in \mathbb{N}$, and let $\alpha, \beta \in \{k\} \times \mathbb{N}^H \times \{l\}$. Then we have that $|||\alpha \boxplus \beta||| \leqslant |||\alpha||| + |||\beta|||$.*

Lemma 2.5. *Assume Setting 2.1, let $H, k, l, n, m \in \mathbb{N}$, and let $\alpha_1, \alpha_2, \ldots, \alpha_m \in \{k\} \times \mathbb{N}^H \times \{l\}$. Then we have that $|||\boxplus_{i=1}^m R_n(\alpha_i)||| \leqslant \max\{\sum_{i=1}^m |||\alpha_i|||, n\}$.*

Proof of Lemma 2.5. Throughout this proof for every $i \in [1, m] \cap \mathbb{Z}$ let $\alpha_{i,j} \in \mathbb{N}$, $j \in [1, H] \cap \mathbb{Z}$, satisfy that $\alpha_i = (k, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,H}, l)$. Then the definition of $\boxplus$, the definition of R_n, and the triangle inequality show that $\boxplus_{i=1}^m R_n(\alpha_i) = (k, \sum_{i=1}^m \alpha_{i,1}, \sum_{i=1}^m \alpha_{i,2}, \ldots, \sum_{i=1}^m \alpha_{i,H}, n)$ and

$$\begin{aligned} \left|\left|\left| \mathop{\boxplus}_{i=1}^m R_n(\alpha_i) \right|\right|\right| &= \sup\left\{ k, \left|\sum_{i=1}^m \alpha_{i,1}\right|, \left|\sum_{i=1}^m \alpha_{i,2}\right|, \ldots, \left|\sum_{i=1}^m \alpha_{i,H}\right|, n \right\} \\ &\leqslant \sup\left\{ k, \sum_{i=1}^m |\alpha_{i,1}|, \sum_{i=1}^m |\alpha_{i,2}|, \ldots, \sum_{i=1}^m |\alpha_{i,H}|, n \right\} \\ &\leqslant \max\left\{ \sum_{i=1}^m |||\alpha_i|||, n \right\}. \end{aligned} \tag{17}$$

This completes the proof of Lemma 2.5. ☐

Lemma 2.6 (DNNs for affine transformations– [31, Lemma 3.7])**.** *Assume Setting 1.1 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 $\lambda\left((\mathcal{R}(\Psi))(\cdot + b) + a\right) \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = \mathcal{D}(\Psi)\})$.*

Lemma 2.7 (DNNs for the multiplication with a vector)**.** *Assume Setting 2.1 and let $d, m \in \mathbb{N}$, $\lambda \in \mathbb{R}^m = \mathbb{R}^{m \times 1}$, $a \in \mathbb{R}$, $b \in \mathbb{R}^d$, $\Psi \in \mathbf{N}$ satisfy that $\mathcal{R}(\Psi) \in C(\mathbb{R}^d, \mathbb{R})$. Then we have that $\lambda\left((\mathcal{R}(\Psi))(\cdot + b) + a\right) \in \mathcal{R}(\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi) = R_m(\mathcal{D}(\Psi))\})$.*

Proof of Lemma 2.7. Throughout this proof let $H, k_0, k_1, \ldots, k_{H+1} \in \mathbb{N}$ satisfy that

$$H + 2 = \dim(\mathcal{D}(\Psi)) \quad \text{and} \quad (k_0, k_1, \ldots, k_H, k_{H+1}) = \mathcal{D}(\Psi), \tag{18}$$

and let $((W_1, B_1), (W_2, B_2), \ldots, (W_H, B_H), (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} \left(\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n}\right)$ satisfy that

$$\Big((W_1, B_1), (W_2, B_2), \ldots, (W_H, B_H), (W_{H+1}, B_{H+1})\Big) = \Psi. \tag{19}$$

Then the fact that $\mathcal{R}(\Psi) \in C(\mathbb{R}^d, \mathbb{R})$ implies that $k_0 = d$ and $k_{H+1} = 1$. Next, let $\phi \in \prod_{n=1}^{H} \left(\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n}\right) \times \left(\mathbb{R}^{m \times k_H} \times \mathbb{R}^m\right)$ satisfy that

$$\phi = \Big((W_1, B_1 + W_1 b), (W_2, B_2), \ldots, (W_H, B_H), (\lambda W_{H+1}, \lambda B_{H+1} + \lambda a)\Big), \tag{20}$$

This, the fact that $k_0 = d$, and (18) show that $\phi \in \mathbf{N}$ and

$$\mathcal{D}(\phi) = (d, k_1, k_2, \ldots, k_H, m) = R_m(\mathcal{D}(\Psi)). \tag{21}$$

Let $x_0, y_0 \in \mathbb{R}^{k_0}, x_1, y_1 \in \mathbb{R}^{k_1}, \ldots, x_H, y_H \in \mathbb{R}^{k_H}$ satisfy for all $n \in \mathbb{N} \cap [1, H]$ that

$$x_n = \mathbf{A}_{k_n}(W_n x_{n-1} + B_n),\ y_n = \mathbf{A}_{k_n}(W_n y_{n-1} + B_n + \mathbb{1}_{\{1\}}(n) W_1 b), \quad \text{and} \quad x_0 = y_0 + b. \tag{22}$$

Then

$$y_1 = \mathbf{A}_{k_1}(W_1 y_0 + B_1 + W_1 b) = \mathbf{A}_{k_1}(W_1(y_0 + b) + B_1) = \mathbf{A}_{k_1}(W_1 x_0 + B_1) = x_1. \tag{23}$$

This and an induction argument prove for all $i \in [2, H] \cap \mathbb{N}$ that

$$y_i = \mathbf{A}_{k_i}(W_i y_{i-1} + B_i) = \mathbf{A}_{k_i}(W_i x_{i-1} + B_i) = x_i. \tag{24}$$

This and the definition of $\mathcal{R}$ prove that

$$\begin{aligned}(\mathcal{R}(\phi))(y_0) &= \lambda W_{H+1} y_H + \lambda B_{H+1} + \lambda a \\ &= \lambda W_{H+1} x_H + \lambda B_{H+1} + \lambda a \\ &= \lambda(W_{H+1} x_H + B_{H+1} + a) \\ &= \lambda((\mathcal{R}(\Psi))(x_0) + a) \\ &= \lambda(\mathcal{R}(\Psi))(y_0 + b) + a.\end{aligned} \tag{25}$$

This and the fact that y_0 was arbitrary prove that

$$\mathcal{R}(\phi) = \lambda((\mathcal{R}(\Psi))(\cdot + b) + a). \tag{26}$$

This and (21) complete the proof of Lemma 2.7. □

Lemma 2.8 (Composition of functions generated by DNNs– [31, Lemma 3.8])**.** *Assume Setting 2.1 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 $(f \circ g) \in \mathcal{R}(\{\Phi \in \mathbf{N} \colon \mathcal{D}(\Phi) = \alpha \odot \beta\})$.*

Lemma 2.9 (Sum of DNNs of the same length– [31, Lemma 3.9])**.** *Assume Setting 2.1 and let $M, H, p, q \in \mathbb{N}$, $h_1, h_2, \ldots, h_M \in \mathbb{R}$, $k_i \in \mathbf{D}$, $f_i \in C(\mathbb{R}^p, \mathbb{R}^q)$, $i \in [1, M] \cap \mathbb{N}$, satisfy for all $i \in [1, M] \cap \mathbb{N}$ that $\dim(k_i) = H + 2$ and $f_i \in \mathcal{R}(\{\Phi \in \mathbf{N} \colon \mathcal{D}(\Phi) = k_i\})$. Then we have that $\sum_{i=1}^M h_i f_i \in \mathcal{R}\left(\{\Phi \in \mathbf{N} \colon \mathcal{D}(\Phi) = \boxplus_{i=1}^M k_i\}\right)$.*

Lemma 2.10 (Existence of DNNs with H hidden layers for $\mathrm{Id}_{\mathbb{R}^d}$– [11, Lemma 3.6])**.** *Assume Setting 2.1 and let $d, H \in \mathbb{N}$. Then we have that $\mathrm{Id}_{\mathbb{R}^d} \in \mathcal{R}(\{\Phi \in \mathbf{N} \colon \mathcal{D}(\Phi) = \mathfrak{n}_{H+2}^d\})$.*

Lemma 2.11 ([11, Lemma 3.7])**.** *Assume Setting 1.1, let $H, p, q \in \mathbb{N}$, and let $g \in C(\mathbb{R}^p, \mathbb{R}^q)$ satisfy that $g \in \mathcal{R}(\{\Phi \in \mathbf{N} \colon \dim(\mathcal{D}(\Phi)) = H + 2\})$. Then for all $n \in \mathbb{N}_0$ we have that $g \in \mathcal{R}(\{\Phi \in \mathbf{N} \colon \dim(\mathcal{D}(\Phi)) = H + 2 + n\})$.*

2.2. **DNN representation of MLP approximations.** In Proposition 2.12 below we prove that the MLP approximations under consideration can be represented by DNNs.

Proposition 2.12. *Assume Setting 1.1. Let $T \in (0, \infty)$, $d, m \in \mathbb{N}$, $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$. Let $c \in (0, \infty)$ satisfy that*

$$c \geqslant \max\{d + 1, |||\mathcal{D}(\Phi_f)|||, |||\mathcal{D}(\Phi_g)|||\}. \tag{27}$$

Let $\varrho \colon [0, T]^2 \to \mathbb{R}$ satisfy for all $t \in [0, T)$, $s \in (t, T)$ that

$$\varrho(t, s) = \frac{1}{\mathrm{B}(\frac{1}{2}, \frac{1}{2})} \frac{1}{\sqrt{(T - s)(s - t)}}. \tag{28}$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $W^\theta \colon [0, T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, be standard $(\mathbb{F}_t)_{t \in [0,T]}$-Brownian motions with continuous sample paths. Let $\mathfrak{t}^\theta \colon \Omega \to (0, 1)$, $\theta \in \Theta$, be i.i.d. random variables. Assume for all $b \in (0, 1)$ that

$$\mathbb{P}(\mathfrak{t}^0 \leqslant b) = \frac{1}{\mathrm{B}(\frac{1}{2}, \frac{1}{2})} \int_0^b \frac{dr}{\sqrt{r(1 - r)}}. \tag{29}$$

Assume that $(W^\theta)_{\theta \in \Theta}$ and $(\mathfrak{t}^\theta)_{\theta \in \Theta}$ are independent. Let $\mathfrak{T}^\theta \colon [0, T) \times \Omega \to [0, T)$, $\theta \in \Theta$, satisfy for all $t \in [0, T)$ that $\mathfrak{T}_t^\theta = t + (T - t)\mathfrak{t}^\theta$. Let $f \in C([0, T] \times \mathbb{R}^d \times \mathbb{R}^{1+d}, \mathbb{R})$, $g \in C(\mathbb{R}^d, \mathbb{R})$. Let

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

$$\begin{aligned} U^\theta_{n,m}(t,x) &= (g(x),0)+\sum_{i=1}^{m^n}\left(1,\frac{W_T^{(\theta,0,-i)}-W_t^{(\theta,0,-i)}}{T-t}\right)^\top \frac{\big(g(x+W_T^{(\theta,0,-i)}-W_t^{(\theta,0,-i)})-g(x)\big)}{m^n}\\ &+\sum_{\ell=0}^{n-1}\sum_{i=1}^{m^{n-\ell}}\left(1,\frac{W^{(\theta,\ell,i)}_{\mathfrak{T}_t^{(\theta,\ell,i)}}-W_t^{(\theta,\ell,i)}}{\mathfrak{T}_t^{(\theta,\ell,i)}-t}\right)^\top \frac{\big(f\circ U^{(\theta,\ell,i)}_{\ell,m}-\mathbb{1}_{\mathbb{N}}(\ell)f\circ U^{(\theta,-\ell,i)}_{\ell-1,m}\big)(\mathfrak{T}_t^{(\theta,\ell,i)},x+W^{(\theta,\ell,i)}_{\mathfrak{T}_t^{(\theta,\ell,i)}}-W_t^{(\theta,\ell,i)})}{m^{n-\ell}\varrho(t,\mathfrak{T}_t^{(\theta,\ell,i)})}, \end{aligned}\tag{30}$$

and let $\omega\in\Omega$. *Then for all* $n\in\mathbb{N}_0$ *there exists* $(\Phi^\theta_{n,t})_{\theta\in\Theta,t\in[0,T)}\subseteq\mathbf{N}$ *such that for all* $t_1,t_2,t\in[0,T)$, $\theta_1,\theta_2,\theta\in\Theta$ *we have that*

$$\mathcal{D}(\Phi^{\theta_1}_{n,t_1})=\mathcal{D}(\Phi^{\theta_2}_{n,t_2}),\tag{31}$$

$$\dim(\mathcal{D}(\Phi^\theta_{n,t}))=n(\dim(\mathcal{D}(\Phi_f))-1)+\dim(\mathcal{D}(\Phi_g)),\tag{32}$$

$$\left\|\!\left|\mathcal{D}(\Phi^\theta_{n,t})\right|\!\right\|\leqslant c(4m)^n,\tag{33}$$

$$U^\theta_{n,m}(t,x,\omega)=(\mathcal{R}(\Phi^\theta_{n,t}))(x).\tag{34}$$

Proof of Proposition 2.12. We prove the lemma by induction on $n\in\mathbb{N}$. Since the zero function can be represented by a DNN of arbitrary length, the base case $n=0$ is clear. For the induction step from $n\in\mathbb{N}_0$ to $n+1\in\mathbb{N}$ let $n\in\mathbb{N}_0$, assume for all $\ell\in[0,n]\cap\mathbb{Z}$, $t_1,t_2,t\in[0,T)$, $\theta_1,\theta_2,\theta\in\Theta$ that

$$\mathcal{D}(\Phi^{\theta_1}_{\ell,t_1})=\mathcal{D}(\Phi^{\theta_2}_{\ell,t_2}),\tag{35}$$

$$\dim(\mathcal{D}(\Phi^\theta_{\ell,t}))=\ell(\dim(\mathcal{D}(\Phi_f))-1)+\dim(\mathcal{D}(\Phi_g)),\tag{36}$$

$$\left\|\!\left|\mathcal{D}(\Phi^\theta_{\ell,t})\right|\!\right\|\leqslant c(3m)^\ell,\tag{37}$$

$$U^\theta_{\ell,m}(t,x,\omega)=(\mathcal{R}(\Phi^\theta_{\ell,t}))(x),\tag{38}$$

and let the notations in Setting 2.1 be given. First, Lemmas 2.6, 2.8, and 2.10 imply for all $i\in[1,m^{n+1}]\cap\mathbb{Z}$, $\theta\in\Theta$, $t\in[0,T)$ that

$$g=\mathrm{Id}_\mathbb{R}\circ g\in\mathcal{R}\big(\big\{\Phi\in\mathbf{N}\colon\mathcal{D}(\Phi)=\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1}\odot\mathcal{D}(\Phi_g)\big\}\big)\tag{39}$$

and[1]

$$\begin{aligned} &g\Big(\cdot+W_T^{(\theta,0,-i)}(\omega)-W_t^{(\theta,0,-i)}(\omega)\Big)=(\mathrm{Id}_\mathbb{R}\circ g)\Big(\cdot+W_T^{(\theta,0,-i)}(\omega)-W_t^{(\theta,0,-i)}(\omega)\Big)\\ &\in\mathcal{R}\left(\big\{\Phi\in\mathbf{N}\colon\mathcal{D}(\Phi)=\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1}\odot\mathcal{D}(\Phi_g)\big\}\right). \end{aligned}\tag{40}$$

This and Lemma 2.7 imply for all $i\in[1,m^{n+1}]\cap\mathbb{Z}$, $\theta\in\Theta$, $t\in[0,T)$ that

$$\begin{aligned} &\left(0,\frac{W_T^{(\theta,0,-i)}(\omega)-W_t^{(\theta,0,-i)}(\omega)}{T-t}\right)^\top g(x)\\ &\in\mathcal{R}\big(\big\{\Phi\in\mathbf{N}\colon\mathcal{D}(\Phi)=R_{d+1}\big(\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1}\odot\mathcal{D}(\Phi_g)\big)\big\}\big), \end{aligned}\tag{41}$$

and

$$\begin{aligned} &\left(1,\frac{W_T^{(\theta,0,-i)}(\omega)-W_t^{(\theta,0,-i)}(\omega)}{T-t}\right)^\top g\Big(\cdot+W_T^{(\theta,0,-i)}(\omega)-W_t^{(\theta,0,-i)}(\omega)\Big)\\ &\in\mathcal{R}\big(\big\{\Phi\in\mathbf{N}\colon\mathcal{D}(\Phi)=R_{d+1}\big(\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1}\odot\mathcal{D}(\Phi_g)\big)\big\}\big). \end{aligned}\tag{42}$$

[1]For every $d_1,d_2\in\mathbb{N}$, $z\in\mathbb{R}^{d_1}$, $h\in\mathbb{R}^{d_1}\to\mathbb{R}^{d_2}$ we denote by $h(\cdot+z)$ the function $\mathbb{R}^{d_1}\ni x\mapsto h(x+z)\in\mathbb{R}^{d_2}$.

Next, the induction hypothesis (see (35) and (38)), the fact that $f = \mathcal{R}(\Phi_f)$, and Lemma 2.8 show for all $i \in [1, m]$, $\theta \in \Theta$, $t \in [0, T)$ that

$$\begin{aligned}&\left(f \circ U^{(\theta,n,i)}_{n,m}\right)\left(\mathfrak{T}^{(\theta,n,i)}_t(\omega), \cdot + W^{(\theta,n,i)}_{\mathfrak{T}^{(\theta,n,i)}_t(\omega)}(\omega) - W^{(\theta,n,i)}_t(\omega), \omega\right)\\ &\in \mathcal{R}\big(\{\Phi \in \mathbf{N}\colon \mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{n,0})\}\big) .\end{aligned} \tag{43}$$

This and Lemma 2.7 imply for all $i \in [1, m]$, $\theta \in \Theta$, $t \in [0, T)$ that

$$\begin{aligned}&\left(1, \frac{W^{(\theta,n,i)}_{\mathfrak{T}^{(\theta,n,i)}_t(\omega)}(\omega) - W^{(\theta,n,i)}_t(\omega)}{\mathfrak{T}^{(\theta,n,i)}_t(\omega) - t}\right)^{\top} \left(f \circ U^{(\theta,n,i)}_{n,m}\right)\left(\mathfrak{T}^{(\theta,n,i)}_t(\omega), \cdot + W^{(\theta,n,i)}_{\mathfrak{T}^{(\theta,n,i)}_t(\omega)}(\omega) - W^{(\theta,n,i)}_t(\omega), \omega\right)\\ &\in \mathcal{R}\big(\{\Phi \in \mathbf{N}\colon R_{d+1}(\mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{n,0}))\}\big) .\end{aligned} \tag{44}$$

Next, Lemmas 2.10 and 2.8, the fact that $f = \mathcal{R}(\Phi_f)$, and the induction hypothesis (see (35) and (38)) prove for all $\ell \in [0, n] \cap \mathbb{Z}$, $\theta \in \Theta$, $i \in [m^{n+1-\ell}] \cap \mathbb{Z}$, $t \in [0, T)$ that

$$\begin{aligned}&\left(f \circ U^{(\theta,\ell,i)}_{\ell,m}\right)\left(\mathfrak{T}^{(\theta,\ell,i)}_t(\omega), \cdot + W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t(\omega)}(\omega) - W^{(\theta,\ell,i)}_t(\omega), \omega\right)\\ &= \left(\mathrm{Id}_{\mathbb{R}} \circ f \circ U^{(\theta,\ell,i)}_{\ell,m}\right)\left(\mathfrak{T}^{(\theta,\ell,i)}_t(\omega), \cdot + W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t(\omega)}(\omega) - W^{(\theta,\ell,i)}_t(\omega), \omega\right)\\ &\in \mathcal{R}\big(\{\Phi \in \mathbf{N}\colon \mathfrak{n}_{(n-\ell)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{\ell,0})\}\big) .\end{aligned} \tag{45}$$

This and Lemma 2.7 demonstrate for all $\ell \in [0, n] \cap \mathbb{Z}$, $\theta \in \Theta$, $i \in [m^{n+1-\ell}] \cap \mathbb{Z}$, $t \in [0, T)$ that

$$\begin{aligned}&\left(1, \frac{W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t(\omega)}(\omega) - W^{(\theta,\ell,i)}_t(\omega)}{\mathfrak{T}^{(\theta,\ell,i)}_t(\omega) - t}\right)^{\top} \left(f \circ U^{(\theta,\ell,i)}_{\ell,m}\right)\left(\mathfrak{T}^{(\theta,\ell,i)}_t(\omega), \cdot + W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t(\omega)}(\omega) - W^{(\theta,\ell,i)}_t(\omega), \omega\right)\\ &\in \mathcal{R}\big(\{\Phi \in \mathbf{N}\colon R_{d+1}(\mathfrak{n}_{(n-\ell)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{\ell,0}))\}\big) .\end{aligned} \tag{46}$$

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

$$\begin{aligned}&\left(f \circ U^{(\theta,-\ell,i)}_{\ell-1,m}\right)\left(\mathfrak{T}^{(\theta,\ell,i)}_t(\omega), \cdot + W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t(\omega)}(\omega) - W^{(\theta,\ell,i)}_t(\omega), \omega\right)\\ &= \left(\mathrm{Id}_{\mathbb{R}} \circ f \circ U^{(\theta,-\ell,i)}_{\ell-1,m}\right)\left(\mathfrak{T}^{(\theta,\ell,i)}_t(\omega), \cdot + W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t(\omega)}(\omega) - W^{(\theta,\ell,i)}_t(\omega)\right)\\ &\in \mathcal{R}\big(\{\Phi \in \mathbf{N}\colon \mathfrak{n}_{(n-\ell+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{\ell-1,0})\}\big)\end{aligned} \tag{47}$$

and

$$\begin{aligned}&\left(1, \frac{W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t(\omega)}(\omega) - W^{(\theta,\ell,i)}_t(\omega)}{\mathfrak{T}^{(\theta,\ell,i)}_t(\omega) - t}\right)^{\top} \left(f \circ U^{(\theta,-\ell,i)}_{\ell-1,m}\right)\left(\mathfrak{T}^{(\theta,\ell,i)}_t(\omega), \cdot + W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t(\omega)}(\omega) - W^{(\theta,\ell,i)}_t(\omega), \omega\right)\\ &\in \mathcal{R}\big(\{\Phi \in \mathbf{N}\colon R_{d+1}(\mathfrak{n}_{(n-\ell+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{\ell-1,0}))\}\big) .\end{aligned} \tag{48}$$

Next, (16), the definition of $\odot$, and (15) imply that

$$\begin{aligned}&\dim\big(R_{d+1}(\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_g))\big)\\ &= \dim\big(\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_g)\big)\\ &= (n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + 1 + \dim(\mathcal{D}(\Phi_g)) - 1\\ &= (n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g)).\end{aligned} \tag{49}$$

Furthermore, (16), the definition of $\odot$, and the induction hypothesis (see (36)) show that

$$\begin{aligned}&\dim\big(R_{d+1}(\mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{n,0}))\big)\\ &= \dim\big(\mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{n,0})\big)\\ &= \dim(\mathcal{D}(\Phi_f)) + \dim(\mathcal{D}(\Phi^0_{n,0})) - 1\\ &= \dim(\mathcal{D}(\Phi_f)) + [n(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g))] - 1\end{aligned}$$

$$= (n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g)). \tag{50}$$

Moreover, (16), the definition of $\odot$, (15), the induction hypothesis (see (36)) imply that

$$\begin{aligned}
&\dim\big(R_{d+1}\big(\mathfrak{n}_{(n-\ell)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{\ell,0})\big)\big) \\
&= \dim\big(\mathfrak{n}_{(n-\ell)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{\ell,0})\big) \\
&= (n-\ell)(\dim(\mathcal{D}(\Phi_f)) - 1) + 1 + \dim(\mathcal{D}(\Phi_f)) + \dim(\mathcal{D}(\Phi^0_{\ell,0})) - 2 \\
&= (n-\ell)(\dim(\mathcal{D}(\Phi_f)) - 1) + 1 + \dim(\mathcal{D}(\Phi_f)) + [\ell(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g))] - 2 \\
&= (n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g)).
\end{aligned} \tag{51}$$

Now, (49)–(51) show, roughly speaking, that the functions in (41), (42), (44), (46), and (48) can be represented by networks with the same number of layers: $(n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g))$. Hence, Lemma 2.9, (41), (42), (44), (46), and (48) imply that there exists $(\Phi^\theta_{n+1,t})_{\theta\in\Theta,t\in[0,T)} \subseteq \mathbf{N}$ such that for all $\theta \in \Theta$, $t \in [0,T)$, $x \in \mathbb{R}^d$ we have that

$$\begin{aligned}
&(\mathcal{R}(\Phi^\theta_{n+1,t}))(x) \\
&= \sum_{i=1}^{m^{n+1}} \left(1, \tfrac{W^{(\theta,0,-i)}_T - W^{(\theta,0,-i)}_t}{T-t}\right)^\top \tfrac{g(x+W^{(\theta,0,-i)}_T - W^{(\theta,0,-i)}_t)}{m^{n+1}} \\
&\quad - \sum_{i=1}^{m^{n+1}} \left(0, \tfrac{W^{(\theta,0,-i)}_T - W^{(\theta,0,-i)}_t}{T-t}\right)^\top \tfrac{g(x)}{m^{n+1}} \\
&\quad + \sum_{i=1}^{m} \left(1, \tfrac{W^{(\theta,n,i)}_{\mathfrak{T}^{(\theta,n,i)}_t} - W^{(\theta,n,i)}_t}{\mathfrak{T}^{(\theta,n,i)}_t - t}\right)^\top \tfrac{(f\circ U^{(\theta,n,i)}_{n,m})\left(\mathfrak{T}^{(\theta,n,i)}_t, x+W^{(\theta,n,i)}_{\mathfrak{T}^{(\theta,n,i)}_t} - W^{(\theta,n,i)}_t\right)}{m^{n+1-n}\varrho(t,\mathfrak{T}^{(\theta,n,i)}_t)} \\
&\quad + \sum_{\ell=0}^{n-1}\sum_{i=1}^{m^{n+1-\ell}} \left(1, \tfrac{W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t} - W^{(\theta,\ell,i)}_t}{\mathfrak{T}^{(\theta,\ell,i)}_t - t}\right)^\top \tfrac{(f\circ U^{(\theta,\ell,i)}_{\ell,m})\left(\mathfrak{T}^{(\theta,\ell,i)}_t, x+W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t} - W^{(\theta,\ell,i)}_t\right)}{m^{n+1-\ell}\varrho(t,\mathfrak{T}^{(\theta,\ell,i)}_t)} \\
&\quad - \sum_{\ell=1}^{n}\sum_{i=1}^{m^{n+1-\ell}} \left(1, \tfrac{W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t} - W^{(\theta,\ell,i)}_t}{\mathfrak{T}^{(\theta,\ell,i)}_t - t}\right)^\top \tfrac{(f\circ U^{(\theta,-\ell,i)}_{\ell-1,m})\left(\mathfrak{T}^{(\theta,\ell,i)}_t, x+W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t} - W^{(\theta,\ell,i)}_t\right)}{m^{n+1-\ell}\varrho(t,\mathfrak{T}^{(\theta,\ell,i)}_t)} \\
&= (g(x), 0) + \sum_{i=1}^{m^{n+1}} \left(1, \tfrac{W^{(\theta,0,-i)}_T - W^{(\theta,0,-i)}_t}{T-t}\right)^\top \tfrac{g(x+W^{(\theta,0,-i)}_T - W^{(\theta,0,-i)}_t) - g(x)}{m^{n+1}} \\
&\quad + \sum_{\ell=0}^{n}\sum_{i=1}^{m^{n+1-\ell}} \left(1, \tfrac{W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t} - W^{(\theta,\ell,i)}_t}{\mathfrak{T}^{(\theta,\ell,i)}_t - t}\right)^\top \tfrac{(f\circ U^{(\theta,\ell,i)}_{\ell,m} - \mathbb{1}_{\mathbb{N}}(\ell) f\circ U^{(\theta,-\ell,i)}_{\ell-1,m})\left(\mathfrak{T}^{(\theta,\ell,i)}_t, x+W^{(\theta,\ell,i)}_{\mathfrak{T}^{(\theta,\ell,i)}_t} - W^{(\theta,\ell,i)}_t\right)}{m^{n+1-\ell}\varrho(t,\mathfrak{T}^{(\theta,\ell,i)}_t)} \\
&= U^\theta_{n+1,m}(t,x,\omega),
\end{aligned} \tag{52}$$

$$\dim(\Phi^\theta_{n+1,t})) = (n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g)), \tag{53}$$

and

$$\begin{aligned}
\mathcal{D}(\Phi^\theta_{n+1,t}) = &\left(\boxplus_{i=1}^{m^{n+1}} R_{d+1}\big(\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_g)\big)\right) \\
&\boxplus \left(\boxplus_{i=1}^{m^{n+1}} R_{d+1}\big(\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_g)\big)\right) \\
&\boxplus \left(\boxplus_{i=1}^{m} R_{d+1}\big(\mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{n,0})\big)\right) \\
&\boxplus \left(\boxplus_{\ell=0}^{n-1}\boxplus_{i=1}^{m^{n+1-\ell}} R_{d+1}\big(\mathfrak{n}_{(n-\ell)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi^0_{\ell,0})\big)\right)
\end{aligned}$$

$$\boxplus\left(\mathop{\boxplus}_{\ell=1}^{n}\mathop{\boxplus}_{i=1}^{m^{n+1-\ell}} R_{d+1}\big(\mathfrak{n}_{(n-\ell+1)(\dim(\mathcal{D}(\Phi_f))-1)+1}\odot\mathcal{D}(\Phi_f)\odot\mathcal{D}(\Phi^0_{\ell-1,0})\big)\right). \tag{54}$$

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

$$\mathcal{D}(\Phi^{\theta_1}_{n+1,t_1})=\mathcal{D}(\Phi^{\theta_2}_{n+1,t_2}). \tag{55}$$

Next, the induction hypothesis (see (36)) and Lemma 2.5 imply for all $\theta\in\Theta$, $t\in[0,T)$ that

$$\begin{aligned}
\left\|\mathcal{D}(\Phi^{\theta}_{n+1,t})\right\| \leqslant \max\Bigg\{&\sum_{i=1}^{m^{n+1}}\left\|\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1}\odot\mathcal{D}(\Phi_g)\right\|\\
&+\sum_{i=1}^{m^{n+1}}\left\|\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1}\odot\mathcal{D}(\Phi_g)\right\|\\
&+\sum_{i=1}^{m}\left\|\mathcal{D}(\Phi_f)\odot\mathcal{D}(\Phi^0_{n,0})\right\|\\
&+\sum_{\ell=0}^{n-1}\sum_{i=1}^{m^{n+1-\ell}}\left\|\mathfrak{n}_{(n-\ell)(\dim(\mathcal{D}(\Phi_f))-1)+1}\odot\mathcal{D}(\Phi_f)\odot\mathcal{D}(\Phi^0_{\ell,0})\right\|\\
&+\sum_{\ell=1}^{n}\sum_{i=1}^{m^{n+1-\ell}}\mathfrak{n}_{(n-\ell+1)(\dim(\mathcal{D}(\Phi_f))-1)+1}\odot\mathcal{D}(\Phi_f)\odot\mathcal{D}(\Phi^0_{\ell-1,0}),d+1\Bigg\}.
\end{aligned} \tag{56}$$

Note that the definition of $\odot$ show that for all $H_1,H_2,\alpha_0,\alpha_1,\dots,\alpha_{H_1+1},\beta_0,\beta_1,\dots,\beta_{H_2+1}\in\mathbb{N}$, $\alpha,\beta\in\mathbf{D}$ with $\alpha=(\alpha_1,\dots,\alpha_{H_1+1})$, $\beta=(\beta_1,\dots,\beta_{H_2+1})$, $\alpha_0=\beta_{H_2+1}=1$ we have that $\|\alpha\odot\beta\|\leqslant\max\{\|\alpha\|,\|\beta\|,2\}$. This, (55), (15), (27), and the induction hypothesis (see (37)) prove for all $\theta\in\Theta$, $t\in[0,T)$ that

$$\begin{aligned}
&\left\|\mathcal{D}(\Phi^{\theta}_{n+1,t})\right\|\\
&\leqslant 2\left[\sum_{i=1}^{m^{n+1}}c\right]+\left[\sum_{i=1}^{m}c(4m)^n\right]+\left[\sum_{\ell=0}^{n-1}\sum_{i=1}^{m^{n+1-\ell}}c(4m)^{\ell}\right]+\left[\sum_{\ell=1}^{n}\sum_{i=1}^{m^{n+1-\ell}}c(4m)^{\ell-1}\right]\\
&\leqslant 2cm^{n+1}+mc(4m)^n+\left[\sum_{\ell=0}^{n-1}m^{n+1-\ell}c(4m)^{\ell}\right]+\left[\sum_{\ell=1}^{n}m^{n+1-\ell}c(4m)^{\ell-1}\right]\\
&\leqslant 2cm^{n+1}+cm^{n+1}4^n+cm^{n+1}\left[\sum_{\ell=0}^{n-1}4^{\ell}\right]+cm^{n+1}\left[\sum_{\ell=1}^{n}4^{\ell-1}\right]\\
&\leqslant cm^{n+1}\left[2+2\sum_{\ell=0}^{n}4^{\ell}\right]\leqslant cm^{n+1}\left[1+3\sum_{\ell=0}^{n}4^{\ell}\right]=cm^{n+1}\left[1+3\frac{4^{n+1}-1}{4-1}\right]\\
&=c(4m)^{n+1}.
\end{aligned} \tag{57}$$

This, (55), (52), and (53) complete the induction step. Induction then completes the proof of Proposition 2.12. ☐

3. Perturbation lemma

Proposition 3.1 (Perturbation lemma). *Let $d\in\mathbb{N}$, $T,\varepsilon\in(0,\infty)$, $p_{\mathrm{v}},p_{\mathrm{z}},p_{\mathrm{x}}\in[1,\infty)$, $q\in[1,2)$, $c\in[1,\infty)$, $(L_i)_{i\in[0,d]\cap\mathbb{Z}}\in\mathbb{R}^{d+1}$ satisfy that $\sum_{i=0}^{d}L_i\leqslant c$, $\frac{2}{p_{\mathrm{v}}}+\frac{1}{p_{\mathrm{x}}}+\frac{1}{p_{\mathrm{z}}}\leqslant 1$, $2q\leqslant p_{\mathrm{v}}$, and $\frac{1}{2q}+\frac{1}{p_{\mathrm{z}}}\leqslant 1$. Let $\|\cdot\|\colon\mathbb{R}^d\to[0,\infty)$ be a norm on $\mathbb{R}^d$. Let $\Lambda=(\Lambda_\nu)_{\nu\in[0,d]\cap\mathbb{Z}}\colon[0,T]\to\mathbb{R}^{1+d}$ satisfy for all $t\in[0,T]$ that $\Lambda(t)=(1,\sqrt{t},\dots,\sqrt{t})$. Let $\mathrm{pr}=(\mathrm{pr}_\nu)_{\nu\in[0,d]\cap\mathbb{Z}}\colon\mathbb{R}^{d+1}\to\mathbb{R}$ satisfy for all $w=(w_\nu)_{\nu\in[0,d]\cap\mathbb{Z}}$, $i\in[0,d]\cap\mathbb{Z}$ that $\mathrm{pr}_i(w)=w_i$. Let $f,\tilde{f}\in C(\mathbb{R}^{d+1},\mathbb{R})$, $g,\tilde{g}\in C(\mathbb{R}^d,\mathbb{R})$, $V\in C([0,T]\times\mathbb{R}^d,[0,\infty))$ satisfy that $\max\{c,48e^{86c^6T^3}\}\leqslant V$. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. Let $(X^{s,x}_t)_{s\in[0,T],t\in[s,T],x\in\mathbb{R}^d}\colon\{(\sigma,\tau)\in[0,T]^2\colon\sigma\leqslant\tau\}\times\mathbb{R}^d\times\Omega\to\mathbb{R}^d$, $(Z^{s,x}_t)_{s\in[0,T),t\in(s,T],x\in\mathbb{R}^d}\colon\{(\sigma,\tau)\in$*

$[0,T]^2\colon \sigma<\tau\}\times\mathbb{R}^d\times\Omega\to\mathbb{R}^{d+1}$ *be measurable. Assume for all* $i\in[0,d]\cap\mathbb{Z}$, $s\in[0,T)$, $t\in[s,T)$, $r\in(t,T]$, $x,y\in\mathbb{R}^d$, $w_1,w_2\in\mathbb{R}^{d+1}$, $A\in(\mathcal{B}(\mathbb{R}^d))^{\otimes\mathbb{R}^d}$, $B\in(\mathcal{B}(\mathbb{R}^d))^{\otimes([t,T)\times\mathbb{R}^d)}$ *that*

$$\max\{|g(x)|,|\tilde g(x)|\}\leqslant V(T,x),\quad \max\Big\{|Tf(0)|,|T\tilde f(0)|\Big\}\leqslant V(t,x), \tag{58}$$

$$\max\Big\{|f(w_1)-f(w_2)|,|\tilde f(w_1)-\tilde f(w_2)|\Big\}\leqslant\sum_{\nu=0}^d\left[L_\nu\Lambda_\nu(T)|\mathrm{pr}_\nu(w_1-w_2)|\right], \tag{59}$$

$$\big\|V(r,X_r^{t,x})\big\|_{p_{\mathrm{v}}}\leqslant V(t,x),\quad \big\|\mathrm{pr}_i(Z_r^{t,x})\big\|_{p_{\mathrm{z}}}\leqslant\frac{c}{\Lambda_i(r-t)}, \tag{60}$$

$$\max\{|g(x)-g(y)|,|\tilde g(x)-\tilde g(y)|\}\leqslant\frac{V(T,x)+V(T,y)}{2}\frac{\|x-y\|}{\sqrt{T}}, \tag{61}$$

$$\big\|\|X_r^{t,x}-X_r^{t,y}\|\big\|_{p_{\mathrm{x}}}\leqslant c\|x-y\|, \tag{62}$$

$$\big\|\mathrm{pr}_i\big(Z_r^{t,x}-Z_r^{t,y}\big)\big\|_{p_{\mathrm{z}}}\leqslant\frac{V(t,x)+V(t,y)}{2}\frac{\|x-y\|}{\sqrt{T}\Lambda_i(r-t)}, \tag{63}$$

$$\mathbb{P}\Big(X_r^{t,X_t^{s,x}}=X_r^{s,x}\Big)=1,\quad \mathbb{P}\Big(X_t^{s,(\cdot)}\in A,X_{(\cdot)}^{t,(\cdot)}\in B\Big)=\mathbb{P}\Big(X_t^{s,(\cdot)}\in A\Big)\,\mathbb{P}\Big(X_{(\cdot)}^{t,(\cdot)}\in B\Big), \tag{64}$$

$$\big\|\mathrm{pr}_i(Z_r^{t,x}-Z_r^{s,x})\big\|_{p_{\mathrm{z}}}\leqslant\frac{V(t,x)+V(s,x)}{2}\frac{\sqrt{t-s}}{\sqrt{r-t}\Lambda_i(r-s)}, \tag{65}$$

$$\big\|\|X_t^{s,x}-x\|\big\|_{p_{\mathrm{x}}}\leqslant V(s,x)\sqrt{t-s}, \tag{66}$$

$$\Big|\tilde f(w)-f(w)\Big|\leqslant\frac{\varepsilon cd^c}{T}\left(1+\sum_{\nu=0}^d(\Lambda_\nu(T))^q|\mathrm{pr}_\nu(w)|^q\right), \tag{67}$$

$$|g(x)-\tilde g(x)|\leqslant\varepsilon cd^cV(T,x), \tag{68}$$

and $\mathbb{P}(X_s^{s,x}=x)=1$. *Then the following items hold.*

(i) *There exist unique continuous functions* $u,\tilde u\colon[0,T)\times\mathbb{R}^d\to\mathbb{R}^{d+1}$ *such that for all* $t\in[0,T)$, $x\in\mathbb{R}^d$ *we have that*

$$\max_{\nu\in[0,d]\cap\mathbb{Z}}\sup_{\tau\in[0,T),\xi\in\mathbb{R}^d}\left[\Lambda_\nu(T-\tau)\frac{|\mathrm{pr}_\nu(u(\tau,\xi))|}{V(\tau,\xi)}\right]<\infty, \tag{69}$$

$$\max_{\nu\in[0,d]\cap\mathbb{Z}}\sup_{\tau\in[0,T),\xi\in\mathbb{R}^d}\left[\Lambda_\nu(T-\tau)\frac{|\mathrm{pr}_\nu(\tilde u(\tau,\xi))|}{V(\tau,\xi)}\right]<\infty, \tag{70}$$

$$\max_{\nu\in[0,d]\cap\mathbb{Z}}\left[\mathbb{E}\big[\big|g(X_T^{t,x})\mathrm{pr}_\nu(Z_T^{t,x})\big|\big]+\int_t^T\mathbb{E}\big[\big|f(u(r,X_r^{t,x}))\mathrm{pr}_\nu(Z_r^{t,x})\big|\big]\,dr\right]<\infty, \tag{71}$$

$$\max_{\nu\in[0,d]\cap\mathbb{Z}}\left[\mathbb{E}\big[\big|\tilde g(X_T^{t,x})\mathrm{pr}_\nu(Z_T^{t,x})\big|\big]+\int_t^T\mathbb{E}\Big[\Big|\tilde f(\tilde u(r,X_r^{t,x}))\mathrm{pr}_\nu(Z_r^{t,x})\Big|\Big]\,dr\right]<\infty, \tag{72}$$

$$u(t,x)=\mathbb{E}\big[g(X_T^{t,x})Z_T^{t,x}\big]+\int_t^T\mathbb{E}\big[f(u(r,X_r^{t,x}))Z_r^{t,x}\big]\,dr, \tag{73}$$

and

$$\tilde u(t,x)=\mathbb{E}\big[\tilde g(X_T^{t,x})Z_T^{t,x}\big]+\int_t^T\mathbb{E}\Big[\tilde f(\tilde u(r,X_r^{t,x}))Z_r^{t,x}\Big]\,dr. \tag{74}$$

(ii) For all $t \in [0,T)$, $x \in \mathbb{R}^d$ we have that

$$\max_{i\in[0,d]\cap\mathbb{Z}} \|\Lambda_i(T-t)\mathrm{pr}_i(\tilde{u}(t,x)-u(t,x))\|_2 \leqslant 10c^2d^{2c}\varepsilon V^{3q+1}(t,x)\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2}). \tag{75}$$

Proof of Proposition 3.1. [42, Lemma 2.7] and the assumptions of Proposition 3.1 show (i).

Next, (58), Hölder's inequality, the fact that $\frac{1}{p_\mathrm{v}}+\frac{1}{p_\mathrm{z}} \leqslant 1$, (60), Jensen's inequality, the fact that $2q \leqslant p_\mathrm{v}$, and the fact that $c \leqslant V$ imply for all $i \in [0,d]\cap\mathbb{Z}$, $s\in[0,T)$, $t \in [s,T)$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\left\|\Lambda_i(T-t)\left\|g(X_T^{t,\tilde{x}})\mathrm{pr}_i(Z_T^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q}\\
&\leqslant \left\|\Lambda_i(T-t)\left\|V(T,X_T^{t,\tilde{x}})\mathrm{pr}_i(Z_T^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q}\\
&\leqslant \left\|\Lambda_i(T-t)\left\|V(T,X_T^{t,\tilde{x}})\right\|_{p_\mathrm{v}}\left\|\mathrm{pr}_i(Z_T^{t,\tilde{x}})\right\|_{p_\mathrm{z}}\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q}\\
&\leqslant \left\|\Lambda_i(T-t)V(t,\tilde{x})\frac{c}{\Lambda_i(T-t)}\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q}\\
&\leqslant c\left\|V(T,X_T^{s,x})\right\|_{p_\mathrm{v}}\\
&\leqslant V^2(s,x).
\end{aligned}\tag{76}$$

Next, Jensen's inequality, (60), the fact that $1\leqslant p_\mathrm{z}$, (58), and the fact that $c \leqslant V$ prove for all $i \in [0,d]\cap\mathbb{Z}$, $s\in[0,T)$, $t\in[s,T)$, $x\in\mathbb{R}^d$ that

$$\begin{aligned}
\int_t^T \left\|\left\|\Lambda_i(T-t)f(0)\mathrm{pr}_i(Z_r^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q} dr &\leqslant \int_t^T \left|\Lambda_i(T-t)f(0)\frac{c}{\Lambda_i(r-t)}\right| dr\\
&\leqslant c|f(0)|\int_t^T \frac{\sqrt{T-t}}{\sqrt{r-t}}\,dr\\
&= c|f(0)|\sqrt{T-t}(2\sqrt{r-t}|_{r=t}^T)\\
&\leqslant 2c|Tf(0)| \leqslant 2cV(s,x) \leqslant 2V^2(s,x).
\end{aligned}\tag{77}$$

Next, Hölder's inequality, the fact that $\frac{1}{2q}+\frac{1}{p_\mathrm{z}} \leqslant 1$, the triangle inequality, (60), the disintegration theorem, the flow property in (64), and the fact that $\sum_{\nu=0}^d L_\nu \leqslant c$ show for all $i\in[0,d]\cap\mathbb{Z}$, $s\in[0,T)$, $t\in[s,T)$, $x\in\mathbb{R}^d$ that

$$\begin{aligned}
&\int_t^T \left\|\Lambda_i(T-t)\sum_{\nu=0}^d L_\nu\Lambda_\nu(T)\left\|\mathrm{pr}_\nu(u(r,X_r^{t,\tilde{x}}))\mathrm{pr}_i(Z_r^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q}\\
&\leqslant \int_t^T \left\|\Lambda_i(T-t)\sum_{\nu=0}^d L_\nu\frac{\sqrt{T}}{\sqrt{T-r}}\Lambda_\nu(T-r)\left\|\mathrm{pr}_\nu(u(r,X_r^{t,\tilde{x}}))\right\|_{2q}\left\|\mathrm{pr}_i(Z_r^{t,\tilde{x}})\right\|_{p_\mathrm{z}}\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q}\\
&\leqslant \int_t^T \Lambda_i(T-t)\sum_{\nu=0}^d L_\nu\frac{\sqrt{T}}{\sqrt{T-r}}\Lambda_\nu(T-r)\left\|\left\|\mathrm{pr}_\nu(u(r,X_r^{t,\tilde{x}}))\right\|_{2q}\left\|\mathrm{pr}_i(Z_r^{t,\tilde{x}})\right\|_{p_\mathrm{z}}\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q}\\
&\leqslant \int_t^T \Lambda_i(T-t)\sum_{\nu=0}^d L_\nu\frac{\sqrt{T}}{\sqrt{T-r}}\Lambda_\nu(T-r)\left\|\left\|\mathrm{pr}_\nu(u(r,X_r^{t,\tilde{x}}))\right\|_{2q}\frac{c}{\Lambda_i(r-t)}\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q}\\
&\leqslant \int_t^T \Lambda_i(T-t)\sum_{\nu=0}^d L_\nu\frac{\sqrt{T}}{\sqrt{T-r}}\Lambda_\nu(T-r)\left\|\mathrm{pr}_\nu(u(r,X_r^{s,x}))\right\|_{2q}\frac{c}{\Lambda_i(r-t)}\,dr\\
&\leqslant \int_t^T \sqrt{T-t}c\frac{\sqrt{T}}{\sqrt{T-r}}\max_{\nu\in[0,d]\cap\mathbb{Z}}\left[\Lambda_\nu(T-r)\left\|\mathrm{pr}_\nu(u(r,X_r^{s,x}))\right\|_{2q}\right]\frac{c}{\sqrt{r-t}}\,dr\\
&\leqslant \int_t^T c^2T\max_{\nu\in[0,d]\cap\mathbb{Z}}\left[\Lambda_\nu(T-r)\left\|\mathrm{pr}_\nu(u(r,X_r^{s,x}))\right\|_{2q}\right]\frac{dr}{\sqrt{(T-r)(r-t)}}.
\end{aligned}\tag{78}$$

This, (73), the triangle inequality, (59), (76), and (77) imply for all $i \in [0,d] \cap \mathbb{Z}$, $s \in [0,T)$, $t \in [s,T)$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\Lambda_i(T-t)\left\|\mathrm{pr}_i(u(t,X_t^{s,x}))\right\|_{2q}\\
&= \left\| \Lambda_i(T-t)\mathbb{E}\Big[g(X_T^{t,\tilde{x}})\mathrm{pr}_i(Z_T^{t,\tilde{x}})\Big]\Big|_{\tilde{x}=X_t^{s,x}} + \Lambda_i(T-t)\int_t^T \mathbb{E}\big[f(u(r,X_r^{t,\tilde{x}}))\mathrm{pr}_i(Z_r^{t,\tilde{x}})\big]\Big|_{\tilde{x}=X_t^{s,x}}dr\right\|_{2q}\\
&\leqslant \left\|\Lambda_i(T-t)\left\|g(X_T^{t,\tilde{x}})\mathrm{pr}_i(Z_T^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q} + \int_t^T \left\|\left\|\Lambda_i(T-t)f(u(r,X_r^{t,\tilde{x}}))\mathrm{pr}_i(Z_r^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q}dr\\
&\leqslant \left\|\Lambda_i(T-t)\left\|g(X_T^{t,\tilde{x}})\mathrm{pr}_i(Z_T^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q} + \int_t^T \left\|\left\|\Lambda_i(T-t)f(0)\mathrm{pr}_i(Z_r^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q}dr\\
&\quad + \int_t^T \left\|\Lambda_i(T-t)\sum_{\nu=0}^d L_\nu\Lambda_\nu(T)\left\|\mathrm{pr}_\nu(u(r,X_r^{t,\tilde{x}}))\mathrm{pr}_i(Z_r^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_{2q}\\
&\leqslant V^2(s,x)+2V^2(s,x)+\int_t^T c^2T\max_{\nu\in[0,d]\cap\mathbb{Z}}\Big[\Lambda_\nu(T-r)\left\|\mathrm{pr}_\nu(u(r,X_r^{s,x}))\right\|_{2q}\Big]\frac{dr}{\sqrt{(T-r)(r-t)}}\\
&\leqslant 3V^2(s,x)+\int_t^T c^2T\max_{\nu\in[0,d]\cap\mathbb{Z}}\Big[\Lambda_\nu(T-r)\left\|\mathrm{pr}_\nu(u(r,X_r^{s,x}))\right\|_{2q}\Big]\frac{dr}{\sqrt{(T-r)(r-t)}}.
\end{aligned}\tag{79}$$

This proves for all $s \in [0,T)$, $t \in [s,T)$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\max_{\nu\in[0,d]\cap\mathbb{Z}}\Big[\Lambda_\nu(T-t)\left\|\mathrm{pr}_\nu(u(t,X_t^{s,x}))\right\|_{2q}\Big]\\
&\leqslant 3V^2(s,x)+\int_t^T c^2T\max_{\nu\in[0,d]\cap\mathbb{Z}}\Big[\Lambda_\nu(T-r)\left\|\mathrm{pr}_\nu(u(r,X_r^{s,x}))\right\|_{2q}\Big]\frac{dr}{\sqrt{(T-r)(r-t)}}.
\end{aligned}\tag{80}$$

This, (69), (60), the fact that $2q \leqslant p_{\mathrm{v}}$, and Grönwall's inequality (see [42, Corollary 2.4]) show for all $s \in [0,T)$, $t \in [s,T)$, $x \in \mathbb{R}^d$ that

$$\max_{\nu\in[0,d]\cap\mathbb{Z}}\Big[\Lambda_\nu(T-t)\left\|\mathrm{pr}_\nu(u(t,X_t^{s,x}))\right\|_{2q}\Big] \leqslant 6V^2(s,x)e^{86c^6T^3}.\tag{81}$$

Next, (68), Hölder's inequality, the fact that $\frac{1}{p_{\mathrm{v}}}+\frac{1}{p_{\mathrm{z}}} \leqslant 1$, (60), Jensen's inequality, and the fact that $\frac{1}{p_{\mathrm{v}}} \leqslant \frac{1}{2}$ imply for all $s \in [0,T)$, $t \in [s,T)$, $x \in \mathbb{R}^d$, $\nu \in [0,d]\cap\mathbb{Z}$ that

$$\begin{aligned}
&\Lambda_\nu(T-t)\left\|\left\|(\tilde{g}(X_T^{t,\tilde{x}})-g(X_T^{t,\tilde{x}}))\mathrm{pr}_\nu(Z_T^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_2\\
&\leqslant \Lambda_\nu(T-t)\left\|\left\|\varepsilon c d^c V(T,X_T^{t,\tilde{x}})\mathrm{pr}_\nu(Z_T^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_2\\
&\leqslant \Lambda_\nu(T-t)\left\|\varepsilon c d^c\left\|V(T,X_T^{t,\tilde{x}})\right\|_{p_{\mathrm{v}}}\left\|\mathrm{pr}_\nu(Z_T^{t,\tilde{x}})\right\|_{p_{\mathrm{z}}}\Big|_{\tilde{x}=X_t^{s,x}}\right\|_2\\
&\leqslant \Lambda_\nu(T-t)\left\|\varepsilon c d^c V(t,X_t^{s,x})\frac{c}{\Lambda_\nu(T-t)}\right\|_{p_{\mathrm{v}}}\\
&\leqslant \varepsilon c^2 d^c V(s,x).
\end{aligned}\tag{82}$$

Next, (59), Hölder's inequality, the fact that $\frac{1}{2}+\frac{1}{p_{\mathrm{z}}} \leqslant 1$, (60), the disintegration theorem, (64), and the fact that $\sum_{i=0}^d L_i \leqslant c$ prove for all $\nu \in [0,d]\cap\mathbb{Z}$, $s \in [0,T)$, $t \in [s,T)$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\int_t^T \Lambda_\nu(T-t)\left\|\left\|\tilde{f}(\tilde{u}(r,X_r^{t,\tilde{x}}))\mathrm{pr}_\nu(Z_r^{t,\tilde{x}})-\tilde{f}(u(r,X_r^{t,\tilde{x}}))\mathrm{pr}_\nu(Z_r^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \int_t^T \Lambda_\nu(T-t)\left\|\left\|\sum_{i=0}^d L_i\Lambda_i(T)\left|\mathrm{pr}_i(\tilde{u}(r,X_r^{t,\tilde{x}})-u(r,X_r^{t,\tilde{x}}))\right|\left|\mathrm{pr}_\nu(Z_r^{t,\tilde{x}})\right|\right\|_1\Bigg|_{\tilde{x}=X_t^{s,x}}\right\|_2 dr
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \int_t^T \Lambda_\nu(T-t)\sum_{i=0}^d L_i\frac{\sqrt{T}}{\sqrt{T-r}}\left\|\left\|\Lambda_i(T-r)\mathrm{pr}_i(\tilde{u}(r,X_r^{t,\tilde{x}})-u(r,X_r^{t,\tilde{x}}))\mathrm{pr}_\nu(Z_r^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \int_t^T \Lambda_\nu(T-t)\sum_{i=0}^d L_i\frac{\sqrt{T}}{\sqrt{T-r}}\left\|\left\|\Lambda_i(T-r)\mathrm{pr}_i(\tilde{u}(r,X_r^{t,\tilde{x}})-u(r,X_r^{t,\tilde{x}}))\right\|_2\left\|\mathrm{pr}_\nu(Z_r^{t,\tilde{x}})\right\|_{p_z}\Big|_{\tilde{x}=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \int_t^T \Lambda_\nu(T-t)\sum_{i=0}^d L_i\frac{\sqrt{T}}{\sqrt{T-r}}\left\|\left\|\Lambda_i(T-r)\mathrm{pr}_i(\tilde{u}(r,X_r^{t,\tilde{x}})-u(r,X_r^{t,\tilde{x}}))\right\|_2\frac{c}{\Lambda_\nu(r-t)}\Big|_{\tilde{x}=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \int_t^T \sqrt{T-t}c\frac{\sqrt{T}}{\sqrt{T-r}}\max_{i\in[0,d]\cap\mathbb{Z}}\left\|\Lambda_i(T-r)\mathrm{pr}_i(\tilde{u}(r,X_r^{s,x})-u(r,X_r^{s,x}))\right\|_2\frac{c}{\sqrt{r-t}}\,dr\\
&\leqslant \int_t^T c^2T\max_{i\in[0,d]\cap\mathbb{Z}}\left\|\Lambda_i(T-r)\mathrm{pr}_i(\tilde{u}(r,X_r^{s,x})-u(r,X_r^{s,x}))\right\|_2\frac{dr}{\sqrt{(T-r)(r-t)}}.
\end{aligned}\tag{83}
$$

Next, (60) shows for all $\nu\in[0,d]\cap\mathbb{Z}$, $s\in[0,T)$, $t\in[s,T)$, $x\in\mathbb{R}^d$ that

$$
\begin{aligned}
\int_t^T \Lambda_\nu(T-t)\left\|\left\|\varepsilon\mathrm{pr}_\nu(Z_r^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_2 dr &\leqslant \int_t^T \Lambda_\nu(T-t)\varepsilon\frac{c}{\Lambda_\nu(r-t)}dr\\
&\leqslant \int_t^T \varepsilon c\frac{\sqrt{T-t}}{\sqrt{r-t}}\,dr\\
&= \varepsilon c\sqrt{T-t}(2\sqrt{r-t}|_{r=t}^T)\\
&\leqslant 2\varepsilon cT.
\end{aligned}\tag{84}
$$

Next, the substitution $s=\frac{r-t}{T-t}$, $ds=\frac{dr}{T-t}$ implies for all $t\in[0,T)$ that

$$
\begin{aligned}
\int_t^T \frac{(T-t)^{\frac{1}{2}}T^{\frac{q}{2}}}{(T-r)^{\frac{q}{2}}(r-t)^{\frac{1}{2}}}\,dr &= \int_0^1 \frac{(T-t)^{\frac{1}{2}}T^{\frac{q}{2}}(T-t)}{[(T-t)(1-s)]^{\frac{q}{2}}[(T-t)s]^{\frac{1}{2}}}\,ds\\
&\leqslant T\int_0^1\frac{ds}{(1-s)^{\frac{q}{2}}s^{\frac{1}{2}}} = T\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2}).
\end{aligned}\tag{85}
$$

This, Hölder's inequality, the fact that $\frac{1}{2}+\frac{1}{p_z}\leqslant 1$, (60), the disintegration theorem, the flow property in (64), (81), and the fact that $6e^{86c^6T^3}\leqslant V$ prove for all $\nu\in[0,d]\cap\mathbb{Z}$, $s\in[0,T)$, $t\in[s,T)$, $x\in\mathbb{R}^d$ that

$$
\begin{aligned}
&\int_t^T \Lambda_\nu(T-t)\left\|\left\|\varepsilon\left(\sum_{i=0}^d(\Lambda_i(T))^q(\mathrm{pr}_i(u(r,X_r^{t,\tilde{x}})))^q\right)\mathrm{pr}_\nu(Z_r^{t,\tilde{x}})\right\|_1\Bigg|_{\tilde{x}=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \int_t^T \Lambda_\nu(T-t)\varepsilon\sum_{i=0}^d\frac{T^{\frac{q}{2}}}{(T-r)^{\frac{q}{2}}}(\Lambda_i(T-r))^q\left\|\left\|(\mathrm{pr}_i(u(r,X_r^{t,\tilde{x}})))^q\mathrm{pr}_\nu(Z_r^{t,\tilde{x}})\right\|_1\Big|_{\tilde{x}=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \int_t^T \Lambda_\nu(T-t)\varepsilon\sum_{i=0}^d\frac{T^{\frac{q}{2}}}{(T-r)^{\frac{q}{2}}}(\Lambda_i(T-r))^q\left\|\left\|(\mathrm{pr}_i(u(r,X_r^{t,\tilde{x}})))^q\right\|_2\left\|\mathrm{pr}_\nu(Z_r^{t,\tilde{x}})\right\|_{p_z}\Big|_{\tilde{x}=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \int_t^T \Lambda_\nu(T-t)\varepsilon\sum_{i=0}^d\frac{T^{\frac{q}{2}}}{(T-r)^{\frac{q}{2}}}(\Lambda_i(T-r))^q\left\|\left\|(\mathrm{pr}_i(u(r,X_r^{t,\tilde{x}})))^q\right\|_2\frac{c}{\Lambda_\nu(T-t)}\Big|_{\tilde{x}=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \int_t^T \Lambda_\nu(T-t)\varepsilon\sum_{i=0}^d\frac{T^{\frac{q}{2}}}{(T-r)^{\frac{q}{2}}}(\Lambda_i(T-r))^q\left\|\mathrm{pr}_i(u(r,X_r^{s,x}))\right\|_{2q}^q\frac{c}{\Lambda_\nu(r-t)}\,dr\\
&\leqslant \int_t^T \sqrt{T-t}\varepsilon(d+1)\frac{T^{\frac{q}{2}}}{(T-r)^{\frac{q}{2}}}\left[6V^2(s,x)e^{86c^6T^3}\right]^q\frac{c}{\sqrt{r-t}}\,dr\\
&\leqslant c\varepsilon(d+1)V^{3q}(s,x)\int_t^T\frac{(T-t)^{\frac{1}{2}}T^{\frac{q}{2}}}{(T-r)^{\frac{q}{2}}(r-t)^{\frac{1}{2}}}\,dr
\end{aligned}
$$

$$\leqslant c\varepsilon(d+1)V^{3q}(s,x)T\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2}). \tag{86}$$

This, (67), the triangle inequality, (84), and the fact that $\mathrm{B}(1-\frac{q}{2},\frac{1}{2})\geqslant 1$ show for all $\nu\in[0,d]\cap\mathbb{Z}$, $s\in[0,T)$, $t\in[s,T)$, $x\in\mathbb{R}^d$ that

$$\begin{aligned}
&\int_t^T \Lambda_\nu(T-t)\left\|\left\|\tilde f(u(r,X_r^{t,\tilde x}))\mathrm{pr}_\nu(Z_r^{t,\tilde x})-f(u(r,X_r^{t,\tilde x}))\mathrm{pr}_\nu(Z_r^{t,\tilde x})\right\|_1\Big|_{\tilde x=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \int_t^T \Lambda_\nu(T-t)\left\|\left\|\frac{\varepsilon c d^c}{T}\left(1+\sum_{\nu=0}^d(\Lambda_i(T))^q(\mathrm{pr}_\nu(u(r,X_r^{t,\tilde x})))^q\right)\mathrm{pr}_\nu(Z_r^{t,\tilde x})\right\|_1\Big|_{\tilde x=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \frac{cd^c}{T}\int_t^T \Lambda_\nu(T-t)\left\|\left\|\varepsilon\mathrm{pr}_\nu(Z_r^{t,\tilde x})\right\|_1\Big|_{\tilde x=X_t^{s,x}}\right\|_2 dr\\
&\quad+\frac{cd^c}{T}\int_t^T \Lambda_\nu(T-t)\left\|\left\|\varepsilon\left(\sum_{i=0}^d(\Lambda_\nu(T))^q(\mathrm{pr}_i(u(r,X_r^{t,\tilde x})))^q\right)\mathrm{pr}_\nu(Z_r^{t,\tilde x})\right\|_1\Big|_{\tilde x=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \frac{cd^c}{T}\left(2\varepsilon cT+c\varepsilon(d+1)V^{3q}(s,x)T\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\right)\\
&\leqslant cd^c c\varepsilon(d+3)V^{3q}(s,x)\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\\
&\leqslant 4dc^2d^c\varepsilon V^{3q}(s,x)\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\\
&\leqslant 4c^2d^{2c}\varepsilon V^{3q}(s,x)\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2}).
\end{aligned} \tag{87}$$

This, (73), (74), the triangle inequality, (82), (83), the fact that $V\leqslant V^{3q}$, and the fact that $\mathrm{B}(1-\frac{q}{2},\frac{1}{2})\geqslant 1$ prove for all $\nu\in[0,d]\cap\mathbb{Z}$, $s\in[0,T)$, $t\in[s,T)$, $x\in\mathbb{R}^d$ that

$$\begin{aligned}
&\Lambda_\nu(T-t)\left\|\mathrm{pr}_\nu(u(t,X_t^{s,x})-\tilde u(t,X_t^{s,x}))\right\|_2\\
&\leqslant \Lambda_\nu(T-t)\left\|\mathbb{E}\Big[\tilde g(X_T^{t,\tilde x})\mathrm{pr}_\nu(Z_T^{t,\tilde x})-g(X_T^{t,\tilde x})\mathrm{pr}_\nu(Z_T^{t,\tilde x})\Big]\Big|_{\tilde x=X_t^{s,x}}\right\|_2\\
&\quad+\int_t^T \Lambda_\nu(T-t)\left\|\left|\mathbb{E}\Big[\tilde f(\tilde u(r,X_r^{t,\tilde x}))\mathrm{pr}_\nu(Z_r^{t,\tilde x})\Big]-\mathbb{E}\big[f(u(r,X_r^{t,\tilde x}))\mathrm{pr}_\nu(Z_r^{t,\tilde x})\big]\right|\Big|_{\tilde x=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \Lambda_\nu(T-t)\left\|\left\|(\tilde g(X_T^{t,\tilde x})-g(X_T^{t,\tilde x}))\mathrm{pr}_\nu(Z_T^{t,\tilde x})\right\|_1\Big|_{\tilde x=X_t^{s,x}}\right\|_2\\
&\quad+\int_t^T \Lambda_\nu(T-t)\left\|\left\|\tilde f(\tilde u(r,X_r^{t,\tilde x}))\mathrm{pr}_\nu(Z_r^{t,\tilde x})-\tilde f(u(r,X_r^{t,\tilde x}))\mathrm{pr}_\nu(Z_r^{t,\tilde x})\right\|_1\Big|_{\tilde x=X_t^{s,x}}\right\|_2 dr\\
&\quad+\int_t^T \Lambda_\nu(T-t)\left\|\left\|\tilde f(u(r,X_r^{t,\tilde x}))\mathrm{pr}_\nu(Z_r^{t,\tilde x})-f(u(r,X_r^{t,\tilde x}))\mathrm{pr}_\nu(Z_r^{t,\tilde x})\right\|_1\Big|_{\tilde x=X_t^{s,x}}\right\|_2 dr\\
&\leqslant \varepsilon c^2d^cV(s,x)+\int_t^T c^2T\max_{i\in[0,d]\cap\mathbb{Z}}\left\|\Lambda_i(T-r)\mathrm{pr}_i(\tilde u(r,X_r^{s,x})-u(r,X_r^{s,x}))\right\|_2\frac{dr}{\sqrt{(T-r)(r-t)}}\\
&\quad+4c^2d^{2c}\varepsilon V^{3q}(s,x)\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\\
&\leqslant 5c^2d^{2c}\varepsilon V^{3q}(s,x)\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\\
&\quad+\int_t^T c^2T\max_{i\in[0,d]\cap\mathbb{Z}}\left\|\Lambda_i(T-r)\mathrm{pr}_i(\tilde u(r,X_r^{s,x})-u(r,X_r^{s,x}))\right\|_2\frac{dr}{\sqrt{(T-r)(r-t)}}.
\end{aligned} \tag{88}$$

Hence, we have for all $s\in[0,T)$, $t\in[s,T)$, $x\in\mathbb{R}^d$ that

$$\begin{aligned}
&\max_{i\in[0,d]\cap\mathbb{Z}}\left\|\Lambda_i(T-t)\mathrm{pr}_i(\tilde u(t,X_t^{s,x})-u(t,X_t^{s,x}))\right\|_2\\
&\leqslant 5c^2d^{2c}\varepsilon V^{3q}(s,x)\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\\
&\quad+\int_t^T c^2T\max_{i\in[0,d]\cap\mathbb{Z}}\left\|\Lambda_i(T-r)\mathrm{pr}_i(\tilde u(r,X_r^{s,x})-u(r,X_r^{s,x}))\right\|_2\frac{dr}{\sqrt{(T-r)(r-t)}}.
\end{aligned} \tag{89}$$

This, Grönwall's inequality (see [42, Corollary 2.4]), and the fact that $10e^{86c^6T^3} \leqslant V$ imply for all $s \in [0,T)$, $t \in [s,T)$, $x \in \mathbb{R}^d$ that

$$\max_{i\in[0,d]\cap\mathbb{Z}} \|\Lambda_i(T-t)\mathrm{pr}_i(\tilde{u}(t,X_t^{s,x}) - u(t,X_t^{s,x}))\|_2 \leqslant 10c^2d^{2c}\varepsilon V^{3q}(s,x)\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})e^{86c^6T^3} \leqslant 10c^2d^{2c}\varepsilon V^{3q+1}(s,x)\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2}). \tag{90}$$

This and the fact that $\forall\, s \in [0,T), x \in \mathbb{R}^d\colon \mathbb{P}(X_s^{s,x}) = 1$ complete the proof of Proposition 3.1. □

4. Proof of the main theorem

Proof of Theorem 1.2. First of all, we need to introduce an MLP setting. Let $\varrho\colon [0,T]^2 \to \mathbb{R}$ satisfy for all $t \in [0,T)$, $s \in (t,T)$ that

$$\varrho(t,s) = \frac{1}{\mathrm{B}(\frac{1}{2},\frac{1}{2})}\frac{1}{\sqrt{(T-s)(s-t)}}. \tag{91}$$

Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. For every $d \in \mathbb{N}$ let $W^{d,\theta}\colon [0,T]\times\Omega \to \mathbb{R}^d$, $\theta \in \Theta$, be standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motions with continuous sample paths. Let $\mathfrak{t}^\theta\colon \Omega \to (0,1)$, $\theta \in \Theta$, be i.i.d. random variables. Assume for all $b \in (0,1)$ that

$$\mathbb{P}(\mathfrak{t}^0 \leqslant b) = \frac{1}{\mathrm{B}(\frac{1}{2},\frac{1}{2})}\int_0^b \frac{dr}{\sqrt{r(1-r)}}. \tag{92}$$

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

$$\begin{aligned} U^{d,\theta}_{n,m,\varepsilon}(t,x) &= (g^d_\varepsilon(x),0) + \sum_{i=1}^{m^n}\left(1, \tfrac{W_T^{d,(\theta,0,-i)}-W_t^{d,(\theta,0,-i)}}{T-t}\right)^\top \tfrac{\left(g^d_\varepsilon(x+W_T^{d,(\theta,0,-i)}-W_t^{d,(\theta,0,-i)})-g^d_\varepsilon(x)\right)}{m^n} \\ &+ \sum_{\ell=0}^{n-1}\sum_{i=1}^{m^{n-\ell}}\left(1, \tfrac{W^{d,(\theta,\ell,i)}_{\mathfrak{T}_t^{(\theta,\ell,i)}}-W_t^{d,(\theta,\ell,i)}}{\mathfrak{T}_t^{(\theta,\ell,i)}-t}\right)^\top \tfrac{\left(f^d_\varepsilon\circ U^{d,(\theta,\ell,i)}_{\ell,m,\varepsilon}-\mathbb{1}_{\mathbb{N}}(\ell)f^d_\varepsilon\circ U^{d,(\theta,-\ell,i)}_{\ell-1,m,\varepsilon}\right)(\mathfrak{T}_t^{(\theta,\ell,i)},x+W^{d,(\theta,\ell,i)}_{\mathfrak{T}_t^{(\theta,\ell,i)}}-W_t^{d,(\theta,\ell,i)})}{m^{n-\ell}\varrho(t,\mathfrak{T}_t^{(\theta,\ell,i)})}. \end{aligned} \tag{93}$$

For every $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $s \in [0,T]$, $t \in [s,T]$ let

$$a(d) = 4 + 48e^{86c^6T^3} + 8\beta d + d^{2c}, \quad \varphi_d(x) = (a(d) + \|x\|^2)^{8\beta}, \tag{94}$$

$$V_d(t,x) = ce^{\beta(T-t)}(\varphi_d(x))^{\frac{1}{8}}, \tag{95}$$

$$X_t^{d,\theta,s,x} = x + W_t^{d,\theta} - W_s^{d,\theta}. \tag{96}$$

For every $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $s \in [0,T)$, $t \in (s,T]$ let

$$Z_t^{d,\theta,s,x} = \left(1, \frac{W_t^{d,\theta}-W_s^{d,\theta}}{t-s}\right)^\top. \tag{97}$$

The fact that $8\beta d \leqslant \varphi^{\frac{1}{8\beta}}$ and [32, Lemma 2.6](applied for every $d \in \mathbb{N}$ with $m \leftarrow d$, $d \leftarrow d$, $p \leftarrow 8\beta$, $a \leftarrow a(d)$, $c \leftarrow 1$, $\mu \leftarrow 0$, $\sigma \leftarrow \mathrm{Id}_{\mathbb{R}^{d\times d}}$ in the notation of [32, Lemma 2.6]) prove for all $d \in \mathbb{N}$ that $\frac{1}{2}\sum_{k=1}^d \frac{\partial^2\varphi_d}{\partial x_k^2} \leqslant 16\beta\varphi_d$ and hence $\frac{1}{2}\sum_{k=1}^d \frac{\partial^2 V_d^8}{\partial x_k^2} \leqslant 16\beta V_d^8$. This and the fact that $\forall\, d \in \mathbb{N}\colon \frac{\partial V_d^8}{\partial t} = -8\beta V_d^8$ show for all $d \in \mathbb{N}$ that

$$\frac{\partial V_d^8}{\partial t} + \frac{1}{2}\sum_{k=1}^d \frac{\partial^2 V_d^8}{\partial x_k^2} \leqslant -8\beta V_d^8 + 16\beta V_d^8 = 8\beta V_d^8. \tag{98}$$

This and [12, Lemma 2.2] (applied for every $s \in [0,T]$, $t \in [s,T]$ with $V \leftarrow V_d^8$, $\alpha \leftarrow 8\beta$, $\tau \leftarrow T$, $\mu \leftarrow 0$, $\sigma \leftarrow \mathrm{Id}_{\mathbb{R}^{d\times d}}$, $t \leftarrow t-s$ in the notation of [12, Lemma 2.2]) imply for all $d \in \mathbb{N}$, $s \in [0,T]$, $t \in [s,T]$ that

$$\begin{aligned} e^{8\beta(T-t+s)}\mathbb{E}\Big[\varphi_d(X_t^{d,0,s,x})\Big] &= \mathbb{E}\Big[e^{8\beta(T-(t-s))}\varphi_d(x+W_t^{d,0}-W_s^{d,0})\Big] \\ &= \mathbb{E}\Big[e^{8\beta(T-(t-s))}\varphi_d(x+W_{t-s}^{d,0})\Big] \\ &= c^{-8}\mathbb{E}\Big[V_d^8(t-s,x+W_{t-s}^{d,0})\Big] \\ &\leqslant c^{-8}V_d^8(0,x) = e^{8\beta T}\varphi_d(x), \end{aligned} \tag{99}$$

hence

$$\mathbb{E}\Big[V_d^8(t,X_t^{d,0,s,x})\Big] = \mathbb{E}\Big[c^8 e^{8\beta(T-t)}\varphi_d(X_t^{d,0,s,x})\Big] \leqslant c^8 e^{8\beta(T-s)}\varphi_d(x) = V_d^8(s,x), \tag{100}$$

and hence

$$\left\|V_d(t,X_t^{d,0,s,x})\right\|_8 \leqslant V_d(s,x). \tag{101}$$

Next, recalling the fact that $c \geqslant 2$ proves for all $d \in \mathbb{N}$, $i \in [1,d] \cap \mathbb{Z}$, $t \in [0,T)$, $r \in (0,T]$ that

$$\left\|\frac{W_r^{d,0,i}-W_t^{d,0,i}}{r-t}\right\|_8 = \frac{1}{\sqrt{r-t}}\left\|\frac{W_r^{d,0,i}-W_t^{d,0,i}}{\sqrt{r-t}}\right\|_8 \leqslant \frac{c}{\sqrt{r-t}}, \tag{102}$$

where $W^{d,0,i}$ is the i-th coordinate of the Brownian motion $W^{d,0}$. This and (97) demonstrate for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $i \in [0,d] \cap \mathbb{Z}$, $t \in [0,T)$, $r \in (0,T]$ that

$$\left\|\mathrm{pr}_i^d(Z_r^{d,0,t,x})\right\|_8 \leqslant \frac{c}{\Lambda_i^d(r-t)}. \tag{103}$$

Next, the triangle inequality implies for all $d \in \mathbb{N}$, $i \in [1,d] \cap \mathbb{N}$, $s \in [0,T)$, $t \in [s,T)$, $r \in (s,T]$ that

$$\begin{aligned} &\left\|\frac{W_r^{d,0,i}-W_t^{d,0,i}}{r-t} - \frac{W_r^{d,0,i}-W_s^{d,0,i}}{r-s}\right\|_8 \\ &= \left\|\left(\frac{1}{r-t}-\frac{1}{r-s}\right)(W_r^{d,0,i}-W_t^{d,0,i}) + \frac{1}{r-s}\left(W_r^{d,0,i}-W_t^{d,0,i}-\left(W_r^{d,0,i}-W_s^{d,0,i}\right)\right)\right\|_8 \\ &\leqslant \frac{t-s}{(r-t)(r-s)}\left\|W_r^{d,0,i}-W_t^{d,0,i}\right\|_8 + \frac{1}{r-s}\left\|W_t^{d,0,i}-W_s^{d,0,i}\right\|_8 \\ &\leqslant \frac{t-s}{(r-t)(r-s)}105^{\frac{1}{8}}\sqrt{r-t} + \frac{1}{r-s}105^{\frac{1}{8}}\sqrt{t-s} \\ &\leqslant \frac{105^{\frac{1}{8}}\sqrt{t-s}}{\sqrt{r-t}\sqrt{r-s}} + \frac{105^{\frac{1}{8}}\sqrt{t-s}}{\sqrt{r-t}\sqrt{r-s}} \\ &\leqslant \frac{4\sqrt{t-s}}{\sqrt{r-t}\sqrt{r-s}}. \end{aligned} \tag{104}$$

This, (97), and the fact that $\forall\, d \in \mathbb{N}\colon 4 \leqslant V_d$ show for all $d \in \mathbb{N}$, $s \in [0,T)$, $t \in [s,T)$, $r \in (s,T]$ that

$$\left\|\mathrm{pr}_i^d(Z_r^{d,0,t,x}-Z_r^{d,0,s,x})\right\|_8 \leqslant \frac{V_d(t,x)+V_d(s,x)}{2}\frac{\sqrt{t-s}}{\sqrt{r-t}\Lambda_i^d(r-s)}. \tag{105}$$

Next, (5)–(9), (94), and (95) imply for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$, $x \in \mathbb{R}^d$, $w_1, w_2 \in \mathbb{R}^{d+1}$ that

$$\max\{|g^d(x)|, |g_\varepsilon^d(x)|\} \leqslant V_d(T,x), \quad \max\{|Tf^d(0)|, |Tf_\varepsilon^d(0)|\} \leqslant V_d(t,x), \tag{106}$$

$$\max\{|f^d(w_1)-f^d(w_2)|, |f_\varepsilon^d(w_1)-f_\varepsilon^d(w_2)|\} \leqslant \sum_{\nu=0}^{d}\left[L_\nu^d\Lambda_\nu^d(T)|\mathrm{pr}_\nu^d(w_1-w_2)|\right], \tag{107}$$

$$\max\{|g^d(x)-g^d(y)|,|g^d_\varepsilon(x)-g^d_\varepsilon(y)|\} \leqslant \frac{V_d(T,x)+V_d(T,y)}{2}\frac{\|x-y\|}{\sqrt{T}}, \tag{108}$$

$$\left|g^d(x)-g^d_\varepsilon(x)\right| \leqslant \varepsilon c d^c V_d(T,x). \tag{109}$$

Proposition 3.1 (applied for every $d\in\mathbb{N}$, $\varepsilon\in(0,1)$ with $p_{\mathrm{v}}\leftarrow 8$, $p_{\mathrm{z}}\leftarrow 8$, $p_{\mathrm{x}}\leftarrow 8$, $(L_i)_{i\in[0,d]\cap\mathbb{Z}}\leftarrow (L_i^d)_{i\in[0,d]\cap\mathbb{Z}}$, $(\Lambda_i)_{i\in[0,d]\cap\mathbb{Z}}\leftarrow(\Lambda_i^d)_{i\in[0,d]\cap\mathbb{Z}}$, $\mathrm{pr}\leftarrow\mathrm{pr}^d$, $f\leftarrow f^d$, $\tilde f\leftarrow f^d_\varepsilon$, $g\leftarrow g^d$, $\tilde g\leftarrow g^d_\varepsilon$, $V\leftarrow V_d$, $X\leftarrow X^{d,0}$, $Z\leftarrow Z^{d,0}$ in the notation of Proposition 3.1), (106), (107), (101), (103), (108), (96), (97), (105), (8), and (109) show that the following items hold.

(a) For all $d\in\mathbb{N}$, $\varepsilon\in(0,1)$ there exist unique continuous functions $u^d,u^d_\varepsilon\colon[0,T)\times\mathbb{R}^d\to\mathbb{R}^{d+1}$ such that for all $t\in[0,T)$, $x\in\mathbb{R}^d$ we have that

$$\max_{\nu\in[0,d]\cap\mathbb{Z}}\sup_{\tau\in[0,T),\xi\in\mathbb{R}^d}\left[\Lambda^d_\nu(T-\tau)\frac{|\mathrm{pr}^d_\nu(u^d(\tau,\xi))|}{V_d(\tau,\xi)}\right]<\infty, \tag{110}$$

$$\max_{\nu\in[0,d]\cap\mathbb{Z}}\sup_{\tau\in[0,T),\xi\in\mathbb{R}^d}\left[\Lambda^d_\nu(T-\tau)\frac{|\mathrm{pr}^d_\nu(u^d_\varepsilon(\tau,\xi))|}{V_d(\tau,\xi)}\right]<\infty, \tag{111}$$

$$\max_{\nu\in[0,d]\cap\mathbb{Z}}\left[\mathbb{E}\Big[\Big|g^d(X^{d,0,t,x}_T)\mathrm{pr}^d_\nu(Z^{d,0,t,x}_T)\Big|\Big]+\int_t^T\mathbb{E}\big[\big|f^d(u^d(r,X^{d,0,t,x}_r))\mathrm{pr}^d_\nu(Z^{d,0,t,x}_r)\big|\big]\,dr\right]<\infty, \tag{112}$$

$$\max_{\nu\in[0,d]\cap\mathbb{Z}}\left[\mathbb{E}\Big[\Big|g^d_\varepsilon(X^{d,0,t,x}_T)\mathrm{pr}^d_\nu(Z^{d,0,t,x}_T)\Big|\Big]+\int_t^T\mathbb{E}\big[\big|f^d_\varepsilon(u^d_\varepsilon(r,X^{d,0,t,x}_r))\mathrm{pr}_\nu(Z^{d,0,t,x}_r)\big|\big]\,dr\right]<\infty, \tag{113}$$

$$u^d(t,x)=\mathbb{E}\Big[g^d(X^{d,0,t,x}_T)Z^{d,0,t,x}_T\Big]+\int_t^T\mathbb{E}\big[f^d(u^d(r,X^{d,0,t,x}_r))Z^{d,0,t,x}_r\big]\,dr, \tag{114}$$

and

$$u^d_\varepsilon(t,x)=\mathbb{E}\Big[g^d_\varepsilon(X^{d,0,t,x}_T)Z^{d,0,t,x}_T\Big]+\int_t^T\mathbb{E}\big[f^d_\varepsilon(u^d_\varepsilon(r,X^{d,0,t,x}_r))Z^{d,0,t,x}_r\big]\,dr. \tag{115}$$

(b) For all $t\in[0,T)$, $x\in\mathbb{R}^d$ we have that

$$\max_{i\in[0,d]\cap\mathbb{Z}}\left\|\Lambda^d_i(T-t)\mathrm{pr}^d_i(u^d_\varepsilon(t,x)-u^d(t,x))\right\|_2\leqslant 10c^2d^{2c}\varepsilon V_d^{3q+1}(t,x)\mathrm{B}(1-\tfrac{q}{2},\tfrac12). \tag{116}$$

Item (a), [43, Theorem 6.9] (applied for every $d\in\mathbb{N}$ with $\mu^d\leftarrow 0$, $\sigma^d\leftarrow \mathrm{Id}_{\mathbb{R}^{d\times d}}$), (6), and (7) imply that for all $d\in\mathbb{N}$ the function u^d satisfies that $v^d:=\mathrm{pr}^d_0(u^d)$ is the unique viscosity solution to the following semilinear PDE of parabolic type:

$$\frac{\partial v^d}{\partial t}(t,x)+\frac12(\Delta v^d)(t,x)+f^d(t,x,v^d(t,x),(\nabla_x v^d)(t,x))=0\quad\forall\, t\in(0,T),x\in\mathbb{R}^d, \tag{117}$$

$$v^d(T,x)=g^d(x)\quad\forall\, x\in\mathbb{R}^d, \tag{118}$$

$\nabla_x v^d=(\mathrm{pr}^d_1(u^d),\mathrm{pr}^d_2(u^d),\ldots,\mathrm{pr}^d_d(u^d))$, and

$$\max_{\nu\in[0,d]\cap\mathbb{Z}}\sup_{\tau\in[0,T),\xi\in\mathbb{R}^d}\left[\Lambda^d_\nu(T-\tau)\frac{|\mathrm{pr}^d_\nu(u^d(\tau,\xi))|}{(1+\|\xi\|^2)^{\frac12}}\right]<\infty. \tag{119}$$

This establishes (i).

Next, [42, Lemma 4.3] (applied for every $d\in\mathbb{N}$, $\varepsilon\in(0,1)$ with $p_{\mathrm{v}}\leftarrow 8$, $p_{\mathrm{z}}\leftarrow 8$, $p_{\mathrm{x}}\leftarrow 8$, $(L_i)_{i\in[0,d]\cap\mathbb{Z}}\leftarrow(L^d_i)_{i\in[0,d]\cap\mathbb{Z}}$, $(\Lambda_i)_{i\in[0,d]\cap\mathbb{Z}}\leftarrow(\Lambda^d_i)_{i\in[0,d]\cap\mathbb{Z}}$, $\mathrm{pr}\leftarrow\mathrm{pr}^d$, $f\leftarrow f^d_\varepsilon$, $g\leftarrow g^d_\varepsilon$, $V\leftarrow V_d$, $X\leftarrow X^{d,0}$, $Z\leftarrow Z^{d,0}$, $(U^\theta_{n,m})\leftarrow(U^{d,\theta}_{n,m,\varepsilon})$, $q_1\leftarrow 3$ in the notation of [42, Lemma 4.3]) demonstrate for all $d,m,n\in\mathbb{N}$, $t\in[0,T)$, $x\in\mathbb{R}^d$, $\varepsilon\in(0,1)$ that

$$\max_{\nu\in[0,d]\cap\mathbb{Z}}\left\|\Lambda^d_\nu(T-t)\mathrm{pr}^d_\nu(U^{d,0}_{n,m,\varepsilon}(t,x)-u^d_\varepsilon(t,x))\right\|_2\leqslant e^{\frac{m^3}{6}}m^{-\frac{n}{2}}8^ne^{nc^2T}V_d^3(t,x). \tag{120}$$

Moreover, (94) and (95) show that there exists $\kappa \in (0,\infty)$ such that

$$\sqrt{d+1}\left(\int_{[0,1]^d} V_d^6(0,x)\,dx\right)^{\frac{1}{2}} + \sqrt{d+1}\left(\int_{[0,1]^d} V_d^{6q+2}(0,x)\,dx\right)^{\frac{1}{2}} \leqslant \kappa d^\kappa. \tag{121}$$

This, the triangle inequality, (120), and (116) prove for all $d,n,m \in \mathbb{N}$, $\varepsilon \in (0,1)$ that

$$\begin{aligned}
&\left(\int_{[0,1]^d} \sum_{\nu=0}^{d} \left\|\Lambda_\nu^d(T)\mathrm{pr}_\nu^d(U^{d,0}_{n,m,\varepsilon}(0,x) - u^d(0,x))\right\|_2^2 dx\right)^{\frac{1}{2}}\\
&\leqslant \left(\int_{[0,1]^d} (d+1) \max_{\nu\in[0,d]\cap\mathbb{Z}} \left\|\Lambda_\nu^d(T)\mathrm{pr}_\nu^d(U^{d,0}_{n,m,\varepsilon}(0,x) - u^d(0,x))\right\|_2^2 dx\right)^{\frac{1}{2}}\\
&\leqslant \sqrt{d+1}\left(\int_{[0,1]^d} \max_{\nu\in[0,d]\cap\mathbb{Z}} \left\|\Lambda_\nu^d(T)\mathrm{pr}_\nu^d(U^{d,0}_{n,m,\varepsilon}(0,x) - u_\varepsilon^d(0,x))\right\|_2^2 dx\right)^{\frac{1}{2}}\\
&\quad + \sqrt{d+1}\left(\int_{[0,1]^d} \max_{\nu\in[0,d]\cap\mathbb{Z}} \left\|\Lambda_\nu^d(T)\mathrm{pr}_\nu^d(u_\varepsilon^d(0,x) - u^d(0,x))\right\|_2^2 dx\right)^{\frac{1}{2}}\\
&\leqslant e^{\frac{m^3}{6}} m^{-\frac{n}{2}} 8^n e^{nc^2T}\sqrt{d+1}\left(\int_{[0,1]^d} V_d^6(0,x)\,dx\right)^{\frac{1}{2}} + 10c^2d^{2c}\varepsilon \mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\sqrt{d+1}\left(\int_{[0,1]^d} V_d^{6q+2}(0,x)\,dx\right)^{\frac{1}{2}}\\
&\leqslant \left(e^{\frac{m^3}{6}} m^{-\frac{n}{2}} 8^n e^{nc^2T} + 10c^2d^{2c}\varepsilon \mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\right)\kappa d^\kappa.
\end{aligned} \tag{122}$$

For the next step for every $d \in \mathbb{N}$, $\epsilon \in (0,1)$ let $N_{d,\epsilon} \in \mathbb{N}$, $\varepsilon_{d,\epsilon} \in (0,1)$ satisfies that

$$N_{d,\epsilon} = \inf\left\{n \in \mathbb{N} \cap [2,\infty) \colon e^{\frac{n^3}{6}} n^{-\frac{n}{2}} 8^n e^{nc^2T}\kappa d^\kappa \leqslant \frac{\epsilon}{2}\right\} \tag{123}$$

and

$$\varepsilon_{d,\epsilon} = \frac{\epsilon}{10c^2d^{2c}\mathrm{B}(1-\frac{q}{2},\frac{1}{2})\kappa d^\kappa}. \tag{124}$$

Then Fubini's theorem and (122) imply for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{[0,1]^d} \sum_{\nu=0}^{d} \left|\Lambda_\nu^d(T)\mathrm{pr}_\nu^d(U^{d,0}_{N_{d,\epsilon},N_{d,\epsilon},\varepsilon_{d,\epsilon}}(0,x) - u^d(0,x))\right|^2 dx\right]\right)^{\frac{1}{2}}\\
&= \left(\int_{[0,1]^d} \sum_{\nu=0}^{d} \mathbb{E}\left[\left|\Lambda_\nu^d(T)\mathrm{pr}_\nu^d(U^{d,0}_{N_{d,\epsilon},N_{d,\epsilon},\varepsilon_{d,\epsilon}}(0,x) - u^d(0,x))\right|^2\right] dx\right)^{\frac{1}{2}}\\
&\leqslant \left(e^{\frac{N_{d,\epsilon}^3}{6}} N_{d,\epsilon}^{-\frac{N_{d,\epsilon}}{2}} 8^{N_{d,\epsilon}} e^{N_{d,\epsilon}c^2T} + 10c^2d^{2c}\varepsilon_{d,\epsilon}\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\right)\kappa d^\kappa\\
&\leqslant \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned} \tag{125}$$

This shows that for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ that there exists $\omega_{d,\epsilon} \in \Omega$ such that

$$\int_{[0,1]^d} \sum_{\nu=0}^{d} \left|\Lambda_\nu^d(T)\mathrm{pr}_\nu^d(U^{d,0}_{N_{d,\epsilon},N_{d,\epsilon},\varepsilon_{d,\epsilon}}(0,x) - u^d(0,x))\right|^2 dx \leqslant \epsilon^2. \tag{126}$$

Next, Proposition 2.12 (applied for every $d \in \mathbb{N}$, $\epsilon \in (0,1)$ with $f \leftarrow f^d_{\varepsilon_{d,\epsilon}}$, $g \leftarrow g^d_{\varepsilon_{d,\epsilon}}$, $(U^\theta_{n,m}) \leftarrow (U^{d,\theta}_{n,m,\varepsilon_{d,\epsilon}})$, $\omega \leftarrow \omega_{d,\epsilon}$ in the notation of Proposition 2.12) imply for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ that there exists $\Psi_{d,\epsilon} \in \mathbf{N}$ such that

$$\dim(\mathcal{D}(\Psi_{d,\epsilon})) = N_{d,\epsilon}(\dim(\mathcal{D}(\Phi_{f^d_{\varepsilon_{d,\epsilon}}})) - 1) + \dim(\mathcal{D}(\Phi_{g^d_{\varepsilon_{d,\epsilon}}})), \tag{127}$$

$$|||\mathcal{D}(\Psi_{d,\epsilon})||| \leqslant \max\left\{d+1, \left|\left|\left|\mathcal{D}(\Phi_{g^d_{\varepsilon_{d,\epsilon}}})\right|\right|\right|, \left|\left|\left|\mathcal{D}(\Phi_{f^d_{\varepsilon_{d,\epsilon}}})\right|\right|\right|\right\}(4N_{d,\epsilon})^{N_{d,\epsilon}}, \tag{128}$$

$$U^{d,0}_{N_{d,\epsilon},N_{d,\epsilon},\varepsilon_{d,\epsilon}}(0,x,\omega_{d,\epsilon}) = (\mathcal{R}(\Psi_{d,\epsilon}))(x). \tag{129}$$

Furthermore, (10) and the fact that $\forall\, \Phi \in \mathbf{N}\colon \max\{\dim(\mathcal{D}(\Phi)), |||\mathcal{D}(\Phi)|||\} \leqslant \mathcal{P}(\Phi)$ imply for all $d \in \mathbb{N}$, $\varepsilon \in (0,1)$ that

$$\max\{\dim(\mathcal{D}(\Phi_{g^d_\varepsilon})), \dim(\mathcal{D}(\Phi_{f^d_\varepsilon})), |||\mathcal{D}(\Phi_{g^d_\varepsilon})|||, |||\mathcal{D}(\Phi_{f^d_\varepsilon})|||\} \leqslant cd^c\varepsilon^{-c}. \tag{130}$$

Hence, we have for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ that

$$\max\Big\{\dim(\mathcal{D}(\Phi_{g^d_{\varepsilon_{d,\epsilon}}})), \dim(\mathcal{D}(\Phi_{f^d_{\varepsilon_{d,\epsilon}}})), \Big|\Big|\Big|\mathcal{D}(\Phi_{g^d_{\varepsilon_{d,\epsilon}}})\Big|\Big|\Big|, \Big|\Big|\Big|\mathcal{D}(\Phi_{f^d_{\varepsilon_{d,\epsilon}}})\Big|\Big|\Big|\Big\} \leqslant cd^c\varepsilon^{-c}_{d,\epsilon}. \tag{131}$$

This, (127), (128), and the fact that $c \geqslant 2$ prove for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ that

$$\dim(\mathcal{D}(\Psi_{d,\epsilon})) \leqslant 2N_{d,\epsilon}cd^c\varepsilon^{-c}_{d,\epsilon} \quad \text{and} \quad |||\mathcal{D}(\Psi_{d,\epsilon})||| \leqslant cd^c\varepsilon^{-c}_{d,\epsilon}(4N_{d,\epsilon})^{N_{d,\epsilon}}. \tag{132}$$

Next, (123) implies for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ that

$$\begin{aligned}\epsilon &\leqslant e^{\frac{(N_{d,\epsilon}-1)^3}{6}}(N_{d,\epsilon}-1)^{-\frac{N_{d,\epsilon}-1}{2}}8^{N_{d,\epsilon}-1}e^{(N_{d,\epsilon}-1)c^2T}\kappa d^\kappa \\ &\leqslant e^{\frac{N^3_{d,\epsilon}}{6}}(N_{d,\epsilon}-1)^{-\frac{N_{d,\epsilon}-1}{2}}8^{N_{d,\epsilon}}e^{N_{d,\epsilon}c^2T}\kappa d^\kappa.\end{aligned} \tag{133}$$

This, the fact that $\forall\, \Phi \in \mathbf{N}\colon \mathcal{P}(\Phi) \leqslant 2\dim(\mathcal{D}(\Phi))|||\mathcal{D}(\Phi)|||^2$, (132), and (124) show for all $d \in \mathbb{N}$, $\epsilon, \gamma \in (0,1)$ that

$$\begin{aligned}\mathcal{P}(\Psi_{d,\epsilon}) &\leqslant 2\dim(\mathcal{D}(\Psi_{d,\epsilon}))|||\mathcal{D}(\Psi_{d,\epsilon})|||^2 \\ &\leqslant 2N_{d,\epsilon}cd^c\varepsilon^{-c}_{d,\epsilon}\left(cd^c\varepsilon^{-c}_{d,\epsilon}(4N_{d,\epsilon})^{N_{d,\epsilon}}\right)^2 \\ &= 32c^3d^{3c}\varepsilon^{-3c}_{d,\epsilon}N^{2N_{d,\epsilon}+1}_{d,\epsilon} \\ &\leqslant 32c^3d^{3c}\left(\frac{\epsilon}{10c^2d^{2c}\mathrm{B}(1-\frac{q}{2},\frac{1}{2})\kappa d^\kappa}\right)^{-3c}N^{2N_{d,\epsilon}+1}_{d,\epsilon} \\ &= 32c^3d^{3c}\left(10c^2d^{2c}\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\kappa d^\kappa\right)^{3c}\epsilon^{-3c}N^{2N_{d,\epsilon}+1}_{d,\epsilon}\epsilon^{4+\gamma}\epsilon^{-4-\gamma} \\ &\leqslant 32c^3d^{3c}\left(10c^2d^{2c}\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\kappa d^\kappa\right)^{3c}\epsilon^{-3c}N^{2N_{d,\epsilon}+1}_{d,\epsilon} \\ &\quad\left[e^{\frac{N^3_{d,\epsilon}}{6}}(N_{d,\epsilon}-1)^{-\frac{N_{d,\epsilon}-1}{2}}8^{N_{d,\epsilon}}e^{N_{d,\epsilon}c^2T}\kappa d^\kappa\right]^{4+\gamma}\epsilon^{-4-\gamma} \\ &= 32c^3d^{3c}\left(10c^2d^{2c}\mathrm{B}(1-\tfrac{q}{2},\tfrac{1}{2})\kappa d^\kappa\right)^{3c}(\kappa d^\kappa)^{4+\gamma}\epsilon^{-3c-4-\gamma}\sup_{n\in[2,\infty)\cap\mathbb{Z}}\frac{n^{2n+1}(e^{\frac{n^3}{6}}8^ne^{nc^2T})^{4+\gamma}}{(n-1)^{\frac{n-1}{2}(\gamma+4)}}.\end{aligned} \tag{134}$$

Next, note that for all $\gamma \in (0,1)$ we have that

$$\begin{aligned}&\sup_{n\in[2,\infty)\cap\mathbb{Z}}\frac{n^{2n+1}(e^{\frac{n^3}{6}}8^ne^{nc^2T})^{4+\gamma}}{(n-1)^{\frac{n-1}{2}(\gamma+4)}} \\ &\leqslant \sup_{n\in[2,\infty)\cap\mathbb{Z}}\frac{n^{2n+1}(e^{\frac{n^3}{6}}8^ne^{nc^2T})^{4+\gamma}n^{\frac{\gamma+4}{2}}}{(n-1)^{\frac{n}{2}(\gamma+4)}} \\ &= \sup_{n\in[2,\infty)\cap\mathbb{Z}}\left[\frac{n^{2n+1}(e^{\frac{n^3}{6}}8^ne^{nc^2T})^{4+\gamma}n^{\frac{\gamma+4}{2}}}{n^{\frac{n}{2}(\gamma+4)}}\left(\frac{n}{n-1}\right)^{\frac{n}{2}(\gamma+4)}\right] \\ &\leqslant \sup_{n\in[2,\infty)\cap\mathbb{Z}}\left[\frac{n(e^{\frac{n^3}{6}}8^ne^{nc^2T})^{4+\gamma}n^{\frac{\gamma+4}{2}}}{n^{\frac{n\gamma}{2}}}2^{\frac{n}{2}(\gamma+4)}\right] \\ &< \infty.\end{aligned} \tag{135}$$

This and (134) prove that there exists $\eta \in (0,\infty)$ such that for all $d \in \mathbb{N}$, $\epsilon \in (0,1)$ we have that $\mathcal{P}(\Psi_{d,\epsilon}) \leqslant \eta d^\eta\epsilon^{-\eta}$. This, (126), and (129) complete the proof of Theorem 1.2. □

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] 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.

[3] 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).

[4] 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.

[5] Beck, C., Becker, S., Grohs, P., Jaafari, N., and Jentzen, A. Solving the Kolmogorov PDE by Means of Deep Learning. *Journal of Scientific Computing 88*, 73 (2021).

[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., 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).

[8] 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.

[9] 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.

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

[11] 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).

[12] Cox, S., Hutzenthaler, M., and Jentzen, A. Local Lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations. *arXiv:1309.5595* (2021).

[13] 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* (2017), 349–380.

[14] 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.

[15] E, W., and Yu, B. The Deep Ritz method: A deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics 6*, 1 (2018), 1–12.

[16] 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.

[17] 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.

[18] Fujii, M., Takahashi, A., and Takahashi, M. Asymptotic Expansion as Prior Knowledge in Deep Learning Method for high dimensional BSDEs. *Asia-Pacific Financial Markets 26* (2019), 391–408.

[19] 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.

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

[21] 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.

[22] 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.

[23] 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.

[24] 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).

[25] 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.

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

[27] HAN, J., ZHANG, L., AND E, W. Solving many-electron Schrödinger equation using deep neural networks. *Journal of Computational Physics 399* (2019), 108929.

[28] HENRY-LABORDÈRE, P. Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM. *Available at SSRN: http://dx.doi.org/10.2139/ssrn.3071506* (2017).

[29] HURÉ, C., PHAM, H., AND WARIN, X. Deep backward schemes for high-dimensional nonlinear pdes. *Mathematics of Computation 89*, 324 (2020), 1547–1579.

[30] 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.

[31] 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*, 10 (2020).

[32] 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.

[33] 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.

[34] JACQUIER, A., AND OUMGARI, M. Deep curve-dependent PDEs for affine rough volatility. *SIAM Journal on Financial Mathematics 14*, 2 (2023), 353–382.

[35] JACQUIER, A., AND ZURIC, Z. Random neural networks for rough volatility. *arXiv preprint arXiv:2305.01035* (2023).

[36] 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.

[37] KHOO, Y., LU, J., AND YING, L. Solving parametric PDE problems with artificial neural networks. *European Journal of Applied Mathematics 32*, 3 (2021), 421–435.

[38] 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.

[39] MISHRA, S. A machine learning framework for data driven acceleration of computations of differential equations. *Mathematics in Engineering 1*, 1 (2019), 118–146.

[40] NABIAN, M. A., AND MEIDANI, H. A deep learning solution approach for high-dimensional random differential equations. *Probabilistic Engineering Mechanics 57* (2019), 14–25.

[41] 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).

[42] 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).

[43] NEUFELD, A., AND WU, S. Multilevel Picard algorithm for general semilinear parabolic PDEs with gradient-dependent nonlinearities. *arXiv:2310.12545* (2023).

[44] NGUWI, J. Y., PENENT, G., AND PRIVAULT, N. A deep branching solver for fully nonlinear partial differential equations. *arXiv preprint arXiv:2203.03234* (2022).

[45] 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).

[46] 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.

[47] RAISSI, M. *Forward–Backward Stochastic Neural Networks: Deep Learning of High-Dimensional Partial Differential Equations*. Peter Carr Gedenkschrift. Research Advances in Mathematical Finance. 2023, ch. 18, pp. 637–655.

[48] 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.

[49] 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.

[50] SIRIGNANO, J., AND SPILIOPOULOS, K. DGM: A deep learning algorithm for solving partial differential equations. *Journal of computational physics 375* (2018), 1339–1364.

[51] 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] Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
Email address: tuananh.nguyen@ntu.edu.sg