Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

Deep Learning in High Dimension: Neural Network Expression Rates for Analytic Functions in $L^2(\R^d,\gamma_d)$

Ch. Schwab and J. Zech

Research Report No. 2021-40
November 2021

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

Deep Learning in High Dimension: Neural Network Expression Rates for Analytic Functions in $L^2(\mathbb{R}^d, \gamma_d)$

Christoph Schwab* and Jakob Zech†

Abstract. For artificial deep neural networks, we prove expression rates for analytic functions $f : \mathbb{R}^d \to \mathbb{R}$ in the norm of $L^2(\mathbb{R}^d, \gamma_d)$ where $d \in \mathbb{N} \cup \{\infty\}$. Here γ_d denotes the Gaussian product probability measure on $\mathbb{R}^d$. We consider in particular ReLU and ReLU^k activations for integer $k \geq 2$. For $d \in \mathbb{N}$, we show exponential convergence rates in $L^2(\mathbb{R}^d, \gamma_d)$. In case $d = \infty$, under suitable smoothness and sparsity assumptions on $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, with γ_∞ denoting an infinite (Gaussian) product measure on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$, we prove dimension-independent expression rate bounds in the norm of $L^2(\mathbb{R}^{\mathbb{N}}, \gamma_\infty)$. The rates only depend on quantified holomorphy of (an analytic continuation of) the map f to a product of strips in $\mathbb{C}^d$ (in $\mathbb{C}^{\mathbb{N}}$ for $d = \infty$, respectively). As an application, we prove expression rate bounds of deep ReLU-NNs for response surfaces of elliptic PDEs with log-Gaussian random field inputs.

1. Introduction.

This paper addresses the approximation of analytic functions $f : \mathbb{R}^d \to \mathbb{R}$ by deep neural networks (DNNs for short) in the space $L^2(\mathbb{R}^d, \gamma_d)$. Here γ_d denotes the d-fold product Gaussian measure, with $d \in \mathbb{N} \cup \{\infty\}$. To quantify DNN expression rates, we assume f to belong to a class of functions that allows holomorphic extensions to certain cartesian products of strips around the real line in the complex plane. This implies summability results on coefficients in Wiener-Hermite polynomial chaos expansions of f. We separately discuss the finite dimensional case $d \in \mathbb{N}$ and the (countably) infinite dimensional case $d = \infty$. Our expression rate analysis is based on expressing such functions through their finite- or infinite-parametric Wiener-Hermite polynomial chaos (gpc) expansion. Reapproximating the gpc expansion, we provide DNN architectures and corresponding DNN size bounds which show that such functions can be approximated at an exponential convergence rate in finite dimension $d \in \mathbb{N}$. For $d = \infty$, i.e. in the infinite dimensional case, our DNN expression rate bounds are free from the so-called curse of dimensionality: we prove that in this case our DNN expression rate bounds are only determined by the summability of the gpc expansion coefficient sequences. Thus, while we concentrate on analytic functions, the scope of our results extends to statistical learning of any object that can be represented as a Wiener-Hermite expansion with bound on summability of the coefficient sequences.

Relevance of the present investigation derives from the fact that functions belonging to the above described class arise in particular as response maps in uncertainty quantification (UQ) for partial differential equations (PDEs for short) with Gaussian random field inputs. Modelling unknown inputs of elliptic or parabolic PDEs by a log-Gaussian random field, the corresponding PDE response surface can under certain assumptions be shown to be of this type [5]. We discuss a standard example in Sec. 6 ahead. As such, our results have broad implications for a wide range of problems in forward and inverse UQ. Dating back to the seminal works [22, 4] the numerical approximation of Gaussian Random Fields (GRFs for short) and response maps with GRF inputs by truncated Hermite polynomial chaos expansions has received substantial attention during recent years, specifically due to the ubiquitous role of GRFs in spatial statistics, theoretical physics, data assimilation, and stochastic Partial Differential Equations (PDEs for short). We refer to the surveys [3, 13, 10], to the recent publications [20, 12] and to the references there for the discussion of GRFs, as well as to, e.g., [8, 2] and the references there for the approximation of PDE response surfaces with log-GRF inputs.

1.1. Previous results.

In recent years, there has been substantial activity in the analysis of expression rates of ReLU-DNNs for various classes of functions. We mention for instance the papers

*Seminar for Applied Mathematics, ETH Zürich, 8092 Zürich, Switzerland, schwab@math.ethz.ch

†Interdisziplinäres Zentrum für wissenschaftliches Rechnen, Universität Heidelberg, 69120 Heidelberg, Germany, jakob.zech@uni-heidelberg.de

[23, 24] which established optimal convergence rates for functions of finite regularity. Approximation in L^p-spaces was discussed in [18]. In [17], DNN expression rates were given for functions from Sobolev- and Besov-spaces, as well as for certain classes of analytic functions. Holomorphic functions of $d \in \mathbb{N}$ many variables on bounded domains were shown to admit exponential expression rates by deep ReLU-NNs in [16]. The case of infinite-parametric holomorphic functions on cartesian products of bounded intervals was discussed in [19]. The analysis there is conceptually closely related to the present work. In this reference, we proved deep ReLU-NN expression rate bounds for gpc representations of countably-parametric functions on $[-1,1]^{\mathbb{N}}$. The obtained approximation rates do not suffer from the curse of dimensionality, and were shown to be governed only by a suitable notion of sparsity, as quantified in terms of summability of gpc coefficients. Importantly, with the exception of [19], all results in these references addressed approximation rate bounds for functions defined on bounded subdomains D of Euclidean space $\mathbb{R}^n$ with moderate, fixed "physical" dimension $n \in \mathbb{N}$. Also in other contexts, ReLU-NN expression rate bounds are often stated and proved for DNNs with bounded input ranges. On bounded intervals, ReLU-NNs afford in particular the efficient emulation of orthogonal Jacobi polynomials.

Our previous paper [19] is conceptually closely related to the present work. In this reference, we proved deep ReLU-NN expression rate bounds for generalized polynomial chaos ("gpc" for short) representations of countably-parametric functions which approximation rates do not suffer from the so-called curse of dimensionality. The DNN expression rates of such functions were shown in [19] to be governed only by a suitable notion of sparsity, as quantified in terms of summability of gpc coefficients. In [19], we only considered bounded parameter domains, and gpc expansions with respect to polynomials that are orthonormal with respect to probability measures on these domains. In particular, Legendre and Jacobi polynomials. Although the present results are in a similar spirit as the results in [19], they do not follow from these results, but differ both in statement and proofs in an essential way from the results in [19]. Similar to [19], the presently obtained expression rate bounds will be based on known (in part rather recent) bounds on approximation rates of n-term Hermite gpc expansions of GRFs, from [5].

Deep Neural Networks (DNNs) have seen intense research activity, mainly driven by successes in practical deep learning approaches in the emerging field of data science. This momentum has also initiated new developments in the numerical solution of PDEs, being based on DNNs as approximation architectures rather than "traditional" approaches built on Finite Element or Spectral methods. In practical applications, at times spectacular performance (in terms of accuracy versus DNN size) has been reported. These practical findings have been recently supported by theory indicating that ReLU DNNs can, indeed, emulate a wide range of linear approximation methods in classical function systems such as splines, multiresolution systems, polynomials, Fourier series, etc. Here, ReLU-NNs with suitable architectures afford with corresponding expression rate bounds which are equal, or only slighly inferior to rates afforded by the mentioned systems (see, e.g., [17, 16] and the references there). Importantly, all results in these references addressed approximation rate bounds for functions defined on bounded subdomains D of euclidean space $\mathbb{R}^n$ with moderate, fixed "physical" dimension $n \in \mathbb{N}$. Also in other contexts, ReLU-NN expression rate bounds are often stated and proved for DNNs with bounded input ranges. On bounded intervals, ReLU-NNs afford in particular the efficient emulation of orthogonal Jacobi polynomials.

The expression rate analysis of polynomial function systems on unbounded domains has received less attention. In view of the wide use of Gaussian process (GP for short) models and of Gaussian random fields in statistical modelling of uncertainty, and in theoretical physics [10], and due to the close connection of Hermite orthogonal polynomials with the Gaussian measure (e.g. [21, 22, 12] and the references there), expression rates of DNNs for Hermite polynomials in mean square with respect to Gaussian measure over $\mathbb{R}$ are crucial for restablishing various approximation rate bounds for Gaussian random fields, and in particular for operator equations with Gaussian random field inputs. The present paper addresses this question. The focus is on DNNs with so-called ReLU activation function. Despite these specificities of ReLU-NNs, our DNN architectures and expression rate bounds, which are explicit in the polynomial degree and in the accuracy, are valid also for wider families of activation functions. We expect that similar arguments allow to prove expression rate bounds also for other (smoother) activation functions.

1.2. Contributions. The present paper has the following principal contributions.

(i) We prove expression rate bounds for deep ReLU-NNs of univariate "probabilistic" Hermite polynomials H_n of polynomial degree $n \in \mathbb{N}_0$ on $\mathbb{R}$, in $L^2(\mathbb{R}, \gamma_1)$, i.e. in mean square with respect to Gaussian measure γ_1 on $\mathbb{R}$. This result is then generalized to multivariate Hermite polynomials, see Theorems 3.5 and 3.7.

(ii) In the case of finite parameter dimension d, we establish exponential convergence in $L^2(\mathbb{R}^d, \gamma_d)$ for the approximation of certain analytic functions by deep ReLU-NNs. See Theorem 4.7.

(iii) In infinite dimension, for a class of infinite parametric functions satisfying an analyticity condition, we prove ReLU-NN expression rate bounds $L^2(\mathbb{R}^{\mathbb{N}}, \gamma_\infty)$ that are free from the curse of dimension with explicit account of the NN size and depth. See Theorem 5.6.

(iv) As an example, we show how our result in infinite dimensions implies ReLU-NN expression rate bounds for response surfaces of elliptic PDEs with infinite-parametric, log-Gaussian random field input. See Proposition 6.2.

1.3. Notation. Throughout $C > 0$ is used to denote a generic constant that may change its value even within the same equation. Moreover, $z = x + \mathrm{i}y \in \mathbb{C}$ indicates that $z \in \mathbb{C}$ and $x = \Re[z] \in \mathbb{R}$ and $y = \Im[z] \in \mathbb{R}$. In particular, i shall denote the imaginary unit.

1.3.1. Gaussian measures. For finite $d \in \mathbb{N}$, denote by γ_d the standard Gaussian measure on $\mathbb{R}^d$. Its density w.r.t. the Lebesgue measure on $\mathbb{R}^d$ is given by

$$\frac{1}{(2\pi)^{d/2}} \mathrm{e}^{-\frac{\|\boldsymbol{y}\|_2^2}{2}} \qquad \forall \boldsymbol{y} \in \mathbb{R}^d,$$

where $\|\cdot\|_2$ is the Euclidean norm. Additionally, $\gamma = \bigotimes_{j\in\mathbb{N}} \gamma_1$ denotes the infinite product (probability) measure on $\mathbb{R}^{\mathbb{N}}$. We refer to [3, Chapter 2] for details. We write $L^2(\mathbb{R}^d, \gamma_d)$ for the usual L^2 space w.r.t. the measure γ_d. For $d = \infty$ we additionally introduce the shorthand notation $U := \mathbb{R}^{\mathbb{N}}$, indicating a countable cartesian product of real lines, the corresponding L^2-space is then $L^2(U, \gamma)$. Similarly, for a Banach space V and $k > 1$, $L^k(U, \gamma; V)$ is the Bochner space of functions with values in V.

1.3.2. Multiindices and polynomials. Throughout, $\mathbb{N} = \{1, 2, ...\}$ and $\mathbb{N}_0 = \{0, 1, 2, ...\}$. Multi-indices in $\mathbb{N}_0^d$ or $\mathbb{N}_0^{\mathbb{N}}$ shall be denoted by $\boldsymbol{\nu}$, i.e. $\boldsymbol{\nu} = (\nu_j)_{j=1}^d$ or $\boldsymbol{\nu} = (\nu_j)_{j\in\mathbb{N}}$ respectively. The size (or total order) of the multi-index $\boldsymbol{\nu}$ is $|\boldsymbol{\nu}| := \sum_{j\geq 1} \nu_j$. For $d = \infty$, by $\mathcal{F} = \{\boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}} : |\boldsymbol{\nu}| < \infty\}$ we denote the countable subset of $\mathbb{N}_0^{\mathbb{N}}$ of multi-indices of "finite support": if $\boldsymbol{\nu} \in \mathcal{F}$, we let $\operatorname{supp} \boldsymbol{\nu} := \{j : \nu_j \neq 0\}$ and $|\boldsymbol{\nu}|_0 := \#(\operatorname{supp} \boldsymbol{\nu})$. Comparison of multi-indices is component-wise: we write $\boldsymbol{\mu} \leq \boldsymbol{\nu}$ iff for every j holds $\mu_j \leq \nu_j$. A finite set $\Lambda \subseteq \mathbb{N}_0^d$ or $\Lambda \subseteq \mathcal{F}$ will be called *downward closed*, iff $\boldsymbol{\nu} \in \Lambda$ implies $\boldsymbol{\mu} \in \Lambda$ whenever $\boldsymbol{\mu} \in \Lambda$.

With $\mathbb{P}_n := \operatorname{span}\{x^j : j \in \{0, \ldots, n\}\}$ we denote the space of all polynomials of degree at most n with real coefficients. In the multivariate case, for a subset $\Lambda \subseteq \mathbb{N}_0^d$ with $d \in \mathbb{N}$ or $\Lambda \subseteq \mathcal{F}$, we write $\mathbb{P}_\Lambda := \operatorname{span}\{\prod_{j\in\operatorname{supp}\boldsymbol{\nu}} x_j^{\nu_j} : \boldsymbol{\nu} \in \Lambda\}$.

1.3.3. Neural networks. We consider feedforward neural networks without skip connections. That is, for a given activation function $\sigma : \mathbb{R} \to \mathbb{R}$, we consider mappings $\Phi : \mathbb{R}^{n_0} \to \mathbb{R}^{n_{L+1}}$ which can be represented via

$$\Phi = A_L \circ \sigma \circ \cdots \circ \sigma \circ A_0 \tag{1.1}$$

for certain linear transformations $A_j : \mathbb{R}^{n_j} \to \mathbb{R}^{n_{j+1}} : x \mapsto W_j x + b_j$. Here $W_j \in \mathbb{R}^{n_{j+1}\times n_j}$ are the weight matrices and $b_j \in \mathbb{R}^{n_{j+1}}$ are the bias vectors, and the application of σ in (1.1) is understood componentwise. Such a function Φ will be called a σ-NN of depth L and size

$$\mathrm{size}(\Phi) := |\{(j,k,l) : (W_j)_{k,l} \neq 0\} \cup \{(j,k) : (b_j)_k \neq 0\}|.$$

We also use the notation $\mathrm{depth}(\Phi) := L$. Hence the depth corresponds to the number of applications of the activation function, and the size corresponds to the number of nonzero weights and biases in the network.

1.4. Layout. The structure of the paper is as follows. In Section 2.1, we recapitulate general definitions and classical properties of Hermite polynomials. Section 2.2 addresses specific properties of Hermite polynomials which are required in the proofs of the ensuing DNN emulation bounds. Section 3 then contains the core results of the present paper: we provide explicit constructions of ReLU and of ReLU^k DNNs which emulate Hermite polynomials in one dimension. We generalize, via the approximate product operator, also to multiple dimensions. Section 4 then has a first application: exponential DNN emulation rate bounds of nonlinear, holomorphic maps on $\mathbb{R}^d$, in finite dimension d. Section 5 addresses the infinite-dimensional case. Section 6 presents an application, dimension-independent expression rate bounds for solutions of linear, elliptic PDEs with a random coefficient, which is a log-Gaussian random field. The final Section 7 reviews the main results, and indicates extensions and further applications of the presently developed theory.

Acknowledgement: Work performed in part in the programme "Mathematics of Deep Learning" (MDL) at the Isaac Newton Institute, Cambridge, UK from July-December 2021. Fertile exchanges, and stimulating workshops are warmly acknowledged.

2. Hermite polynomials and functions.

2.1. Basic definitions and properties. For $n \in \mathbb{N}_0$ we denote by H_n the nth **probabilists' Hermite polynomial**[1] normalized in $L^2(\mathbb{R}, \gamma_1)$, i.e.

$$H_n(x) := \frac{(-1)^n}{\sqrt{n!}} \mathrm{e}^{\frac{x^2}{2}} \frac{d^n}{dx^n} \mathrm{e}^{-\frac{x^2}{2}} \qquad \forall\, n \in \mathbb{N}_0 \tag{2.1}$$

with the usual convention $0! = 1$. Since for any $f \in C^1(\mathbb{R})$ holds

$$\frac{d}{dx}\left(f(x)\,\mathrm{e}^{-\frac{x^2}{2}}\right) = f'(x)\,\mathrm{e}^{-\frac{x^2}{2}} - x f(x)\,\mathrm{e}^{-\frac{x^2}{2}}, \tag{2.2}$$

it is easy to see that $H_n \in \mathbb{P}_n$.

Next, we introduce the **Hermite functions** via

$$h_n(x) := \frac{(-1)^n}{\sqrt{\pi^{1/2} 2^n n!}} \mathrm{e}^{\frac{x^2}{2}} \frac{d^n}{dx^n} \mathrm{e}^{-x^2} \qquad \forall\, n \in \mathbb{N}_0. \tag{2.3}$$

The relation between the Hermite polynomials and the Hermite functions is made clear by the following lemma.

Lemma 2.1. *The map*

$$\Theta : L^2(\mathbb{R}, \gamma_1) \to L^2(\mathbb{R}) : f(x) \mapsto f(2^{1/2}x) \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\pi^{\frac{1}{4}}}$$

is an isometric isomorphism and $\Theta(H_n) = h_n$ for all $n \in \mathbb{N}_0$.

[1] The *physicists'* Hermite polynomials are defined as $x \mapsto H_n(2^{1/2}x)$. Since we shall not use them in this manuscript, we simply refer to the $(H_n)_{n\in\mathbb{N}_0}$ in the following as the Hermite polynomials.

Proof. Let $f \in L^2(\mathbb{R},\gamma_1)$. Using the change of variables $x = 2^{1/2}y$

$$\|f\|^2_{L^2(\mathbb{R},\gamma_1)} = \int_{\mathbb{R}} f(x)^2 \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \mathrm{d}x = \int_{\mathbb{R}} f(2^{1/2}y)^2 \frac{\mathrm{e}^{-y^2}}{\pi^{\frac{1}{2}}} \mathrm{d}y = \|\Theta(f)\|^2_{L^2(\mathbb{R})}. \tag{2.4}$$

Thus Θ is an isometry. By a similar argument $\tilde{\Theta}(F)(y) := F(\frac{y}{2^{1/2}})\exp(\frac{y^2}{4})\pi^{1/4}$ defines an isometry from $L^2(\mathbb{R}) \to L^2(\mathbb{R},\gamma_1)$ and $\Theta \circ \tilde{\Theta}$ is the identity. In all, Θ is an isometric isomorphism.

To show $\Theta(H_n) = h_n$, denote $r(x) := \mathrm{e}^{-x^2}$ and $r(\mathrm{i}x) = \mathrm{e}^{x^2}$. Due to $\frac{d^n}{dx^n} r(\frac{x}{2^{1/2}}) = 2^{-n/2} r^{(n)}(\frac{x}{2^{1/2}})$ we have by (2.1)

$$H_n(x) = \frac{(-1)^n}{\sqrt{2^n n!}} r\left(\mathrm{i}\frac{x}{2^{1/2}}\right) r^{(n)}\left(\frac{x}{2^{1/2}}\right)$$

and by (2.3)

$$h_n(x) = \frac{(-1)^n}{\sqrt{\pi^{1/2}2^n n!}} \mathrm{e}^{-\frac{x^2}{2}} r(\mathrm{i}x) r^{(n)}(x).$$

Thus

$$h_n(x) = H_n(2^{1/2}x)\frac{\mathrm{e}^{-\frac{x^2}{2}}}{\pi^{\frac{1}{4}}} = \Theta(H_n).$$

■

As is well-known, these sequences are orthonormal bases in the respective spaces. We recall the classical proof for the convenience of the reader.

Proposition 2.2. *It holds*

(i) $(H_n)_{n\in\mathbb{N}_0}$ *is an ONB of* $L^2(\mathbb{R},\gamma_1)$*,*

(ii) $(h_n)_{n\in\mathbb{N}_0}$ *is an ONB of* $L^2(\mathbb{R})$*.*

Proof. We start by showing orthonormality of $(H_n)_{n\in\mathbb{N}_0}$ in $L^2(\mathbb{R},\gamma_1)$. Let $n \in \mathbb{N}_0$, $m \in \mathbb{N}$ and $n \le m$. Integrating by parts we have

$$\int_{\mathbb{R}} H_n(x)H_m(x)\mathrm{d}\gamma_1(x) = \frac{(-1)^{n+m}}{\sqrt{2\pi n!m!}} \int_{\mathbb{R}} H_n(x)\frac{d^m}{dx^m}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x = \frac{(-1)^n}{\sqrt{2\pi n!m!}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{x^2}{2}} \frac{d^m}{dx^m} H_n(x)\mathrm{d}x.$$

Since $H_n \in \mathbb{P}_n$, for $n < m$ the integrand vanishes. In case $n = m$, $\frac{d^n}{dx^n}H_n(x)$ equals $n!$ times the leading coefficient of H_n. Using (2.1)-(2.2) one obtains $\frac{d^n}{dx^n}H_n(x) = (-1)^n n!$. Since $\int_{\mathbb{R}} H_0(x)^2\mathrm{d}\gamma_1(x) = \int_{\mathbb{R}} 1\mathrm{d}\gamma_1(x) = 1$, we have shown

$$\int_{\mathbb{R}} H_n(x)H_m(x)\mathrm{d}\gamma_1(x) = \delta_{n,m} \qquad \forall\, n,m \in \mathbb{N}_0.$$

We show completeness of $(H_n)_{n\in\mathbb{N}_0}$ in $L^2(\mathbb{R},\gamma_1)$. Since $H_n \in \mathbb{P}_n$ (with nonzero leading coefficient), it suffices to show density of all polynomials in $L^2(\mathbb{R},\gamma_1)$. Let $f \in L^2(\mathbb{R},\gamma_1)$ be such that $\int_{\mathbb{R}} f(x)x^n\mathrm{d}\gamma_1(x) = 0$ for all $n \in \mathbb{N}_0$. Define $g(z) := \int_{\mathbb{R}} f(x)\exp(zx)\mathrm{d}\gamma_1(x)$, which yields an entire function on $\mathbb{C}$. It holds $g^{(n)}(0) = \int_{\mathbb{R}} x^n f(x)\mathrm{d}\gamma_1(x) = 0$ for all $n \in \mathbb{N}$. Thus $g \equiv 0$. However, $\mathbb{R} \ni x \mapsto g(-\mathrm{i}x)$ is the Fourier transform of f. This implies $f \equiv 0$ and consequently the Hermite polynomials $(H_n)_{n\in\mathbb{N}_0}$ are dense in $L^2(\mathbb{R},\gamma_1)$.

Finally, since $\Theta : L^2(\mathbb{R},\gamma_1) \to L^2(\mathbb{R})$ in Lemma 2.1 is an isometric isomorphism, it transforms the ONB $(H_n)_{n\in\mathbb{N}_0}$ of $L^2(\mathbb{R},\gamma_1)$ to an ONB $(\Theta(H_n))_{n\in\mathbb{N}_0} = (h_n)_{n\in\mathbb{N}_0}$ of $L^2(\mathbb{R})$. ■

2.2. Some preliminary bounds.

We will use *Cramer's bound* [9] on the Hermite functions,

$$\sup_{x\in\mathbb{R}} |h_n(x)| \le \pi^{-1/4} \qquad \forall n \in \mathbb{N}_0. \tag{2.5}$$

The Hermite polynomials allow the explicit representation, see, e.g., [21, Eqn. (5.5.4)][2]

$$H_n(x) = \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{\sqrt{n!}(-1)^j}{j!(n-2j)!2^j} x^{n-2j}. \tag{2.6}$$

In the following we also write $H_n(x) = \sum_{j=0}^{n} c_{n,j} x^n$.

Lemma 2.3. *For all $n \in \mathbb{N}_0$*

$$\sum_{j=0}^{n} |c_{n,j}| \le 6^{n/2} \le 3^n. \tag{2.7}$$

Proof. One checks (e.g. with Stirling's inequality) that $\sqrt{n!} \le 2^n(\lfloor n/2 \rfloor)!$ for all $n \in \mathbb{N}_0$. By (2.6) the term $\sum_{j=0}^{n} |c_{n,j}|$ is bounded by

$$\sum_{j=0}^{\lfloor n/2 \rfloor} \frac{\sqrt{n!}}{j!(n-2j)!2^j} \le 2^n \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{\lfloor \frac{n}{2} \rfloor !}{j!(\lfloor \frac{n}{2} \rfloor - j)!2^j} = 2^n \sum_{j=0}^{\lfloor n/2 \rfloor} \binom{\lfloor \frac{n}{2} \rfloor}{j} 2^{-j} = 2^n \left(1 + \frac{1}{2}\right)^{\lfloor \frac{n}{2} \rfloor}.$$

The last term is bounded by $2^n(3/2)^{n/2} \le 3^n$ which concludes the proof. ∎

Lemma 2.3 implies the (crude) bound

$$|H_n(x)| \le (3 \max\{1, |x|\})^n \qquad \forall x \in \mathbb{R}. \tag{2.8}$$

In the following for $n \in \mathbb{N}_0 \cup \{-1\}$, $n!!$ denotes the double factorial, i.e. $-1!! = 0!! = 1!! = 1$ and $n!! = n \cdot (n-2)!!$ if $n \ge 2$.

Lemma 2.4. *Let $M \ge 2$ and $n \in \mathbb{N}_0$. Then*

$$\int_{|x|>M} \mathrm{e}^{-\frac{x^2}{2}} x^n \mathrm{d}x \le n!! M^n \mathrm{e}^{-\frac{M^2}{2}}. \tag{2.9}$$

Proof. Set $a_n := \int_{x>M} x^n \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x$. For $n = 0$

$$a_0 = \int_{x>M} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x = \int_{y>0} \mathrm{e}^{-\frac{(y+M)^2}{2}} \mathrm{d}y = \mathrm{e}^{-\frac{M^2}{2}} \int_{y>0} \mathrm{e}^{-\frac{y^2}{2} - My} \mathrm{d}y \le \frac{1}{2} \mathrm{e}^{-\frac{M^2}{2}},$$

where we used

$$\int_{y>0} \mathrm{e}^{-\frac{y^2}{2} - My} \mathrm{d}y \le \int_{y>0} \mathrm{e}^{-\frac{y^2}{2} - 2y} \mathrm{d}y \le \left(\int_{y>0} \mathrm{e}^{-y^2} \mathrm{d}y\right)^{1/2} \left(\int_{y>0} \mathrm{e}^{-4y} \mathrm{d}y\right)^{1/2} = \frac{\pi^{1/4}}{2^{1/2}} \frac{1}{\sqrt{4}} < \frac{1}{2},$$

which follows by the well-known fact $\int_{y>0} \mathrm{e}^{-y^2} \mathrm{d}y = \sqrt{\pi}/2$. For $n = 1$

$$a_1 = \int_{x>M} x \,\mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x = -\int_{x>M} (\mathrm{e}^{-\frac{x^2}{2}})' \mathrm{d}x = \mathrm{e}^{-\frac{M^2}{2}}.$$

This shows (2.9) for $n \in \{0, 1\}$. For any $n \ge 2$, using integration by parts

$$a_n = \int_{x>M} x^n \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x = -\int_{x>M} x^{n-1} (\mathrm{e}^{-\frac{x^2}{2}})' \mathrm{d}x = M^{n-1} \mathrm{e}^{-\frac{M^2}{2}} + (n-1) \int_{x>M} x^{n-2} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x,$$

[2] A $\sqrt{n!}$ factor is due to a different scaling, compare [21, Eqn. (5.5.3)] with (2.1).

so that $a_n = M^{n-1}\exp(-M^2/2) + (n-1)a_{n-2}$. For $n \in \{0,1\}$ we have in particular shown $a_n \le (n-1)!!M^n\exp(-M^2/2)$. Using $M^{n-1} + M^{n-2} \le M^n$ since $M \ge 2$, by induction we get

$$\forall n \ge 2 : \quad a_n \le \mathrm{e}^{-\frac{M^2}{2}} M^{n-1} + (n-1)!!M^{n-2}\mathrm{e}^{-\frac{M^2}{2}} \le (n-1)!!M^n\mathrm{e}^{-\frac{M^2}{2}} \ .$$

Using this bound and again the recurrence we obtain for $n \ge 2$

$$a_n \le \mathrm{e}^{-\frac{M^2}{2}}(M^{n-1} + (n-1)!!M^{n-2}) \le \mathrm{e}^{-\frac{M^2}{2}} M^{n-2}(M + (n-1)!!). \tag{2.10}$$

For all $x \ge 1$ holds $M + x \le \frac{3}{4}M^2 x$ because $M \ge 2$. Furthermore $(n-1)!!(3/2) \le n!!$ for all $n \ge 2$. Hence, with $x = (n-1)!! \ge 1$ we get $M + (n-1)!! \le 3M^2(n-1)!!/4 \le \frac{1}{2}M^2 n!!$. Together with (2.10) this finally implies $a_n \le \frac{1}{2}n!!M^n\exp(-M^2/2)$ and concludes the proof. ■

We note in passing that Lemma 2.4 and (2.8) imply for every $M \ge 2$, $p \ge 1$ and $n \in \mathbb{N}_0$

$$\int_{|x|>M} |H_n(x)|^p \mathrm{d}\gamma_1(x) = \frac{1}{\sqrt{2\pi}} \int_{|x|>M} |H_n(x)|^p \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x \le \frac{1}{\sqrt{2\pi}}(pn)!!(3M)^{pn}\mathrm{e}^{-\frac{M^2}{2}} \ . \tag{2.11}$$

3. DNN emulation of Hermite polynomials.

A key technical step in the DNN expression rate analysis of Gaussian random fields is the ReLU NN expression of Hermite polynomials. Due to general representation of GRFs in terms of Hermite-expansions (e.g. [4, 10, 3] and the references there) quantitative bounds for ReLU NN expression rates of GRFs will follow from assumptions on summability of Hermite coefficient sequences of the GRFs and from ReLU DNN expression rates of Hermite polynomials H_n. To establish the latter is the purpose of the present section. Due to the goal of expressing truncated Hermite gpc expansions, our main result in the present section, Theorem 3.7, will provide quantitative bounds of expression of *(collections of tensor products of) Hermite polynomials* by *by one common ReLU NN architecture*.

3.1. Univariate Hermite polynomials.

We start by recalling that univariate, continuous piecewise linear functions can be realized exactly by shallow ReLU-NNs, see, e.g., [19, Lemma 4.5].

Lemma 3.1. *Let $-\infty < x_0 < x_1 < \cdots < x_{n-1} < x_n < \infty$ induce a partition of $\mathbb{R}$ into $n+2 \in \mathbb{N}$ intervals. For any continuous piecewise linear function $f : \mathbb{R} \to \mathbb{R}$ w.r.t. this partition, there exist a ReLU NN $\phi : \mathbb{R} \to \mathbb{R}$ such that $\phi(x) = f(x)$ for $x \in \mathbb{R}$ and $\mathrm{size}(\phi) \le 2(n+2)+1$, $\mathrm{depth}(\phi) = 1$.*

Next, we address truncation of ReLU-NNs to finite support in $\mathbb{R}$.

Lemma 3.2. *Let $M > 0$ and let $\phi : \mathbb{R} \to \mathbb{R}$ be a ReLU NN. For every $\delta \in (0, M)$ there exists a ReLU NN $\psi : \mathbb{R} \to \mathbb{R}$ satisfying $\sup_{x\in[-M,M]} |\psi(x)| \le \sup_{x\in[-M,M]} |\phi(x)|$,*

$$\psi(x) = \begin{cases} \phi(x) & x \in [-M+\delta, M-\delta] \\ 0 & x \in \mathbb{R}\backslash[-M, M] \end{cases} \tag{3.1}$$

$\mathrm{size}(\psi) \le C(1 + \mathrm{size}(\phi))$ and $\mathrm{depth}(\psi) \le C(1 + \mathrm{depth}(\phi))$, with $C > 0$ independent of M, δ, ϕ.

Proof. Since ϕ is a ReLU NN, there exists $N > 0$ such that $\phi|_{(-\infty,-N]}$ and $\phi|_{[N,\infty)}$ are linear. We now construct a ReLU NN η such that $\eta|_{[-M,M]} = \phi|_{[-M,M]}$ and $\eta|_{(-\infty,-M]}$ and $\eta|_{[M,\infty)}$ are linear.

Set

$$p(x) = \frac{\phi(M) + \phi(-M)}{2} + x\frac{\phi(M) - \phi(-M)}{2M},$$

i.e. $p : \mathbb{R} \to \mathbb{R}$ is linear and $p(-M) = \phi(-M)$, $p(M) = \phi(M)$. Then $(\phi-p)(M) = (\phi-p)(-M) = 0$. For $x \in \mathbb{R}$

$$q(x) := M - \sigma(x) - \sigma(-x) = \begin{cases} x + M & x < 0 \\ -x + M & x \ge 0 \end{cases}$$

is a ReLU NN satisfying $q(-M) = q(M) = 0$, $q|_{(-M,M)} > 0$, $q|_{(-\infty,-M)} < 0$ and $q|_{(M,\infty)} < 0$. Since $(\phi - p)|_{(-\infty,N]}$ and $(\phi - p)|_{[N,\infty)}$ are linear, we can find $\alpha > 0$ such that $(\phi - p) + \alpha q$ is positive on $(-M, M)$ and negative on $\mathbb{R}\backslash[-M, M]$. Then

$$\eta(x) = \sigma(\phi(x) - p(x) + \alpha q(x)) + p(x) - \alpha q(x)$$

equals $\phi(x)$ for $x \in [-M, M]$ and $\eta|_{(-\infty,-M]}$ and $\eta|_{[M,\infty)}$ are linear. Since we only added and subtracted continuous, piecewise linear functions from ϕ and composed them with σ, the function η can be expressed by a ReLU NN.

Now we construct ψ. Wlog let $\delta \in (0, M)$ be so small that $\eta|_{[-M,-M+\delta]}$ and $\eta|_{[M-\delta,M]}$ are linear (which is possible because η is a continuous, piecewise linear function). Then both, $\eta|_{[M-\delta,M]}$ and $\eta|_{[M,\infty)}$ are linear, and by Lemma 3.1 the function $r : \mathbb{R} \to \mathbb{R}$ that is continuous, piecewise linear on the partition $x_0 = -\infty$, $x_1 = M - \delta$, $x_2 = M$, $x_3 = \infty$ and satisfies $r|_{(-\infty,M-\delta]} = 0$, $r(M) = \eta(M)$ and $r|_{[M,\infty)} = \eta|_{[M,\infty)}$ is expressed by a network of size $O(1)$. Then $\eta - r|_{(-\infty,M-\delta]} = \eta|_{(-\infty,M-\delta]}$ and $\eta - r|_{[M,\infty)} \equiv 0$. Furthermore $(\eta - r)(M - \delta) = \eta(M - \delta)$, $(\eta - r)(M) = 0$ and $(\eta - r)|_{[M-\delta,M]}$ is linear so that $\sup_{x\in[M-\delta,M]} |\eta(x) - r(x)| \leq |\eta(M - \delta)|$. Similarly, we can construct $s : \mathbb{R} \to \mathbb{R}$ continuous, piecewise affine such that $s|_{[-M+\delta,\infty)} \equiv 0$, $s(-M) = \eta(-M)$ and $s|_{(-\infty,-M]} = \eta|_{(-\infty,-M]}$. Then $\psi = \eta - r - s$ is as claimed. ■

We are now in position to state our main result on architecture and quantitative bounds for emulations of Hermite polynomials by deep ReLU-NNs.

Proposition 3.3. *Let $n \in \mathbb{N}_0$, $M > 0$ and $\varepsilon \in (0, \mathrm{e}^{-1})$ be arbitrary. Then there exists a ReLU NN $\tilde{H}_{n,M,\varepsilon} : \mathbb{R} \to \mathbb{R}$ such that*

(i) $\|H_n - \tilde{H}_{n,M,\varepsilon}\|_{L^2(\mathbb{R},\gamma_1)} \leq \varepsilon + \sqrt{2n!!}(3M)^n \mathrm{e}^{-\frac{M^2}{4}}$,

(ii) $\tilde{H}_{n,M,\varepsilon}(x) = 0$ *for* $|x| > M$ *and* $\sup_{x\in\mathbb{R}} |\tilde{H}_{n,M,\varepsilon}(x)| \leq 1 + (3M)^n$,

(iii) for a constant $C > 0$ independent of n, M, ε

$$\mathrm{size}(\tilde{H}_{n,M,\varepsilon}) \leq C\Big(1 + n^2 \log(M) + n \log\left(\frac{n}{\varepsilon}\right)\Big),$$
$$\mathrm{depth}(\tilde{H}_{n,M,\varepsilon}) \leq C((1 + \log(n))(n \log(M) - \log(\varepsilon))).$$

Proof. In this proof we will need the following result shown in [17, Prop. 4.2]: for any polynomial $p(x) = \sum_{j=0}^n c_j x^j$, there exists a neural network $\tilde{p}$ such that $|p(x) - \tilde{p}(x)| \leq \varepsilon$ for all $x \in [-1, 1]$, and with $C_0 := \max\{2, \sum_{j=0}^n |c_j|\}$ it holds

$$\mathrm{size}(\tilde{p}) \leq C\Big((1 + n) \log\left(\frac{C_0}{\varepsilon}\right) + n \log(n)\Big), \qquad \mathrm{depth}(\tilde{p}) \leq C\Big((1 + \log(n)) \log\left(\frac{C_0}{\varepsilon}\right) + \log(n)^3\Big),$$

where the constant C is independent of $\varepsilon \in (0, \mathrm{e}^{-1})$ and of $n \in \mathbb{N}_0$.

Denote by $H_{n,M}(x) = H_n(Mx)$ the rescaled Hermite polynomial. Then $H_{n,M}(x) = \sum_{j=0}^n c_{n,j} M^j x^j$. By Lemma 2.3 it holds $C_0 := \sum_{j=0}^n |c_{n,j} M^j| \leq M^n \sum_{j=0}^n |c_{n,j}| \leq (3M)^n$. Thus by [17, Prop. 4.2] there exists a neural network $\hat{H}_{n,M,\varepsilon}$ such that

$$\sup_{x\in[-1,1]} |H_{n,M}(x) - \hat{H}_{n,M,\varepsilon}(x)| \leq \frac{\varepsilon}{2} \tag{3.2}$$

and

$$\mathrm{size}(\hat{H}_{n,M,\varepsilon}) \leq C\Big(1 + n^2 \log(M) + n \log\left(\frac{n}{\varepsilon}\right)\Big),$$
$$\mathrm{depth}(\hat{H}_{n,M,\varepsilon}) \leq C\Big((1 + \log(n))(n \log(M) - \log(\varepsilon))\Big), \tag{3.3}$$

for some constant $C > 0$ independent of $M \geq 2$, $n \in \mathbb{N}_0$ and $\varepsilon \in (0, \mathrm{e}^{-1})$ (for the bound on the depth we could absorb the term $\log(n)^3$ in $n \log(M)$, since $\log(M) > 0$ due to $M \geq 2$).

With $\delta > 0$ for the moment fixed, but to be chosen shortly (in dependence of n and ε), by Lemma 3.2 there exists a NN $\tilde{H}_{n,M,\varepsilon}$ such that $\sup_{x\in[-M,M]} |\tilde{H}_{n,M,\varepsilon}(x)| \le \sup_{x\in[-1,1]} |\hat{H}_{n,M,\varepsilon}(x)|$ and

$$\tilde{H}_{n,M,\varepsilon}(x) := \begin{cases} \hat{H}_{n,M,\varepsilon}(\frac{x}{M}) & x \in [-M+\delta, M-\delta] \\ 0 & x \in \mathbb{R}\backslash[-M,M]. \end{cases}$$

For $x \in [-M+\delta, M-\delta]$

$$|H_n(x) - \tilde{H}_{n,M,\varepsilon}(x)| = \left| H_{n,M}\Big(\frac{x}{M}\Big) - \hat{H}_{n,M,\varepsilon}\Big(\frac{x}{M}\Big) \right| \le \varepsilon.$$

By Lemma 3.2, the depth and size bounds for $\hat{H}_{n,M,\varepsilon}$ from (3.3) are also valid for $\tilde{H}_{n,M,\varepsilon}$ (possibly for a different constant C), which shows (iii).

Next, for $x \in \mathbb{R}\backslash[-M,M]$ it holds $|H_n(x) - \tilde{H}_{n,M,\varepsilon}(x)| = |H_n(x)|$. By (2.8), (3.2) and Lemma 3.2 we find

$$\sup_{x\in\mathbb{R}} |\tilde{H}_{n,M,\varepsilon}(x)| \le \sup_{x\in[-1,1]} |\hat{H}_{n,M,\varepsilon}(x)| \le \varepsilon + \sup_{x\in[-M,M]} |H_n(x)| \le 1 + (3M)^n, \tag{3.4}$$

and thus for $x \in [-M,M]\backslash[-M+\delta, M-\delta]$ we get $|H_n(x) - \tilde{H}_{n,M,\varepsilon}(x)| \le 1 + 2(3M)^n$. Hence using (2.11)

$$\begin{aligned} \|H_n - \tilde{H}_{n,M,\varepsilon}\|_{L^2(\mathbb{R},\gamma_1)} &\le \|H_n - \tilde{H}_{n,M,\varepsilon}\|_{L^2([-M,M],\gamma)} + \|H_n - \tilde{H}_{n,M,\varepsilon}\|_{L^2([-M,M]\backslash[-M+\delta,M-\delta],\gamma)} \\ &\quad + \|H_n\|_{L^2(\mathbb{R}\backslash[-M,M],\gamma)} \\ &\le \frac{\varepsilon}{2} + \sqrt{\delta}(1 + 2(3M)^n) + \sqrt{(2n)!!}(3M)^n \,\mathrm{e}^{-\frac{M^2}{4}}. \end{aligned}$$

Choosing $\delta > 0$ small enough it holds $\sqrt{\delta}(1 + 2(3M)^n) \le \frac{\varepsilon}{2}$, which shows (i).

Finally, (ii) holds by (3.4) and the construction of $\tilde{H}_{n,M,\varepsilon}$. ■

Lemma 3.4. *For all $n \in \mathbb{N}$ it holds $\sup_{x\in\mathbb{R}} x^n \mathrm{e}^{-\frac{x^2}{2}} = n^{\frac{n}{2}} \mathrm{e}^{-\frac{n}{2}} = \mathrm{e}^{\frac{n(\log(n)-1)}{2}}$.*

Proof. We have $(x^n \mathrm{e}^{-x^2/2})' = (nx^{n-1} - x^{n+1})\mathrm{e}^{-x^2/2} = x^{n-1}(n - x^2)\mathrm{e}^{-x^2/2}$. The only positive root of this term is $x = \sqrt{n}$, which implies the lemma. ■

Corollary 3.5. *Consider the setting of Proposition 3.3 and set, for $\varepsilon \in (0, \mathrm{e}^{-1})$,*

$$M(n,\varepsilon) := \sqrt{24(n\log(2n) - \log(\varepsilon))}, \quad n \in \mathbb{N}. \tag{3.5}$$

With this choice of M, define the ReLU*-NN $\tilde{H}_{n,\varepsilon} := \tilde{H}_{n,M,\varepsilon} : \mathbb{R} \to \mathbb{R}$. It satisfies*

(i) $\|H_n - \tilde{H}_{n,\varepsilon}\|_{L^2(\mathbb{R},\gamma_1)} \le 2\varepsilon$,
(ii) $\tilde{H}_{n,\varepsilon}(x) = 0$ *for* $|x| > M$ *and* $\sup_{x\in\mathbb{R}} |\tilde{H}_{n,\varepsilon}(x)| \le 1 + (3M)^n$,
(iii) for some $C > 0$ independent of n and ε

$$\begin{aligned} \mathrm{size}(\tilde{H}_{n,\varepsilon}) &\le C\Big(1 + n^2(\log(n) + \log(-\log(\varepsilon))) + n\log\Big(\frac{n}{\varepsilon}\Big)\Big), \\ \mathrm{depth}(\tilde{H}_{n,\varepsilon}) &\le C(1 + n\log(n)^2 + n\log(n)\log(-\log(\varepsilon)) - \log(n)\log(\varepsilon)). \end{aligned}$$

Proof. Inserting M from (3.5) into the bound in Proposition 3.3 (iii) we get

$$\mathrm{size}(\tilde{H}_{n,\varepsilon}) \le C\left(1 + \frac{1}{2}n^2 \log\left(24(n\log(2n) - \log(\varepsilon))\right) + n\log\Big(\frac{n}{\varepsilon}\Big)\right).$$

For $a, b \geq 1$ it holds

$$\log(a+b) = \log(a) + \int_a^{a+b} \frac{1}{x} \mathrm{d}x \leq \log(a) + \int_1^{1+b} \frac{1}{x} \mathrm{d}x = \log(a) + \log(1+b) \leq 1 + \log(a) + \log(b).$$

With $a = 24n\log(2n)$ and $b = -24\log(\varepsilon)$ we get

$$\begin{aligned} \operatorname{size}(\tilde{H}_{n,\varepsilon}) &\leq C\left(1 + \frac{1}{2}n^2\big(1 + \log(24n) + \log(\log(2n)) + \log(-24\log(\varepsilon))\big) + n\log\left(\frac{n}{\varepsilon}\right)\right) \\ &\leq C\Big(1 + n^2(\log(n) + \log(-\log(\varepsilon))) + n\log\left(\frac{n}{\varepsilon}\right)\Big) \end{aligned}$$

for a constant C independent of n and ε. This shows the bound on the size in (iii). The bound on the depth is obtained similarly.

To show (i) we use Proposition 3.3 (i) and claim that $\sqrt{(2n)!!}(3M)^n \,\mathrm{e}^{-\frac{M^2}{4}} \leq \varepsilon$. Since $\sqrt{(2n)!!} \leq (2n)^n$ it is sufficient to show that

$$-\frac{M^2}{4} + n\log(2n) + n\log(3M) - \log(\varepsilon) \leq 0. \tag{3.6}$$

The definition of M implies $\frac{M^2}{24} \geq n\log(2n) - \log(\varepsilon)$ and thus

$$-\frac{M^2}{24} + n\log(2n) - \log(\varepsilon) \leq 0. \tag{3.7}$$

Next we show

$$-\frac{M^2}{6} + n\log(3M) \leq 0,$$

which will then imply (3.6) due to $\frac{1}{24} + \frac{1}{6} \leq \frac{1}{4}$. The last inequality is equivalent to $\frac{M^2}{\log(3M)} \geq 6n$. The function $x \mapsto \frac{x^2}{\log(3x)}$ is monotonically increasing for $x \geq 1$ and by (3.5) it holds $M \geq \sqrt{24n\log(2n)} =: x$. Hence

$$\frac{M^2}{\log(3M)} \geq \frac{x^2}{\log(3x)} = 6n\frac{4\log(2n)}{\log(3\sqrt{24n\log(2n)})}.$$

It suffices to show that $\frac{4\log(2n)}{\log(3\sqrt{24n\log(2n)})} \geq 1$ for all $n \in \mathbb{N}$. It is checked directly that this holds for $n = 1$. Furthermore, this term is monotonically increasing for $n \geq 1$, so that it is true for all $n \in \mathbb{N}$. Together with (3.7) this verifies (3.6). In all, together with Proposition 3.3 (i) we get $\|H_n - \tilde{H}_{n,\varepsilon}\|_{L^2(\mathbb{R},\gamma_1)} \leq 2\varepsilon$. ■

3.2. Multivariate Hermite polynomials. We proceed to show ReLU-NN expression bounds for multivariate, tensorized Hermite polynomials.

Recall that $\mathcal{F} = \{\boldsymbol{\nu} \in \mathbb{N}_0^\infty : |\boldsymbol{\nu}| < \infty\}$ denotes the (countable) set of all finitely supported multiindices. For a finite index set $\Lambda \subseteq \mathcal{F}$, we define

$$\operatorname{supp}\Lambda := \{j \in \operatorname{supp}\boldsymbol{\nu} \,:\, \boldsymbol{\nu} \in \Lambda\} \tag{3.8}$$

and we introduce the *maximum order* $m(\Lambda)$ and the *effective dimension* $d(\Lambda)$ of Λ as

$$m(\Lambda) := \max_{\boldsymbol{\nu}\in\Lambda} |\boldsymbol{\nu}|_1, \qquad d(\Lambda) := \max_{\boldsymbol{\nu}\in\Lambda} |\boldsymbol{\nu}|_0. \tag{3.9}$$

Proposition 3.6 ([19, Proposition 3.3]). *For any $\varepsilon \in (0, \mathrm{e}^{-1})$, for every $d \in \mathbb{N}$ and every $A > 0$, there exists a* ReLU*-NN* $\widetilde{\prod}_{d,A,\varepsilon} : [-A, A]^d \to \mathbb{R}$ *such that*

$$\sup_{(x_i)_{i=1}^d \in [-A,A]^d} \left| \prod_{j=1}^{d} x_j - \widetilde{\prod}_{d,A,\varepsilon}(x_1, \dots, x_d) \right| \le \varepsilon. \tag{3.10}$$

There exists a constant C independent of $\varepsilon \in (0, \mathrm{e}^{-1})$, $d \in \mathbb{N}$ and $A \ge 1$ such that

$$\operatorname{size}\Big(\widetilde{\prod}_{d,A,\varepsilon}\Big) \le C\Big(1 + d\log\Big(\frac{dA^d}{\varepsilon}\Big)\Big) \quad \text{and} \quad \operatorname{depth}\Big(\widetilde{\prod}_{d,A,\varepsilon}\Big) \le C\Big(1 + \log(d)\log\Big(\frac{dA^d}{\varepsilon}\Big)\Big). \tag{3.11}$$

Theorem 3.7. *Let $\Lambda \subseteq \mathcal{F}$ be finite and downward closed. Then for every $\varepsilon \in (0, \mathrm{e}^{-1})$ there exists a neural network $\Phi = \{\tilde{H}_{\varepsilon,\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\Lambda} : \mathbb{R}^{|\operatorname{supp}\Lambda|} \to \mathbb{R}^{|\Lambda|}$ such that*

$$\max_{\boldsymbol{\nu}\in\Lambda} \; \|H_{\boldsymbol{\nu}} - \tilde{H}_{\varepsilon,\boldsymbol{\nu}}\|_{L^2(U,\gamma)} \le \varepsilon,$$

and there exists a positive constant C (independent of $m(\Lambda)$, $d(\Lambda)$ and of $\varepsilon \in (0, \mathrm{e}^{-1})$) such that

$$\operatorname{size}(\Phi) \le C|\Lambda| m(\Lambda)^3 \log(1 + m(\Lambda)) d(\Lambda)^2 \log(\varepsilon^{-1}),$$
$$\operatorname{depth}(\Phi) \le C m(\Lambda) \log(1 + m(\Lambda))^2 d(\Lambda) \log(1 + d(\Lambda)) \log(\varepsilon^{-1}).$$

Proof. Fix $\varepsilon \in (0, \mathrm{e}^{-1})$. Throughout this proof, we write $m := m(\Lambda)$, $d := d(\Lambda)$ and we assume w.l.o.g. that $m \ge 1$ and $d \ge 1$ (otherwise $\Lambda = \emptyset$ or $\Lambda = \{\mathbf{0}\}$, and these cases are trivial). Furthermore, with the constant M as defined in (3.5), set

$$A := 1 + (3M)^m, \quad \tilde{\varepsilon} := \varepsilon 2^{-d}.$$

Step 1. We define $\tilde{H}_{\varepsilon,\boldsymbol{\nu}}$ and show that $\|H_{\boldsymbol{\nu}} - \tilde{H}_{\varepsilon,\boldsymbol{\nu}}\|_{L^2(U,\gamma)} \le \varepsilon$.
Let $H_{\mathbf{0}} := 1$ and for $\mathbf{0} \ne \boldsymbol{\nu} \in \Lambda$

$$\tilde{H}_{\tilde{\varepsilon},\boldsymbol{\nu}} := \widetilde{\prod}_{|\boldsymbol{\nu}|_0,A,\varepsilon}((\tilde{H}_{\tilde{\varepsilon},\nu_j}(x_j))_{j\in\operatorname{supp}\boldsymbol{\nu}}).$$

Then for $\mathbf{0} \ne \boldsymbol{\nu} \in \Lambda$

$$\begin{aligned} \|H_{\boldsymbol{\nu}} - \tilde{H}_{\varepsilon,\boldsymbol{\nu}}\|_{L^2(U,\gamma)} \le & \left\| H_{\boldsymbol{\nu}} - \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \tilde{H}_{\tilde{\varepsilon},\nu_j} \right\|_{L^2(U,\gamma)} \\ & + \left\| \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \tilde{H}_{\tilde{\varepsilon},\nu_j} - \widetilde{\prod}_{|\boldsymbol{\nu}|_0,A,\varepsilon}\big((\tilde{H}_{\tilde{\varepsilon},\nu_j})_{j\in\operatorname{supp}\boldsymbol{\nu}}\big) \right\|_{L^2(U,\gamma)}. \end{aligned} \tag{3.12}$$

By Corollary 3.5 (ii) it holds $\sup_{x\in\mathbb{R}} |\tilde{H}_{\tilde{\varepsilon},\nu_j}(x)| \le 1 + (3M)^{\nu_j} \le A$ for all $j \in \operatorname{supp}\boldsymbol{\nu}$. Proposition 3.6 thus implies

$$\left| \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \tilde{H}_{\tilde{\varepsilon},\nu_j}(y_j) - \widetilde{\prod}_{|\boldsymbol{\nu}|_0,A,\varepsilon}\big((\tilde{H}_{\tilde{\varepsilon},\nu_j}(y_j))_{j\in\operatorname{supp}\boldsymbol{\nu}}\big) \right| \le \varepsilon$$

for all $\boldsymbol{y} \in U$ so that $\| \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \tilde{H}_{\nu_j} - \tilde{H}_{\varepsilon,\boldsymbol{\nu}}\|_{L^2(U,\gamma)} \le \varepsilon$. To bound the first term in (3.12) we compute

$$\left\| \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} H_{\nu_j} - \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \tilde{H}_{\tilde{\varepsilon},\nu_j} \right\|_{L^2(U,\gamma)} \le \sum_{j\in\operatorname{supp}\boldsymbol{\nu}} \prod_{\substack{i\in\operatorname{supp}\boldsymbol{\nu}\\ i<j}} \|H_{\nu_i}\|_{L^2(\mathbb{R},\gamma_1)} \cdot \|H_{\nu_j} - \tilde{H}_{\tilde{\varepsilon},\nu_j}\|_{L^2(\mathbb{R},\gamma_1)} \cdot \prod_{\substack{i\in\operatorname{supp}\boldsymbol{\nu}\\ i>j}} \|\tilde{H}_{\tilde{\varepsilon},\nu_i}\|_{L^2(\mathbb{R},\gamma_1)}.$$

For all i it holds $\|H_{\nu_i}\|_{L^2(\mathbb{R},\gamma_1)} = 1$, $\|H_{\nu_i} - \tilde{H}_{\tilde{\varepsilon},\nu_i}\|_{L^2(\mathbb{R},\gamma_1)} \le \tilde{\varepsilon}$ by Corollary 3.5 and thus $\|\tilde{H}_{\tilde{\varepsilon},\nu_i}\|_{L^2(\mathbb{R},\gamma_1)} \le 1 + \tilde{\varepsilon} \le \frac{3}{2}$ (since $\tilde{\varepsilon} \le \varepsilon < \mathrm{e}^{-1}$). Hence

$$\left\| \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} H_{\nu_j} - \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \tilde{H}_{\tilde{\varepsilon},\nu_j} \right\|_{L^2(U,\gamma)} \le |\boldsymbol{\nu}|_0 \tilde{\varepsilon}(1+\tilde{\varepsilon})^{|\boldsymbol{\nu}|_0 - 1} \le \tilde{\varepsilon} d \left(\frac{3}{2}\right)^{d-1} \le \tilde{\varepsilon} 2^d \le \varepsilon,$$

where we used $d(\frac{3}{2})^{d-1} \le 2^d$ for all $d \in \mathbb{N}$.

Step 2. We construct $\Phi = \{\tilde{H}_{\varepsilon,\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\Lambda}$ and provide bounds on the size and depth of Φ.

Let $\Phi_1 : \mathbb{R}^{|\operatorname{supp}\Lambda|} \to \mathbb{R}^{m|\Lambda|}$, with output

$$\Phi_1(\boldsymbol{y}) = \left\{ \tilde{H}_{\tilde{\varepsilon},j}(y_i) \right\}_{i\in\operatorname{supp}\Lambda, j\in\{1,\dots,m\}}. \tag{3.13}$$

Due to Corollary 3.5 for each $j \le m$

$$\begin{aligned} \operatorname{size}(\tilde{H}_{\tilde{\varepsilon},j}) &\le C\Big(1 + j^2\big(\log(j) + \log(-\log(\tilde{\varepsilon}))\big) + j\log\Big(\frac{j}{\tilde{\varepsilon}}\Big)\Big) \\ &= C\bigg(1 + j^2\big(\log(j) + \log(d\log(2) - \log(\varepsilon))\big) + j\Big(\log(j) + d\log(2) - \log(\varepsilon)\Big)\bigg) \\ &\le C\bigg(1 + m^2\log(m) + m^2\log(d) + m^2\log(-\log(\varepsilon)) + md - m\log(\varepsilon)\bigg) \\ &=: CC_0(m,d,\varepsilon), \end{aligned} \tag{3.14}$$

with $C_0(m,d,\varepsilon)$ denoting the term in brackets, and C being a constant independent of m, d and ε. Note that $\log(-\log(\varepsilon))$ is well defined since $-\log(\varepsilon) > 1$ due to $\varepsilon < \mathrm{e}^{-1}$.

To derive a bound on the depth, we observe that by Corollary 3.5

$$\begin{aligned} \operatorname{depth}(\tilde{H}_{\tilde{\varepsilon},j}) &\le C(1 + j\log(j)^2 + j\log(j)\log(-\log(\tilde{\varepsilon})) - \log(j)\log(\tilde{\varepsilon})) \\ &\le C\Big(1 + m\log(m)^2 + m\log(m)\log(-\log(\varepsilon)) + m\log(m)\log(d) - \log(m)\log(\varepsilon)\Big) \\ &=: CC_1(m,d,\varepsilon), \end{aligned} \tag{3.15}$$

where $C_1(m,d,\varepsilon)$ is the term in parentheses, and C is a positive constant that is independent of m, d and ε. Concatenating $\tilde{H}_{\tilde{\varepsilon},j}$ with $CC_1(m,d,\varepsilon) - \operatorname{depth}(\tilde{H}_{\tilde{\varepsilon},j})$ times the identity network $x = \sigma(x) - \sigma(-x)$, we may and will assume that each $\tilde{H}_{\tilde{\varepsilon},j}(y_i)$ in (3.13) has the same depth $CC_1(m,d,\varepsilon)$, and the size is bounded by $CC_0(m,d,\varepsilon) + CC_1(m,d,\varepsilon) \le CC_0(m,d,\varepsilon)$ for a suitable constant C that is independent of m, d and ε.

Next, we let $\Phi_2 : \mathbb{R}^{m|\operatorname{supp}\Lambda|} \to \mathbb{R}^{|\Lambda|}$ denote the network

$$\Phi_2 := \left\{ \widetilde{\prod}_{|\boldsymbol{\nu}|_0, A, \varepsilon} \right\}_{\boldsymbol{\nu}\in\Lambda}. \tag{3.16}$$

Then

$$\Phi_2 \circ \Phi_1 = \left\{ \widetilde{\prod}_{|\boldsymbol{\nu}|_0, A, \varepsilon} \left((\tilde{H}_{\varepsilon/2^d, \nu_j})_{j \in \operatorname{supp} \boldsymbol{\nu}} \right) \right\}_{\boldsymbol{\nu} \in \Lambda} = \Phi.$$

It remains to estimate the size and depth of Φ_2. By Proposition 3.6

$$\begin{aligned} \operatorname{size}\left(\widetilde{\prod}_{|\boldsymbol{\nu}|_0, A, \varepsilon} \right) &\leq C(1 + |\boldsymbol{\nu}|_0 \log(|\boldsymbol{\nu}|_0) + |\boldsymbol{\nu}|_0^2 \log(A) - |\boldsymbol{\nu}|_0 \log(\varepsilon)) \\ &\leq C(1 + d \log(d) + d^2 \log(A) - d \log(\varepsilon)). \end{aligned}$$

By definition of A and M, using $\log(1+x) \leq 1 + \log(x)$ for $x \geq 1$,

$$\log(A) \leq 1 + m \log(3\sqrt{24(m \log(2m) - \log(\varepsilon))}) \leq C(1 + m \log(m) + m \log(-\log(\varepsilon))).$$

Hence

(3.17)

$$\operatorname{size}\left(\widetilde{\prod}_{|\boldsymbol{\nu}|_0, A, \varepsilon} \right) \leq C\Big(1 + d \log(d) + d^2 m \log(m) + d^2 m \log(-\log(\varepsilon)) - d \log(\varepsilon)\Big) =: C D_0(m, d, \varepsilon).$$

In addition, by Propsition 3.6

$$\begin{aligned} \operatorname{depth}\left(\widetilde{\prod}_{|\boldsymbol{\nu}|_0, A, \varepsilon} \right) &\leq C\Big(1 + \log(d)(\log(d) + d \log(A) - \log(\varepsilon))\Big) \\ &\leq C\Big(1 + \log(d)^2 + d \log(d) m \log(m) + d \log(d) m \log(-\log(\varepsilon)) - \log(d) \log(\varepsilon)\Big) \\ &=: C D_1(m, d, \varepsilon). \end{aligned} \tag{3.18}$$

Similar as before, by concatenating $\widetilde{\prod}_{|\boldsymbol{\nu}|_0, A, \varepsilon}$ a suitable number of times with the identity network $x = \sigma(x) - \sigma(-x)$, we can assume that all networks $\widetilde{\prod}_{|\boldsymbol{\nu}|_0, A, \varepsilon}$, $\boldsymbol{\nu} \in \Lambda$, have the same depth, and a uniform bound on the size given by (3.17).

We now sum the size of all subnetworks. First note that the downward closedness of Λ implies $|\operatorname{supp} \Lambda| \leq |\Lambda|$ (cf. (3.8)).

Hence, by (3.13), (3.14) and (3.16), (3.17) it follows that there exists a constant $C > 0$ such that for all $0 < \varepsilon < \mathrm{e}^{-1}$ holds

$$\begin{aligned} \operatorname{size}(\Phi) &\leq C(1 + \operatorname{size}(\Phi_1) + \operatorname{size}(\Phi_2)) \\ &\leq C(1 + (|\operatorname{supp} \Lambda| m) C_0(m, d, \varepsilon) + |\Lambda| D_0(m, d, \varepsilon)) \\ &\leq C|\Lambda| \Big(\big(1 + m^3(\log(m) + d) + m^3 \log(\varepsilon)\big) + \big(d^2 m \log(m) \log(\varepsilon^{-1})\big) \Big) \\ &\leq C(1 + |\Lambda| m^3 \log(1+m) d^2 \log(\varepsilon^{-1})). \end{aligned}$$

Similarly by (3.15) and (3.18)

$$\begin{aligned} \operatorname{depth}(\Phi) &\leq C(1 + \operatorname{depth}(\Phi_1) + \operatorname{depth}(\Phi_2)) \\ &\leq C(1 + C_1(m, d, \varepsilon) + D_1(m, d, \varepsilon)) \\ &\leq C(1 + d \log(1+d) m \log(1+m)^2 \log(\varepsilon^{-1})). \end{aligned}$$

■

Remark 3.8. *The preceding analysis was based on* approximating *Hermite polynomials by* ReLU-*NNs. The so-called "polynomial* ReLU*" activation* ReLU^k*, sometimes also referred to as "rectified power unit" ("RePU"), is capable of* exactly *expressing multivariate polynomials, i.e.* without emulation error. *For an integer $k \geq 2$, this activation function is given by* $\mathrm{ReLU}^k(x) := \max\{x, 0\}^k$.

Evidently, $\mathrm{ReLU}^k \in W^{k,\infty}_{\mathrm{loc}}(\mathbb{R})$*, so that the resulting DNNs will inherit this regularity in the input-output maps arising as their realizations. From [16, Prop. 2.14], we have the following statement.*

Fix $d \in \mathbb{N}$ *and* $k \in \mathbb{N}$*,* $k \geq 2$ *arbitrary. Then there exists a constant* $C > 0$ *(depending on* k *but independent of* d*) such that for any finite, downward closed* $\Lambda \subseteq \mathbb{N}_0^d$ *and for any* $p \in \mathbb{P}_\Lambda$ *there is a* ReLU^k*-NN* $\tilde{p} : \mathbb{R}^d \to \mathbb{R}$ *which realizes* p *exactly and such that* $\mathrm{size}(\tilde{p}) \leq C|\Lambda|$ *and* $\mathrm{depth}(\tilde{p}) \leq C \log(|\Lambda|)$.

4. DNN approximation of analytic functions in $L^2(\mathbb{R}^d, \gamma_d)$.

In this section, we show that certain analytic functions $f \in L^2(\mathbb{R}^d, \gamma_d)$ with finite $d \in \mathbb{N}$ can be approximated at an exponential rate by ReLU-NNs. To state the precise assumption on f, for $\kappa > 0$ introduce the complex open strip

$$S_\kappa := \{z = x + \mathrm{i}y \in \mathbb{C} : |y| < \kappa\} \subset \mathbb{C}\,. \tag{4.1a}$$

For $\boldsymbol{\tau} = (\tau_j)_{j=1}^d \in (0,\infty)^d$, we define the product domains

$$S_{\boldsymbol{\tau}} := \otimes_{j=1}^d S_{\tau_j} \subset \mathbb{C}^d. \tag{4.1b}$$

Assumption 4.1. *There exists* $\boldsymbol{\tau} \in (0,\infty)^d$ *so that* $f : S_{\boldsymbol{\tau}} \to \mathbb{C}$ *is holomorphic. For every* $\mathbf{0} \leq \boldsymbol{\beta} \leq \boldsymbol{\tau}$ *there exists* $B(\boldsymbol{\beta}) > 0$ *such that for all* $\boldsymbol{x} + \mathrm{i}\boldsymbol{y} = (x_j + \mathrm{i}y_j)_{j=1}^d \in S_{\boldsymbol{\beta}}$ *it holds*

$$|f(\boldsymbol{x} + \mathrm{i}\boldsymbol{y})| \leq B(\boldsymbol{\beta}) \exp\left(\sum_{j=1}^d \left(\frac{x_j^2}{4} - 2^{-1/2}|x_j|(\beta_j^2 - \frac{1}{2}y_j^2)^{1/2}\right)\right). \tag{4.2}$$

Condition (4.2) is a growth condition on f on the domains $S_{\boldsymbol{\tau}} \subset \mathbb{C}^d$. It states that f should increase along the real axis in x_j slower than $\exp(\frac{x_j^2}{4})$. The parameters β_j quantify this further, and will determine the rate of convergence. The occurence of the factor $\exp(\frac{x_j^2}{4})$ stems from the fact, that we wish to approximate f in $L^2(\mathbb{R}^d, \gamma_d)$, where the Gaussian γ_d has Lebesgue density $(2\pi)^{-d/2} \exp(-\sum_{j=1}^d \frac{x_j^2}{2})$. Hence f increasing faster than $\exp(\frac{x_j^2}{4})$ would imply $f \notin L^2(\mathbb{R}^d, \gamma_d)$.

4.1. Polynomial approximation.

Recall that the Hermite functions $(h_n)_{n \in \mathbb{N}_0}$ in (2.3) form an ONB of $L^2(\mathbb{R})$. Our analysis in the finite dimensional case is based on the classical paper [7] of E. Hille.

Theorem 4.2 ([7, Theorem 1]). *Let* $\tau > 0$ *and let* $g : S_\tau \to \mathbb{C}$ *be holomorphic and satisfy: for every* $\beta \in (0,\tau)$ *exists* $B(\beta) < \infty$ *such that for all* $x + \mathrm{i}y \in S_\beta$

$$|g(x + \mathrm{i}y)| \leq B(\beta) \exp(-|x|(\beta^2 - y^2)^{1/2}). \tag{4.3}$$

Then for every $\beta \in (0,\tau)$ *exists a constant* $K(\beta)$ *depending on* β *(but independent of* g*,* τ *and* n*) such that for every* $n \in \mathbb{N}$

$$\left| \int_{\mathbb{R}} g(x) h_n(x) \mathrm{d}x \right| \leq K(\beta) B(\beta) \exp(-\beta(2n+1)^{1/2}). \tag{4.4}$$

We recall part of the proof of the theorem in Appendix A. The reason is that the result in [7, Theorem 1] does not explicitly state the dependence of the occurring constants. In the following we wish to repeatedly apply (4.4) coordinatewise to obtain a multivariate version. To this end we need (4.4) to hold for some $K(\beta)$, where $K(\beta)$ is only a function of β but does not depend on g.

To state the multivariate version of Theorem 4.2, with H_n and h_n from (2.1), (2.3), for all $\boldsymbol{\nu} \in \mathbb{N}_0^d$ in the following

$$H_{\boldsymbol{\nu}}(\boldsymbol{x}) := \prod_{j=1}^d H_{\nu_j}(x_j) \qquad \text{and} \qquad h_{\boldsymbol{\nu}}(\boldsymbol{x}) := \prod_{j=1}^d h_{\nu_j}(x_j).$$

Moreover we use standard multivariate notation such as $(0,\boldsymbol{\tau})$ to denote the cube $\times_{j=1}^d(0,\tau_j)\subset\mathbb{R}^d$ for $\boldsymbol{\tau}=(\tau_j)_{j=1}^d\in(0,\infty)^d$.

Corollary 4.3. *Let $d\in\mathbb{N}$, $\boldsymbol{\tau}\in(0,\infty)^d$ and let $F:S_{\boldsymbol{\tau}}\to\mathbb{C}$ be holomorphic and satisfy: for every $\boldsymbol{\beta}\in(0,\boldsymbol{\tau})$ exists $B(\boldsymbol{\beta})>0$ such that for all $\boldsymbol{x}+\mathrm{i}\boldsymbol{y}=(x_j+\mathrm{i}y_j)_{j=1}^d\in S_{\boldsymbol{\beta}}$*

$$|F(\boldsymbol{x}+\mathrm{i}\boldsymbol{y})|\le B(\boldsymbol{\beta})\exp\left(-\sum_{j=1}^d|x_j|(\beta_j^2-y_j^2)^{1/2}\right). \tag{4.5}$$

With $K(\beta_j)>0$ as in Theorem 4.2 then holds for every $\boldsymbol{\nu}\in\mathbb{N}_0^d$ and every $\boldsymbol{\beta}\in(0,\boldsymbol{\tau})$

$$|\langle F,h_{\boldsymbol{\nu}}\rangle_{L^2(\mathbb{R}^d)}|\le\left(\prod_{j=1}^d K(\beta_j)\right)B(\boldsymbol{\beta})\exp\left(-\sum_{j=1}^d\beta_j(2\nu_j+1)^{1/2}\right). \tag{4.6}$$

Proof. Fix $\boldsymbol{\nu}\in\mathbb{N}_0^d$ and $\boldsymbol{\beta}\in(0,\boldsymbol{\tau})$. Then (4.5) and Theorem 4.2 imply for all $z_j=x_j+\mathrm{i}y_j\in S_{\tau_j}$, $j\in\{2,\dots,d\}$,

$$\left|\int_{\mathbb{R}}h_{\nu_1}(x_1)F(x_1,z_2,\dots,z_d)\mathrm{d}x_1\right|\le K(\beta_1)B(\boldsymbol{\beta})\exp\left(-\sum_{j=2}^d|x_j|(\beta_j^2-y_j^2)^{1/2}\right)\exp(-\beta_1(2\nu_1+1)^{1/2}). \tag{4.7}$$

The function $(z_2,\dots,z_d)\mapsto\int_{\mathbb{R}}h_{\nu_1}(x_1)F(x_1,z_2,\dots,z_d)\mathrm{d}x_1$ is well-defined and holomorphic (e.g., by the theorem in [15]) on $S_{\tau_2}\times\cdots\times S_{\tau_d}$. Thus (4.7) and Theorem 4.2 imply for all $z_j=x_j+\mathrm{i}y_j\in S_{\tau_j}$, $j\in\{3,\dots,d\}$,

$$\begin{aligned}&\left|\int_{\mathbb{R}}h_{\nu_2}(x_2)\int_{\mathbb{R}}h_{\nu_1}(x_1)F(x_1,x_2,z_3,\dots,z_d)\mathrm{d}x_1\mathrm{d}x_2\right|\\&\qquad\le K(\beta_1)K(\beta_2)B(\boldsymbol{\beta})\exp\left(-\sum_{j=3}^d|x_j|(\beta_j^2-y_j^2)^{1/2}\right)\exp(-\beta_1(2\nu_1+1)^{1/2}-\beta_2(2\nu_2+1)^{1/2}).\end{aligned}$$

Repeating the argument another $d-2$ times concludes the proof. ■

Our goal is to bound the Fourier coefficients w.r.t. the orthonormal Hermite polynomials $(H_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathbb{N}_0^d}$ in $L^2(\mathbb{R}^d,\gamma_d)$. Theorem 4.2 and Corollary 4.3 instead provide bounds on the Fourier coefficients w.r.t. the Hermite functions $(h_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathbb{N}_0^d}$ in $L^2(\mathbb{R})$. The following multivariate version of Lemma 2.1 relates the two.

Lemma 4.4. *Let $d\in\mathbb{N}$ and set*

$$\Theta:L^2(\mathbb{R}^d,\gamma_d)\to L^2(\mathbb{R}^d):f(\boldsymbol{x})\mapsto f(2^{1/2}\boldsymbol{x})\frac{\exp\left(-\frac{\|\boldsymbol{x}\|_2^2}{2}\right)}{\pi^{\frac{d}{4}}}.$$

Then Θ is an isometric isomorphism and $\Theta(H_{\boldsymbol{\nu}})=h_{\boldsymbol{\nu}}$ for all $\boldsymbol{\nu}\in\mathbb{N}_0^d$. In particular, for every $f\in L^2(\mathbb{R}^d,\gamma_d)$

$$\langle f,H_{\boldsymbol{\nu}}\rangle_{L^2(\mathbb{R}^d,\gamma_d)}=\langle\Theta(f),h_{\boldsymbol{\nu}}\rangle_{L^2(\mathbb{R}^d)}\qquad\forall\boldsymbol{\nu}\in\mathbb{N}_0^d. \tag{4.8}$$

Equation (4.8) shows that, as long as $\Theta(f)$ satisfies the Assumptions of Corollary 4.3, we have a bound of the type (4.4) on the Hermite coefficients $|\langle f,H_{\boldsymbol{\nu}}\rangle_{L^2(\mathbb{R},\gamma_1)}|$. Upon observing that $\Theta(f)$ satisfies Assumption 4.1 if f satisfies the assumptions of Corollary 4.3, A version of this theorem has already been shown with essentially the same argument in [1, Lemma 4.6, Thm. 4.1]. For completeness and because our statement and assumptions slightly differ[3] from [1], we provide the proof in the appendix.

[3] In particular we allow for stronger growth of f as $x\to\pm\infty$.

Theorem 4.5. *Let $f : \mathbb{R}^d \to \mathbb{R}$ satisfy Assumption 4.1 for some $\boldsymbol{\tau} \in (0,\infty)^d$.*

Then for all $\boldsymbol{\beta} \in (0,\boldsymbol{\tau})$ exist $C > 0$ (depending on $\boldsymbol{\beta}$) such that there holds, with $K(\beta_j)$ as in Theorem 4.2,

(i) for all $\boldsymbol{\nu} \in \mathbb{N}_0^d$

$$|\langle f, H_{\boldsymbol{\nu}}\rangle_{L^2(\mathbb{R}^d,\gamma_d)}| \le \pi^{-d/4} B(\boldsymbol{\beta}) \left(\prod_{j=1}^d K(\beta_j)\right) \exp\left(-\sum_{j=1}^d \beta_j (2\nu_j+1)^{1/2}\right), \tag{4.9}$$

(ii) for all $\varepsilon \in (0,1)$ with

$$\Lambda_\varepsilon := \left\{\boldsymbol{\nu} \in \mathbb{N}_0^d \,:\, \nu_j \le \left(\frac{\log(\varepsilon)}{\beta_j}\right)^2\right\} \tag{4.10}$$

and

$$\delta(\boldsymbol{\beta}) := \left(\prod_{j=1}^d \beta_j\right)^{\frac{1}{d}} \tag{4.11}$$

holds

$$\left\| f - \sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} \langle f, H_{\boldsymbol{\nu}}\rangle H_{\boldsymbol{\nu}} \right\|_{L^2(\mathbb{R}^d,\gamma_d)} \le C\varepsilon \le C \exp\left(-2^{-\frac{1}{2}}\delta(\boldsymbol{\beta})|\Lambda_\varepsilon|^{\frac{1}{2d}}\right). \tag{4.12}$$

4.2. ReLU neural network approximation. The polynomial approximation result in the previous subsection together with the ReLU approximation result of Hermite polynomials provided in Sec. 3 yield exponential convergence in $L^2(\mathbb{R}^d,\gamma_d)$ of DNN approximations with ReLU activations. We prepare the proof of the theorem by showing two basic properties of Λ_ε.

Lemma 4.6. *Fix $\boldsymbol{\beta} \in (0,\infty)^d$ and let Λ_ε be as in (4.10). Then for all*

$$\varepsilon \in \left(0, \exp\left(-\max_{j\in\{1,\dots,d\}} \beta_j\right)\right) \tag{4.13}$$

holds $|\Lambda_\varepsilon| \le 2^d \frac{\log(\varepsilon)^{2d}}{\prod_{j=1}^d \beta_j^2}$. Furthermore, $m(\Lambda_\varepsilon) \le \log(\varepsilon)^2 (\sum_{j=1}^d \beta_j^{-2})$.

Proof. Due to (4.13) we have $(\log(\varepsilon)/\beta_j)^2 \ge 1$ and therefore

$$|\Lambda_\varepsilon| \le \prod_{j=1}^d \left(1 + \left(\frac{\log(\varepsilon)}{\beta_j}\right)^2\right) \le 2^d \frac{\log(\varepsilon)^{2d}}{\prod_{j=1}^d \beta_j^2}.$$

In addition

$$m(\Lambda_\varepsilon) = \max_{\boldsymbol{\nu}\in\Lambda_\varepsilon} |\boldsymbol{\nu}|_1 \le \sum_{j=1}^d \frac{\log(\varepsilon)^2}{\beta_j^2}.$$

■

Theorem 4.7. *Let $f : \mathbb{R}^d \to \mathbb{R}$ satisfy Assumption 4.1 for some $\boldsymbol{\tau} \in (0,\infty)^d$.*

Then for all $\boldsymbol{\beta} \in (0,\boldsymbol{\tau})$ exists $C > 0$ (depending on $\boldsymbol{\beta}$ and d) such that for all $N \in \mathbb{N}$ exists a ReLU *network $\tilde{f}_N : \mathbb{R}^d \to \mathbb{R}$ such that*

$$\mathrm{size}(\tilde{f}_N) \le CN(1+\log(N)), \qquad \mathrm{depth}(\tilde{f}_N) \le CN^{\frac{3}{2d+7}}(1+\log(N))^2, \tag{4.14}$$

and

$$\|f - \tilde{f}_N\|_{L^2(\mathbb{R}^d,\gamma_d)} \le C \exp\left(-2^{-\frac{1}{2}}\delta(\boldsymbol{\beta})N^{\frac{1}{2d+7}}\right). \tag{4.15}$$

Proof. For $M > 0$, define with $\delta(\boldsymbol{\beta})$ as in (4.11)

$$\varepsilon_M := \exp\left(-\left(\frac{M\prod_{j=1}^d \beta_j^2}{2^d}\right)^{\frac{1}{2d}}\right) = \exp\left(-2^{-\frac{1}{2}}\delta(\boldsymbol{\beta})M^{\frac{1}{2d}}\right) . \tag{4.16}$$

Assume $M > 2$ is chosen so large that $\varepsilon_M < \exp\left(-\max_{j\in\{1,\dots,d\}}\beta_j\right)$, i.e. (4.13) holds.

By Lemma 4.6 it holds $|\Lambda_{\varepsilon_M}| \le M$. Furthermore, (4.12) with ε_M in place of ε implies

$$\left\| f - \sum_{\boldsymbol{\nu}\in\Lambda_{\varepsilon_M}} \langle f, H_{\boldsymbol{\nu}}\rangle H_{\boldsymbol{\nu}} \right\|_{L^2(\mathbb{R}^d,\gamma_d)} \le C\varepsilon_M \le C\exp(-2^{-\frac{1}{2}}\delta(\boldsymbol{\beta})M^{\frac{1}{2d}}). \tag{4.17}$$

Next, let $(\tilde{H}_{\varepsilon_M,\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_{\varepsilon_M}}$ be the ReLU approximation from Theorem 3.7. As the coefficients $\langle f, H_{\boldsymbol{\nu}}\rangle$ are summable according to (4.9), we get

$$\left\| \sum_{\boldsymbol{\nu}\in\Lambda_{\varepsilon_M}} \langle f, H_{\boldsymbol{\nu}}\rangle |H_{\boldsymbol{\nu}} - \tilde{H}_{\varepsilon_M,\boldsymbol{\nu}}| \right\|_{L^2(\mathbb{R}^d,\gamma_d)} \le \varepsilon_M \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^d} |\langle f, H_{\boldsymbol{\nu}}\rangle| \le C\exp(-2^{-\frac{1}{2}}\delta(\boldsymbol{\beta})M^{\frac{1}{2d}}).$$

Together with (4.17) we observe that the network

$$\tilde{g}_M := \sum_{\boldsymbol{\nu}\in\Lambda_{\varepsilon_M}} \langle f, H_{\boldsymbol{\nu}}\rangle \tilde{H}_{\varepsilon_M,\boldsymbol{\nu}}$$

satisfies the error bound

$$\|f - \tilde{g}_M\|_{L^2(\mathbb{R}^d,\gamma_d)} \le C\exp\left(-2^{-\frac{1}{2}}\delta(\boldsymbol{\beta})M^{\frac{1}{2d}}\right). \tag{4.18}$$

Next we bound the size and depth of $\tilde{g}_M$. By Lemma 4.6 and (4.16)

$$m(\Lambda_{\varepsilon_M}) \le \log(\varepsilon_M)^2 \sum_{j=1}^d \beta_j^{-2} \le M^{\frac{1}{d}}\delta(\boldsymbol{\beta})^2 \sum_{j=1}^d \beta_j^{-2} \le CM^{\frac{1}{d}}$$

for some $C = C(d, \boldsymbol{\beta})$. It holds $\operatorname{size}(\tilde{g}_M) \le C|\Lambda_{\varepsilon_M}| + \operatorname{size}(\tilde{H}_{\varepsilon_M,\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_{\varepsilon_M}}$. By Theorem 3.7

$$\begin{aligned}\operatorname{size}(\tilde{g}_M) &\le C|\Lambda_{\varepsilon_M}| + C|\Lambda_{\varepsilon_M}| m(\Lambda_{\varepsilon_M})^3 \log(1 + m(\Lambda_{\varepsilon_M})) d(\Lambda_{\varepsilon_M})^2 |\log(\varepsilon_M)| \\ &\le CM + CMM^{\frac{3}{d}}\log(CM)d^2M^{\frac{1}{2d}} \\ &\le C(1+M)^{1+\frac{3}{d}+\frac{1}{2d}}(1+\log(M)),\end{aligned} \tag{4.19}$$

where C depends on $\boldsymbol{\beta}$ and d and may change its value after each inequality in the above computation. Similarly, using again Theorem 3.7,

$$\begin{aligned}\operatorname{depth}(\tilde{g}_M) &\le C + Cm(\Lambda_{\varepsilon_M})\log(1 + m(\Lambda_{\varepsilon_M}))^2 d(\Lambda_{\varepsilon_M})\log(1 + d(\Lambda_{\varepsilon_M}))|\log(\varepsilon_M)| \\ &\le C + CM^{\frac{1}{d}}\log(1+M)^2 d\log(1+d)M^{\frac{1}{2d}} \\ &\le CM^{\frac{1}{d}+\frac{1}{2d}}(1+\log(M))^2.\end{aligned} \tag{4.20}$$

Setting $\tilde{f}_N := \tilde{g}_M$ with $M := N^{\frac{2d}{2d+7}} - 1$, (4.18), (4.19), (4.20) imply the error, size and depth bounds (4.14) and (4.15).

Finally, the condition $\varepsilon_M \le \exp\left(-\max_{j\in\{1,\dots,d\}}\beta_j\right)$ corresponds to $M \ge M_0$ for some fixed M_0 depending on d and $\boldsymbol{\beta}$. Since the theorem holds for all $M \ge M_0$, it remains true for all $M \in \mathbb{N}$ after possibly adjusting the constant C. ■

5. DNN approximation of infinite-parametric, analytic functions in $L^2(\mathbb{R}^{\mathbb{N}}, \gamma)$.

In this section we consider the ReLU-NN approximation of certain *countably-parametric*, analytic maps from $U = \mathbb{R}^{\mathbb{N}}$ to $\mathbb{R}$ in $L^2(\mathbb{R}^{\mathbb{N}}, \gamma)$. Such maps arise as solutions of operator equations with Gaussian random field inputs, which are represented in an affine-parametric fashion, via a Parseval frame [14] such as e.g. a Karhunen-Loève or a Lévy-Cieselskii expansion of the GRF. We discuss an example in Sec. 6. The proof of NN approximation bounds proceeds in two stages. First, a polynomial chaos approximation is constructed based on the results in [5], and second, this approximation is emulated by a deep ReLU-NN using our results from the preceding sections.

5.1. Wiener polynomial chaos approximation.

We recall the notion of $(\boldsymbol{b}, \xi, \delta)$-holomorphy from [5, Def. 6.1].

Definition 5.1 ($(\boldsymbol{b}, \xi, \delta)$-Holomorphy). *Let $\boldsymbol{b} = (b_j)_{j\in\mathbb{N}} \in (0,\infty)^{\mathbb{N}}$ and let $\xi > 0$, $\delta > 0$. We say that $\boldsymbol{\varrho} \in (0,\infty)^N$ is $(\boldsymbol{b}, \xi)$-**admissible** if for every $N \in \mathbb{N}$*

$$\sum_{j=1}^{N} b_j \varrho_j \leq \xi \,. \tag{5.1}$$

*A real-valued function $u \in L^2(U, \gamma)$ is called $(\boldsymbol{b}, \xi, \delta)$-**holomorphic** if*

(i) for every finite $N \in \mathbb{N}$ there exists $u^N : \mathbb{R}^N \to \mathbb{R}$, which, for every $(\boldsymbol{b}, \xi)$-admissible $\boldsymbol{\varrho} \in (0,\infty)^N$, admits a holomorphic extension (denoted again by u^N) from $\mathcal{S}_{\boldsymbol{\varrho}} \to \mathbb{C}$; moreover for all $N < M$

$$u^N(y_1, \ldots, y_N) = u^M(y_1, \ldots, y_N, 0, \ldots, 0) \qquad \forall (y_j)_{j=1}^N \in \mathbb{R}^N, \tag{5.2}$$

(ii) for every $N \in \mathbb{N}$ there exists $\varphi_N : \mathbb{R}^N \to \mathbb{R}_+$ such that $\|\varphi_N\|_{L^2(\mathbb{R}^N, \gamma_N)} \leq \delta$ and

$$\sup_{\boldsymbol{\varrho}\in(0,\infty)^N \text{ is } (\boldsymbol{b},\xi)-adm.} \ \sup_{\boldsymbol{z}\in\mathcal{B}(\boldsymbol{\varrho})} |u^N(\boldsymbol{y} + \boldsymbol{z})| \leq \varphi_N(\boldsymbol{y}) \qquad \forall \boldsymbol{y} \in \mathbb{R}^N, \tag{5.3}$$

(iii) with $\hat{u}^N : U \to \mathbb{R}$ defined by $\hat{u}^N(\boldsymbol{y}) := u^N(y_1, \ldots, y_N)$ for $\boldsymbol{y} \in U$ it holds

$$\lim_{N\to\infty} \|u - \hat{u}^N\|_{L^2(U,\gamma)} = 0. \tag{5.4}$$

In the following, for $u : U \to \mathbb{R}$ as in Def. 5.1, we set

$$u_{\boldsymbol{\nu}} := \int_U u(\boldsymbol{y}) H_{\boldsymbol{\nu}}(\boldsymbol{y}) \mathrm{d}\gamma(\boldsymbol{y}) \in \mathbb{R}, \quad \boldsymbol{\nu} \in \mathcal{F},$$

which are the so-called Wiener-Hermite polynomial chaos (PC) expansion coefficients. They are well-defined since $u \in L^2(U, \gamma)$ and $H_{\boldsymbol{\nu}} \in L^2(U, \gamma)$, and thus $\boldsymbol{y} \mapsto u(\boldsymbol{y}) H_{\boldsymbol{\nu}}(\boldsymbol{y}) \in L^1(U, \gamma)$.

The following theorem specifies Hermite PC coefficient summability, see [5, Corollary 7.9].

Theorem 5.2. *Let u be $(\boldsymbol{b}, \xi, \delta)$-holomorphic for some $\boldsymbol{b} \in \ell^p(\mathbb{N})$ and some $p \in (0, \frac{2}{3})$. Then $(u_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^{2p/(2-p)}(\mathcal{F})$.*

Since $p \in (0, \frac{2}{3})$, Theorem 5.2 implies $(|u_{\boldsymbol{\nu}}|)_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^1(\mathcal{F}) \hookrightarrow \ell^2(\mathcal{F})$. Since $(H_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ is an orthonormal basis of $L^2(U, \gamma)$, the expansion

$$u = \sum_{\boldsymbol{\nu}\in\mathcal{F}} u_{\boldsymbol{\nu}} H_{\boldsymbol{\nu}}, \tag{5.5}$$

converges in $L^2(U, \gamma)$. Truncating this expansion yields an approximation to u. Proving convergence rates of N-term truncated Wiener-Hermite pc expansions requires a more specific result however. It is given in the next theorem that is shown in [5, Thm. 7.8, Lemmata 9.5 and 9.6].

Theorem 5.3. *Let u be $(\boldsymbol{b},\xi,\delta)$-holomorphic for some $\boldsymbol{b}\in\ell^p(\mathbb{N})$ and some $p\in(0,\frac{2}{3})$. Let $r>2/p-1$.*

Then there exists $K>0$ such that with

$$c_{\boldsymbol{\nu}} := \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \max\left\{1, K b_j^{p-1}\right\}^2 \nu_j^r \qquad \boldsymbol{\nu}\in\mathcal{F}, \tag{5.6}$$

it holds

(i) $(c_{\boldsymbol{\nu}}^{-1})_{\boldsymbol{\nu}\in\mathcal{F}}\in\ell^{\frac{p}{2(1-p)}}$,

(ii) $\sum_{\boldsymbol{\nu}\in\mathcal{F}} c_{\boldsymbol{\nu}} u_{\boldsymbol{\nu}}^2<\infty$.

In the following, for $c_{\boldsymbol{\nu}}$ as in (5.6) we let similar to (4.10) for $\varepsilon>0$

$$\Lambda_\varepsilon := \{\boldsymbol{\nu}\in\mathcal{F} \,:\, c_{\boldsymbol{\nu}}^{-1}\ge\varepsilon\}. \tag{5.7}$$

It is easy to see that the definition of $c_{\boldsymbol{\nu}}$ in (5.6) implies Λ_ε to be finite and downward closed.

Corollary 5.4. *Consider the setting of Theorem 5.3. Then for every $\varepsilon>0$*

$$\left\| u-\sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} u_{\boldsymbol{\nu}} H_{\boldsymbol{\nu}}\right\|_{L^2(U,\gamma)} \le \varepsilon^{1/2}\left(\sum_{\boldsymbol{\nu}\in\mathcal{F}} c_{\boldsymbol{\nu}} u_{\boldsymbol{\nu}}^2\right)^{1/2}.$$

In addition,

$$\varepsilon \le \|(c_{\boldsymbol{\nu}}^{-1})_{\boldsymbol{\nu}\in\mathcal{F}}\|_{\ell^{p/(2(1-p))}} |\Lambda_\varepsilon|^{-\frac{2(1-p)}{p}} \tag{5.8}$$

so that in particular with the finite constant $C:=(\|(c_{\boldsymbol{\nu}}^{-1})_{\boldsymbol{\nu}\in\mathcal{F}}\|_{\ell^{p/(2(1-p))}}\sum_{\boldsymbol{\nu}\in\mathcal{F}} c_{\boldsymbol{\nu}} u_{\boldsymbol{\nu}}^2)^{1/2}$ holds

$$\left\| u-\sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} u_{\boldsymbol{\nu}} H_{\boldsymbol{\nu}}\right\|_{L^2(U,\gamma)} \le C|\Lambda_\varepsilon|^{-\frac{1}{p}+1}.$$

Proof. By (5.5) and the orthogonality of the $(H_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ in $L^2(U,\gamma)$,

$$\left\| u-\sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} u_{\boldsymbol{\nu}} H_{\boldsymbol{\nu}}\right\|_{L^2(U,\gamma)} = \left(\sum_{\boldsymbol{\nu}\in\mathcal{F}\backslash\Lambda_\varepsilon} u_{\boldsymbol{\nu}}^2\right)^{1/2}.$$

It holds

$$\sum_{\boldsymbol{\nu}\in\mathcal{F}\backslash\Lambda_\varepsilon} u_{\boldsymbol{\nu}}^2 = \sum_{\boldsymbol{\nu}\in\mathcal{F}\backslash\Lambda_\varepsilon} u_{\boldsymbol{\nu}}^2 c_{\boldsymbol{\nu}} c_{\boldsymbol{\nu}}^{-1} \le \sum_{\boldsymbol{\nu}\in\mathcal{F}} u_{\boldsymbol{\nu}}^2 c_{\boldsymbol{\nu}} \sup_{\boldsymbol{\nu}\in\mathcal{F}\backslash\Lambda_\varepsilon} c_{\boldsymbol{\nu}}^{-1} \le \varepsilon \sum_{\boldsymbol{\nu}\in\mathcal{F}} u_{\boldsymbol{\nu}}^2 c_{\boldsymbol{\nu}}$$

by definition of $\Lambda_\varepsilon=\{\boldsymbol{\nu}\,:\,c_{\boldsymbol{\nu}}^{-1}\ge\varepsilon\}$.

By Theorem 5.3 we have $\sum_{\boldsymbol{\nu}\in\mathcal{F}} u_{\boldsymbol{\nu}}^2 c_{\boldsymbol{\nu}}<\infty$ and $(c_{\boldsymbol{\nu}}^{-1})_{\boldsymbol{\nu}\in\mathcal{F}}\in\ell^{p/(2(1-p))}$. Hence, using $c_{\boldsymbol{\nu}}\le\varepsilon^{-1}$ for all $\boldsymbol{\nu}\in\Lambda_\varepsilon$,

$$|\Lambda_\varepsilon| = \sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} c_{\boldsymbol{\nu}}^{-\frac{p}{2(1-p)}} c_{\boldsymbol{\nu}}^{\frac{p}{2(1-p)}} \le \varepsilon^{-\frac{p}{2(1-p)}} \sum_{\boldsymbol{\nu}\in\mathcal{F}} c_{\boldsymbol{\nu}}^{-\frac{p}{2(1-p)}}$$

and consequently

$$\varepsilon \le |\Lambda_\varepsilon|^{-\frac{2(1-p)}{p}} \|(c_{\boldsymbol{\nu}}^{-1})_{\boldsymbol{\nu}\in\mathcal{F}}\|_{\ell^{p/(2(1-p))}}.$$

In all

$$\left\| u - \sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} u_{\boldsymbol{\nu}} H_{\boldsymbol{\nu}} \right\|_{L^2(U,\gamma)} \le \varepsilon^{1/2} \left(\sum_{\boldsymbol{\nu}\in\mathcal{F}} c_{\boldsymbol{\nu}} u_{\boldsymbol{\nu}}^2 \right)^{1/2} \le |\Lambda_\varepsilon|^{-\frac{1}{p}+1} \left(\|(c_{\boldsymbol{\nu}}^{-1})_{\boldsymbol{\nu}\in\mathcal{F}}\|_{\ell^{p/(2(1-p))}} \sum_{\boldsymbol{\nu}\in\mathcal{F}} c_{\boldsymbol{\nu}} u_{\boldsymbol{\nu}}^2 \right)^{1/2}$$

as claimed. ■

5.2. ReLU **neural network approximation.** Let again $(c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ be as in (5.6) with some $\boldsymbol{b} \in \ell^p(\mathbb{N})$ and $K > 0$, and let Λ_ε be as in (5.7). As in [25], we investigate the quantities $m(\Lambda_\varepsilon)$ and $d(\Lambda_\varepsilon)$ defined in (3.9), as $\varepsilon \searrow 0$.

Lemma 5.5. *Assume that there exists $C_0 > 0$, $s > 0$ and $p > 0$ such that $\boldsymbol{b} = (b_j)_{j\in\mathbb{N}} \in \ell^p(\mathbb{N})$ and $b_j \ge C_0 j^{-\frac{s}{2(1-p)}}$ for all $j \in \mathbb{N}$. Let $(c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ be as in (5.6) for this $\boldsymbol{b}$ and some $K > 0$. Then*

$$d(\Lambda_\varepsilon) = o(\log(|\Lambda_\varepsilon|)) \quad \text{and} \quad m(\Lambda_\varepsilon) = O\left(|\Lambda_\varepsilon|^{\frac{s}{r}}\right) \qquad \text{as } \varepsilon \searrow 0. \tag{5.9}$$

Proof. With K as in (5.6) set

$$\hat{\varrho}_j := \max\{1, K b_j^{p-1}\}.$$

Throughout this proof we assume wlog that $(b_j)_{j\in\mathbb{N}}$ is monotonically decreasing (otherwise permute the sequence $(b_j)_{j\in\mathbb{N}}$ accordingly).

Denote by $(x_j)_{j\in\mathbb{N}}$ a monotonically decreasing rearrangement of $(c_{\boldsymbol{\nu}}^{-1})_{\boldsymbol{\nu}\in\mathcal{F}}$. Since $b_j^{p-1} \to \infty$, there exists $C_1 > 0$ such that $\hat{\varrho}_j \le C_1 b_j^{p-1}$ for all j. We have $c_{\boldsymbol{e}_j}^{-1} = \hat{\varrho}_j^{-2} \ge C_1^{-2} b_j^{-2(p-1)} = C_1^{-2} b_j^{2(1-p)}$. Since b_j is monotonically decreasing, by definition of x_j it must hold $x_j \ge C_1^{-2} b_j^{2(1-p)}$. With the assumption $b_j \ge C_0 j^{-\frac{s}{2(1-p)}}$ we get

$$x_j \ge C_1^{-2} C_0 j^{-s} = C_2 j^{-s}. \tag{5.10}$$

We will show that there are fixed constants C_3, C_4, $C_5 > 0$ depending on $(\hat{\varrho}_j)_{j\in\mathbb{N}}$ but independent of $d \in \mathbb{N}$ so that there holds

$$\max_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,|\boldsymbol{\nu}|_0 = d\}} c_{\boldsymbol{\nu}}^{-1} \le C_3 d^{-C_4 d} \qquad \text{and} \qquad \max_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,|\boldsymbol{\nu}| = m\}} c_{\boldsymbol{\nu}}^{-1} \le C_5 m^{-r}\ . \tag{5.11}$$

Denote $F(t) := C_3 t^{-C_4 t}$. Then $F : [1, \infty) \to (0, C_3]$ is strictly monotonically decreasing and bijective. Hence $F^{-1} : (0, C_3] \to [1, \infty)$ is strictly decreasing and bijective. Using (5.11) and (5.10) it holds

$$F(d(\Lambda_N)) \ge \max_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,|\boldsymbol{\nu}|_0 = d(\Lambda_N)\}} c_{\boldsymbol{\nu}}^{-1} \ge \max_{\{\boldsymbol{\nu}\in\Lambda_N\,:\,|\boldsymbol{\nu}|_0 = d(\Lambda_N)\}} c_{\boldsymbol{\nu}}^{-1} \ge \min_{\boldsymbol{\nu}\in\Lambda_N} c_{\boldsymbol{\nu}}^{-1} = x_N \ge C_2 N^{-s}.$$

If N is so large that $C_2 N^{-s} \le C_3$, we may apply F^{-1} on both sides and conclude that $d(\Lambda_N) \le F^{-1}(C_2 N^{-s})$. Since $F^{-1}(t) = o(-\log(t))$ as $t \to 0$, we obtain

$$d(\Lambda_N) = o(-\log(N^{-s})) = o(\log(N))$$

as $N \to \infty$. Similarly, letting $G(t) := C_5 t^{-r}$ and observing that $G^{-1}(t) = (t/C_5)^{-1/r} = O(t^{-1/r})$ as $t \to 0$, one shows that

$$m(\Lambda_N) \le G^{-1}(C_2 N^{-s}) = O(N^{\frac{s}{r}})$$

as $N \to \infty$.

It remains to verify (5.11). Without loss of generality we assume $(\hat{\varrho}_j^{-1})_{j\in\mathbb{N}}$ to be monotonically decreasing. Using Hölder's inequality and the fact that $(\hat{\varrho}_j^{-1})_{j\in\mathbb{N}} \in \ell^q(\mathbb{N})$ with $q := p/(1-p)$ (since $\hat{\varrho}_j^{-1} \sim b_j^{1-p}$ and $(b_j)_{j\in\mathbb{N}} \in \ell^p(\mathbb{N})$) one can show that $\hat{\varrho}_j^{-1} \le \|(\hat{\varrho}_j^{-1})_{j\in\mathbb{N}}\|_{\ell^q(\mathbb{N})} j^{-1/q}$ for all $j \in \mathbb{N}$ (see for example [25, Lemma 2.9]). Therefore with $C_6 := \|(\hat{\varrho}_j^{-1})_{j\in\mathbb{N}}\|_{\ell^q(\mathbb{N})}$

$$\max_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,|\boldsymbol{\nu}|_0=d\}} c_{\boldsymbol{\nu}}^{-1} = \prod_{j=1}^{d} \hat{\varrho}_j^{-2} \le \prod_{j=1}^{d} (C_6 j^{-\frac{1}{q}})^2 \le C_6^{2d} (d!)^{-\frac{2}{q}} \le C_6^{2d} (\mathrm{e}^{-d} d^d)^{-\frac{2}{q}} \le C_6^{2d} \mathrm{e}^{d\frac{2}{q}} d^{-d\frac{2}{q}}.$$

This implies the first inequality in (5.11). To show the second inequality we note that

$$\max_{|\boldsymbol{\nu}|_1=m} c_{\boldsymbol{\nu}}^{-1} = \max_{|\boldsymbol{\nu}|_1=m} \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \hat{\varrho}_j^{-1} \nu_j^{-r}. \tag{5.12}$$

Observe that for $\boldsymbol{\nu} = \boldsymbol{\mu} + \boldsymbol{e}_i$

$$\frac{c_{\boldsymbol{\nu}}^{-1}}{c_{\boldsymbol{\mu}}^{-1}} = \frac{\prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \hat{\varrho}_j^{-1}\nu_j^{-r}}{\prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \hat{\varrho}_j^{-1}\mu_j^{-r}} = \begin{cases} \hat{\varrho}_i^{-1} & \text{if } \nu_i = 0, \\ \left(\frac{\nu_i}{\nu_i+1}\right)^r & \text{otherwise.} \end{cases}$$

By definition $\hat{\varrho}_j = \max\{1, K\varrho_j\} \ge 1$ for all j and thus $\hat{\varrho}_j^{-1} \le 1$ for all $j \in \mathbb{N}$. Now suppose that $J \in \mathbb{N}$ is so large that $\hat{\varrho}_j^{-1} \le 2^{-r}$ for all $j \ge J$. Then for all $m \ge J$, since $(n/(n+1))^r$ is monotonically increasing as a function of $n \in \mathbb{N}$,

$$\max_{|\boldsymbol{\nu}|_1=m} \prod_{j\in\operatorname{supp}\boldsymbol{\nu}} \hat{\varrho}_j^{-1}\nu_j^{-r} \le \prod_{n=1}^{m-J} \left(\frac{n}{n+1}\right)^r = (m-J+1)^{-r}.$$

Together with (5.12) this implies the second inequality in (5.11). ■

We are now in position to state our main result in this section. It provides ReLU-NN expression rates for countably-parametric, $(\boldsymbol{b}, \xi, \delta)$-holomorphic maps.

Theorem 5.6. *Let $u : U \to \mathbb{R}$ be $(\boldsymbol{b}, \xi, \delta)$-holomorphic for some $\boldsymbol{b} \in \ell^p(\mathbb{N})$ with a $p \in (0, \frac{2}{3})$. Fix $\theta > 0$ arbitrarily small.*

Then there exists a constant $C > 0$ (depending on u and on θ) such that for every $N \in \mathbb{N}$ there exists a ReLU*-NN $\tilde{u}_N$ with*

$$\|u(\boldsymbol{y}) - \tilde{u}_N(\boldsymbol{y})\|_{L^2(U,\gamma)} \le C N^{-\frac{1}{p}+1}, \tag{5.13}$$

and it holds

$$\operatorname{size}(\tilde{u}_N) \le C N^{1+\theta}, \qquad \operatorname{depth}(\tilde{u}_N) \le C N^{\theta}. \tag{5.14}$$

Proof. Define $\hat{b}_j := \max\{b_j, j^{-2/p}\}$. Then $\hat{\boldsymbol{b}} \in \ell^p(\mathbb{N})$ and $\hat{b}_j \ge b_j$ for all $j \in \mathbb{N}$. The definition of $(\boldsymbol{b}, \xi, \delta)$-holomorphy implies that u is also $(\hat{\boldsymbol{b}}, \xi, \delta)$-holomorphic. As in (5.7) we let $\Lambda_\varepsilon = \{\boldsymbol{\nu} : c_{\boldsymbol{\nu}}^{-1} \ge \varepsilon\}$, with $c_{\boldsymbol{\nu}}$ as in (5.6) defined with $\hat{\boldsymbol{b}}$ in place of $\boldsymbol{b}$. We fix $r > 2/p - 1 > 1$ in (5.6) large enough such that with $s := (2/p)/(2(1-p)) > 0$ it holds $\frac{4s}{r} < \theta$.

For $\varepsilon \in (0,1]$ set $u_\varepsilon := \sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} u_{\boldsymbol{\nu}} H_{\boldsymbol{\nu}}$. By Corollary 5.4 with $C_1 := (\sum_{\boldsymbol{\nu}\in\mathcal{F}} c_{\boldsymbol{\nu}} u_{\boldsymbol{\nu}}^2)^{1/2} < \infty$ holds $\|u - u_\varepsilon\|_{L^2(U,\gamma)} \le C_1 \varepsilon^{1/2}$. By Theorem 3.7, there exists a ReLU-NN $\Phi = (\tilde{H}_{\varepsilon,\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_\varepsilon} : \mathbb{R}^{|\operatorname{supp}\Lambda_\varepsilon|} \to \mathbb{R}^{|\Lambda_\varepsilon|}$ such that $\|H_{\boldsymbol{\nu}} - \tilde{H}_{\varepsilon,\boldsymbol{\nu}}\|_{L^2(U,\gamma)} \le \varepsilon$ for each $\boldsymbol{\nu} \in \Lambda_\varepsilon$. Then the NN $\tilde{u}_\varepsilon := \sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} u_{\boldsymbol{\nu}} \tilde{H}_{\varepsilon,\boldsymbol{\nu}}$ satisfies

$$\|u_\varepsilon - \tilde{u}_\varepsilon\|_{L^2(U,\gamma)} \le \sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} |u_{\boldsymbol{\nu}}|\varepsilon = \varepsilon \sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} |u_{\boldsymbol{\nu}}|.$$

By Theorem 5.2 it holds $C_2 := \sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} |u_{\boldsymbol{\nu}}| < \infty$. Hence (using $0 < \varepsilon \le \varepsilon^{1/2} \le 1$)

$$\|u - \tilde{u}_\varepsilon\|_{L^2(U,\gamma)} \le \|u - u_\varepsilon\|_{L^2(U,\gamma)} + \|u_\varepsilon - \tilde{u}_\varepsilon\|_{L^2(U,\gamma)} \le (C_1 + C_2)\varepsilon^{1/2}. \tag{5.15}$$

Next, by Lemma 5.5

$$d(\Lambda_\varepsilon) \le C\log(|\Lambda_\varepsilon|), \qquad m(\Lambda_\varepsilon) \le C|\Lambda_\varepsilon|^{\frac{s}{r}},$$

where $s = (2/p)/(2(1-p)) > 0$. Theorem 3.7 thus implies the bounds (here we use $4s/r < \theta$)

$$\mathrm{size}(\tilde{u}_\varepsilon) \le |\Lambda_\varepsilon| + \mathrm{size}(\Phi) \le C|\Lambda_\varepsilon| m(\Lambda_\varepsilon)^4 d(\Lambda_\varepsilon)^2 \log(\varepsilon^{-1}) \le C|\Lambda_\varepsilon||\Lambda_\varepsilon|^{\frac{4s}{r}} \log(|\Lambda_\varepsilon|)^4 \le C|\Lambda_\varepsilon|^{1+\theta}.$$

Similarly

$$\mathrm{depth}(\tilde{u}_\varepsilon) \le 1 + \mathrm{depth}(\Phi) \le 1 + Cm(\Lambda_\varepsilon)^3 d(\Lambda_\varepsilon)^2 \log(\varepsilon^{-1}) \le 1 + C|\Lambda_\varepsilon|^{\frac{3s}{r}} \log(|\Lambda_\varepsilon|)^3 \le 1 + C|\Lambda_\varepsilon|^{\theta}.$$

Now using (5.8) we have with $C_3 := \|(c_{\boldsymbol{\nu}}^{-1})_{\boldsymbol{\nu}\in\mathcal{F}}\|_{\ell^{p/(2(1-p))}}$

$$|\Lambda_\varepsilon| \le \frac{1}{C_3}\varepsilon^{-\frac{p}{2(1-p)}}.$$

Finally, for $N \in \mathbb{N}$ so large that $\varepsilon_N := (C_3 N)^{-2(1-p)/p}$ is less or equal to 1 we have in particular $|\Lambda_{\varepsilon_N}| \le N$. This choice yields a network $\tilde{u}_N := \tilde{u}_{\varepsilon_N}$ satisfying the size and depth bounds (5.14), as well as the error bound (5.13) due to (5.15) and the definition of ε_N. ■

Remark 5.7. *The network $\tilde{u}_N$ in Theorem 5.6 has size upper bounded by $CN^{1+\theta}$ but takes infinitely many inputs $\boldsymbol{y} \in \mathbb{R}^{\mathbb{N}}$. This is to be understood as follows: $\tilde{u}_N$ is a network with only finitely many inputs $(y_j)_{j\in S_N}$ for some finite set $S_N \subseteq \mathbb{N}$ of cardinality $|S_N| \le CN^{1+\theta}$. All other inputs are ignored. If $(b_j)_{j\in\mathbb{N}}$ in Def. 5.1 is monotonically decreasing, the proof shows that one can choose $S_N = \{j \in \mathbb{N} : j \le \tilde{C}N^{1+\theta}\}$.*

We remark that inspection of the proof actually reveals slightly more precise bounds on $\mathrm{size}(\tilde{u}_\varepsilon)$ and on $\mathrm{depth}(\tilde{u}_\varepsilon)$ than the claim (5.14).

6. DNN Expression rate bounds for response-surfaces of PDEs with GRF input.

We illustrate the expression rate bounds for the infinite-parametric case obtained in Sec. 5.2, by applying them to pushforwards of Gaussian measures under PDE solution maps. For definiteness, we consider standard, linear elliptic second order diffusion in a bounded Lipschitz domain $\mathrm{D} \subseteq \mathbb{R}^d$. For a given source term $f \in H^{-1}(\mathrm{D})$, and for a log-Gaussian diffusion coefficient $a = \exp(g)$ with a GRF g taking values in $L^\infty(\mathrm{D})$, consider the Dirichlet problem

$$\nabla \cdot (a\nabla u) + f = 0 \quad \text{in} \quad \mathrm{D}\,, \quad u|_{\partial\mathrm{D}} = 0\,. \tag{6.1}$$

We assume the log-Gaussian random field $g = \log(a)$ to admit a representation in terms of a Karhunen-Loève expansion

$$\log(a(x,\boldsymbol{y})) = g(x,\boldsymbol{y}) = \sum_{j\ge1} y_j\psi_j(x)\,, \qquad x \in \mathrm{D}, \tag{6.2}$$

where $\boldsymbol{y} = (y_j)_{j\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ with the $y_j \in \mathbb{R}$ iid centered standard Gaussian, and for certain $\psi_j \in L^\infty(\mathrm{D})$.

Remark 6.1. *The functions g and u in (6.1)-(6.2) are well-defined for instance under the following assumptions: Assume that for every $j \in \mathbb{N}$ $\psi_j \in L^\infty(\mathrm{D})$ and there exists $(\lambda_j)_{j\ge1} \in [0,\infty)^{\mathbb{N}}$ such that (i) $(\exp(-\lambda_j^2))_{j\ge1} \in \ell^1(\mathbb{N})$ and (ii) $\sum_{j\in\mathbb{N}} \lambda_j|\psi_j|$ converges in $L^\infty(\mathrm{D})$.*

Then the set $U_0 := \{\boldsymbol{y} \in \mathbb{R}^{\mathbb{N}} : g(\boldsymbol{y}) \in L^\infty(\mathrm{D})\}$ *has full measure, i.e.* $\gamma(U_0) = 1$*, and for every* $k \in \mathbb{N}$ *holds* $\mathbb{E}(\exp(k\|g\|_{L^\infty(\mathrm{D})}) < \infty$ *(see [2, Thm. 2.2]). Furthermore, for all* $f \in H^{-1}(\mathrm{D})$ *and for every* $\boldsymbol{y} \in U_0$*, (6.1) has a unique solution* $u(\boldsymbol{y}) \in H_0^1(\mathrm{D})$*, and* $u \in L^k(U, \gamma; H_0^1(\mathrm{D}))$ *for all finite* $k \in \mathbb{N}$*.*

For an *observable* $G \in H^{-1}(\mathrm{D})$, we consider the *countably-parametric, deterministic PDE response map* $G \circ u$ with u denoting the solution to (6.1) for the log-Gaussian random field a as in (6.2). This map can be formally expressed as

$$G(u(\boldsymbol{y})) = \mathcal{U}\left(\exp\left(\sum_{j\in\mathbb{N}} y_j\psi_j\right)\right), \qquad \boldsymbol{y} \in U, \tag{6.3}$$

for some mapping $\mathcal{U} : L^\infty(\mathrm{D}) \to \mathbb{R}$. More precisely, $\mathcal{U}$ maps a diffusion coefficient $a \in L^\infty(\mathrm{D})$ to the observable G applied to the solution of (6.1). By the complex Lax-Milgram Lemma, the map $\mathcal{U}$ is in particular well-defined on the set $\{a \in L^\infty(\mathrm{D}, \mathbb{C}) : \operatorname{ess\,inf}_{x\in\mathrm{D}} \Re(a) > 0\}$.

An abstract result shown in [5, Lemma 7.10], implies that functions of the type $\boldsymbol{y} \mapsto \mathcal{U}(\exp(\sum_j y_j\psi_j))$ as in (6.3) are $(\boldsymbol{b}, \xi, \delta)$-holomorphic with $b_j = \|\psi_j\|_{L^\infty(\mathrm{D})}$, as long as $\mathcal{U}$ is a holomorphic map between two Banach spaces and it holds $\boldsymbol{b} \in \ell^p(\mathbb{N})$ for some $p \in (0, \frac{2}{3})$. More precisely, [5, Lemma 7.10] shows that (under certain additional assumptions) the functions

$$v^N(y_1, \dots, y_N) := \mathcal{U}\left(\exp\left(\sum_{j=1}^N y_j\psi_j\right)\right)$$

converge towards some $v \in L^2(U, \gamma)$ as $N \to \infty$, and this v is $(\boldsymbol{b}, \xi, \delta)$-holomorphic. In this sense $v(\boldsymbol{y}) = G(u(\boldsymbol{y})) \in L^2(U, \gamma)$ is well-defined. We emphasize that the crucial assumption of $\mathcal{U}$ being holomorphic can be shown for the diffusion problem (6.1), but the result is far from limited to this specific PDE: similar statements can be shown for instance for the Maxwell's equations [11] or for well-posed parabolic PDEs [5] (see [5, Section 7], where well-definedness and $(\boldsymbol{b}, \xi, \delta)$-holomorphy of $U \ni \boldsymbol{y} \mapsto G(u(\boldsymbol{y}))$ is verified in the current setting). ReLU-NN expression rates then follow with Theorem 5.6. We collect these results in the following proposition.

Proposition 6.2. *Let* $f \in H^{-1}(\mathrm{D})$ *and* $g(\boldsymbol{y}) = \log(a(\boldsymbol{y}))$ *be as in (6.2). Suppose that* $(\psi_j)_{j\in\mathbb{N}} \subseteq L^\infty(\mathrm{D})$ *in (6.2) is such that with* $b_j := \|\psi_j\|_{L^\infty(\mathrm{D})}$ *holds* $\boldsymbol{b} \in \ell^p(\mathbb{N})$ *for some* $0 < p < \frac{2}{3}$*. Denote the solution of (6.1) by* $u(\boldsymbol{y}) \in H_0^1(\mathrm{D})$ *whenever* $g(\boldsymbol{y}) \in L^\infty(\mathrm{D})$*.*

For a given observable $G \in H^{-1}(\mathrm{D})$*, the map* $\boldsymbol{y} \mapsto G(u(\boldsymbol{y}))$ *is well-defined as the limit*

$$\lim_{N\to\infty} G(u(y_1, \dots, y_N, 0, 0, \dots)) \in L^2(U, \gamma).$$

Moreover, for every $\theta > 0$ *(arbitrarily small) there exists* $C < \infty$ *such that for every* $N \in \mathbb{N}$ *there exists a* ReLU*-NN* $\tilde{R}_N$ *satisfying*

$$\|G \circ u - \tilde{R}_N\|_{L^2(U,\gamma)} \le CN^{-\frac{1}{p}+1},$$

and

$$\operatorname{size}(\tilde{R}_N) \le CN^{1+\theta}, \qquad \operatorname{depth}(\tilde{R}_N) \le CN^{\theta}.$$

7. Conclusions and extensions. In this paper we discussed the approximation of functions in $L^2(\mathbb{R}^d, \gamma_d)$ with deep ReLU-neural networks. We proved that the Hermite polynomials can be approximated at an exponential convergence rate (in terms of the network size). From this, and classical bounds on the Hermite coefficients, we deduced that ReLU-NNs are capable of approximating analytic functions on $\mathbb{R}^d$ that allow holomorphic extensions onto certain strips in the complex plane at an exponential convergence rate. This result was extended to the infinite dimensional case $d = \infty$, in which case we showed algebraic convergence rates for the class of so-called "$(\boldsymbol{b}, \xi, \delta)$-holomorphic functions". This notion has previously occurred in the literature predominantly for functions with domain $[-1, 1]^{\mathbb{N}}$. We recently extended this definition to functions with domain $\mathbb{R}^{\mathbb{N}}$, and analysed the sparsity properties of this function class in [5]. The present analysis in the case $d = \infty$ strongly draws from these results. Notably, while the investigation of the expressivity of ReLU-NNs on function classes over bounded domains has drawn widespread attention in recent years (see, e.g., the survey [6] and the references there), we provide such results on high-dimensional inputs with unbounded parameter range.

As an application, we discussed the response map of an elliptic PDE, whose input is given in the form of a Karhunen-Loève expansion of a log-Gaussian random field, and proved that this map can be approximated at an algebraic convergence rate with ReLU-DNNs. We emphasize, that similar results will hold also for other well-posed PDE models with log-GRF input. Moreover, as shown in [5], also Bayesian posterior densities for certain PDE based inverse problems belong to the class of $(\boldsymbol{b}, \xi, \delta)$-holomorphic functions. Hence our approximation result may also be applied to such densities. Therefore our analysis could serve as a starting point for developing and analysing neural network driven algorithms for parameter estimation in physical systems.

Appendix A. Proof of Theorem 4.2.

We recall some of the main steps of the proof of [7, Theorem 1], to exhibit the specific bound (4.4), in particular the claimed dependence of the constants on β and f.

As in [7, (3.5)-(3.6)], let $n \in \mathbb{N}$,

$$N := (2n + 1)^{1/2}$$

and define for $z \in S_N := \{z \in \mathbb{C} : \Re[z] \in [-N + 1, N - 1], \ \Im[z] \geq 0\}$

$$\xi(z) := \int_N^z (N^2 - t^2)^{1/2} \mathrm{d}t. \tag{A.1}$$

Due to $\Re[z] \in [-N + 1, N - 1]$, we have $\Re[N^2 - t^2] \geq 0$ for all complex t in the straight line connecting N and z. Throughout what follows, for all $x \in \mathbb{C}$ with $\Re[x] \geq 0$, $x^{1/2}$ is understood as the complex root with nonnegative real part (cp. [7, (3.8)]). Then (A.1) uniquely defines $\xi(z) \in \mathbb{C}$ for all $z \in S_N$.

There hold the following properties:

(i) As recalled in [7, (3.3)-(3.4)], there exist holomorphic functions[4] $(\tilde{h}_n)_{n \in \mathbb{Z}}$ such that for all $n \in \mathbb{N}$

$$h_n(z) = \pi^{-3/4} (2^n n!)^{1/2} \left(\exp\left(\frac{n\pi \mathrm{i}}{2}\right) \tilde{h}_{-n-1}(\mathrm{i}z) + \exp\left(\frac{-n\pi \mathrm{i}}{2}\right) \tilde{h}_{-n-1}(-\mathrm{i}z) \right). \tag{A.2}$$

(ii) As argued in [7, (3.11)], there exists an absolute constant $M > 0$ such that for all $n \in \mathbb{N}$ and for all $z \in S_N$[5] holds

$$|\tilde{h}_{-n-1}(-\mathrm{i}z)| \leq M N^{-n-1} \left| \exp\left(\frac{N^2}{4}\right) \left(1 - \frac{z^2}{N^2}\right)^{-1/4} \exp(\mathrm{i}\xi(z)) \right|. \tag{A.3}$$

(iii) By [7, Lemma 2], with the ellipse

$$E(N, \beta) := \left\{ x + \mathrm{i}y \in \mathbb{C} : \frac{x^2}{N^2} + \frac{y^2}{\beta^2} = 1 \right\}$$

[4] These functions are denoted by h_n in [7].

[5] This bound holds outside of a neighbourhood of the points $\pm N$, which are excluded in our definition of S_N.

it holds for all $z \in E(N,\beta) \cap S_N$ that

$$\Im[\xi(z)] + |x|(\beta^2 - y^2)^{1/2} \geq \beta N - \frac{5}{24}\frac{\beta^3}{N}. \tag{A.4}$$

For fixed $n \in \mathbb{N}$, we bound $|\int_{\mathbb{R}} g(x)h_n(x)\mathrm{d}x|$. Since the integrand is holomorphic in the strip S_τ, the path of integration may be changed within the strip. Using (A.2) we can write

$$\begin{aligned}\int_{\mathbb{R}} g(x)h_n(x)\mathrm{d}x &= \int_{|x|>N-1} g(x)h_n(x)\mathrm{d}x + \pi^{-3/4}(2^n n!)^{1/2}\int_{C_1} g(z)\exp\left(\frac{n\pi \mathrm{i}}{2}\right)\tilde{h}_{-n-1}(\mathrm{i}z)\mathrm{d}z \\ &+ \pi^{-3/4}(2^n n!)^{1/2}\int_{C_2} g(z)\exp\left(\frac{n\pi \mathrm{i}}{2}\right)\tilde{h}_{-n-1}(-\mathrm{i}z)\mathrm{d}z, \end{aligned} \tag{A.5}$$

where the contours C_1 and C_2 are sketched in Fig. A. In the following fix $\beta \in (0,\tau)$.

First, by (2.5) holds $\sup_{x\in\mathbb{R}} |h_n(x)| \leq \pi^{-1/4}$. Using (4.3), we get

$$\left|\int_{|x|>N-1} g(x)h_n(x)\mathrm{d}x\right| \leq 2\pi^{-1/4}\int_{N-1}^{\infty} \exp(-\beta|x|)\mathrm{d}x = \frac{2\pi^{-1/4}\exp(\beta)}{\beta}\exp(-N\beta). \tag{A.6}$$

Next we bound the integral over C_2 in (A.5). By symmetry, the one over C_1 can be treated in the same way. Denote the intersection of $E(N,\beta)$ with $\{z \in \mathbb{C} : \Re[z] = -N+1\}$ in the second quadrant with P, and the intersection of $E(N,\beta)$ with $\{z \in \mathbb{C} : \Re[z] = N-1\}$ in the first quadrant with Q. Denote the vertical line connecting $-N+1$ with with P by v_1, and the one connecting $N-1$ with Q by v_2. We start with the integral over v_1 and compute P. We have $\Re[P] = -N+1$. The imaginary part of P is obtained by solving $\frac{(N-1)^2}{N^2} + \frac{y^2}{\beta^2} = 1$ for y. This yields

$$\Im[P] = \beta\frac{(2N-1)^{1/2}}{N}. \tag{A.7}$$

We note in passing that [7] claims the length of the vertical parts of the path of integration is $O(n^{-3/4})$, but we obtain $O(N^{-1/2}) = O(n^{-1/4})$. This shall be, as we show, sufficient to conclude. By (4.3) for all $z = -(N-1) + \mathrm{i}y \in v_1$

$$\begin{aligned}|g(z)| &\leq B(\beta)\exp\left(-(N-1)(\beta^2-y^2)^{1/2}\right) \leq B(\beta)\exp\left(-(N-1)\beta\left(1-\frac{2N-1}{N^2}\right)^{1/2}\right) \\ &= B(\beta)\exp\left(-(N-1)\beta\frac{N-1}{N}\right) \leq B(\beta)\exp(2\beta)\exp(-N\beta).\end{aligned} \tag{A.8}$$

Next observe that for $z = -N+1+\mathrm{i}y \in v_1$

$$\begin{aligned}\Im[\xi(z)] &= \Im\left[\int_N^{-N+1}(N^2-t^2)^{1/2}\mathrm{d}t + \int_{-N+1}^{-N+1+\mathrm{i}y}(N^2-t^2)^{1/2}\mathrm{d}t\right] \\ &= \Im\left[\mathrm{i}\int_0^y (N^2-(N-1+\mathrm{i}t)^2)^{1/2}\mathrm{d}t\right].\end{aligned}$$

The last term is equal to $\int_0^y \Re[(N^2-(N-1+\mathrm{i}t)^2)^{1/2}]\mathrm{d}t$, and this term is nonnegative by our choice of the branch for the square root and since $\Re[N^2-(N-1+\mathrm{i}t)^2] = N^2-(N-1)^2+t^2 \geq 0$. Hence

$$|\exp(i\xi(z))| \leq 1 \qquad \forall z \in v_1. \tag{A.9}$$

Next, we bound the term $|1-\frac{z^2}{N^2}|^{-1/4}$ occurring in (A.3). Assume $z = x+\mathrm{i}y \in S_N$, i.e. $x \in [-N+1, N-1]$ and $y \geq 0$. Then

$$\begin{aligned}|N^2-z^2|^2 &= |N^2-(x^2+2\mathrm{i}xy-y^2)|^2 = |N^2-x^2+y^2-2\mathrm{i}xy|^2 \\ &= (N^2-x^2+y^2)^2+4x^2y^2 \geq (N^2-x^2)^2,\end{aligned}$$

since the minimum is reached for $y = 0$. Hence, using that $N^2 - x^2 \geq 2N - 1$ if $|x| \leq N - 1$,

$$\left|1 - \frac{z^2}{N^2}\right|^{-1/4} = \left|\frac{N^2}{N^2 - z^2}\right|^{1/4} \leq \left(\frac{N^2}{2N-1}\right)^{1/4} \leq N^{1/4} \qquad \forall z \in S_N \tag{A.10}$$

where we used $N \geq 1$ so that $\frac{N}{2N-1} \leq 1$. Stirling's formula $n! < \mathrm{e}(2\pi n)^{1/2}(\frac{n}{\mathrm{e}})^n$ implies $(\frac{n}{\mathrm{e}})^{-1/2} n^{-1/2} < (n!)^{-1/2}(2\pi)^{1/4}\,\mathrm{e}^{1/2}$. Hence

$$N^{-n-1}\exp\left(\frac{N^2}{4}\right) = (2n+1)^{-\frac{n+1}{2}}\exp\left(\frac{2n+1}{4}\right) < 2^{-n/2}\left(\frac{n}{\mathrm{e}}\right)^{-1/2} n^{-1/2} < (2\pi\,\mathrm{e})^{1/2}(2^n n!)^{-1/2}. \tag{A.11}$$

Combining (A.7)-(A.11) with (A.3) we get

$$\begin{aligned}\left|\int_{v_1} g(x)\tilde{h}_{-n-1}(-\mathrm{i}z)\mathrm{d}z\right| &\leq MB(\beta)\exp(2\beta)\exp(-N\beta)\left(\beta\frac{(2N-1)^{1/2}}{N}\right)N^{1/4}\exp\left(\frac{N^2}{4}\right)N^{-n-1} \\ &\leq 2M(2\pi\,\mathrm{e})^{1/2}B(\beta)\beta\exp(2\beta)(2^n n!)^{-1/2}\exp(-\beta(2n+1)^{1/2}),\end{aligned} \tag{A.12}$$

where we used $\frac{(2N-1)^{1/2}N^{1/4}}{N} \leq 2$ for all $N \geq 1$. The integral over v_2 can be treated in the same way.

Finally, denote by $a = E(N, \beta) \cap S_N$ the arc of the ellipse $E(N, \beta)$ connecting P and Q. By (4.3), (A.3), (A.4), (A.10) and (A.11) we have with $z = x + \mathrm{i}y$ and because the length of the arc a is bounded by $2(\beta + N)$

$$\begin{aligned}\int_a |g(z)\tilde{h}_{-n-1}(-\mathrm{i}z)|\mathrm{d}z &\leq MB(\beta)N^{-n-1}\exp\left(\frac{N^2}{4}\right)\int_a |\exp(-\mathrm{i}\xi(z) - |x|(\beta^2 - y^2))|\left|1 - \frac{z^2}{N^2}\right|^{-1/4}\mathrm{d}z \\ &\leq M(2\pi\,\mathrm{e})^{1/2}(2^n n!)^{-1/2}B(\beta)2(\beta + N)N^{1/4}\exp\left(-\beta N + \frac{5}{24}\frac{\beta^3}{N}\right) \\ &\leq M(2\pi\,\mathrm{e})^{1/2}(2^n n!)^{-1/2}B(\beta)2(1+\beta)\exp\left(\frac{5\beta^3}{24}\right)(1+N)^{5/4}\exp(-\beta N).\end{aligned} \tag{A.13}$$

Using (A.5) and adding up all upper bounds in (A.6), (A.12) and (A.13) we obtain with

$$\tilde{K}(\beta) := \frac{2\pi^{-1/4}\exp(\beta)}{\beta} + 4\left[M\pi^{-3/4}(2\pi\,\mathrm{e})^{1/2}\beta\exp(2\beta)\right] + 2\left[M\pi^{-3/4}(2\pi\,\mathrm{e})^{1/2}2(1+\beta)\exp\left(\frac{5\beta^3}{24}\right)\right]$$

the bound

$$\left|\int_{\mathbb{R}} g(x)h_n(x)\mathrm{d}x\right| \leq B(\beta)K(\beta)(1 + (2n+1)^{1/2})^{5/4}\exp(-\beta(2n+1)^{1/2}).$$

Since this holds for all $\beta \in (0, \tau)$, absorbing[6] $(1 + (2n+1)^{1/2})^{5/4}$ in the exponentially decreasing term, we find that for all $\beta \in (0, \tau)$ exists $K(\beta)$ depending on β (but not on n or g) such that (4.4) holds.

Appendix B. Proof of Theorem 4.5. There holds the following Lemma [1, Lemma A.2][7]:

Lemma B.1. *Let $r \in (0, 1)$ and $s > 0$. Then with $a := r^{\sqrt{2}}$*

$$\sum_{\{k \in \mathbb{N}_0\,:\,k > s\}} r^{\sqrt{2k+1}} \leq \frac{2}{a(1-a)}(\sqrt{s+2} + 4)a^{\sqrt{s}}.$$

[6]Here [7] obtains a term $O(n^{1/4})$ instead of $O(n^{5/8})$.

[7]Lemma A.2 in [1] is stated only for $s \in \mathbb{N}$ and with different constants. The current lemma follows by the same argument after adjusting some constants.

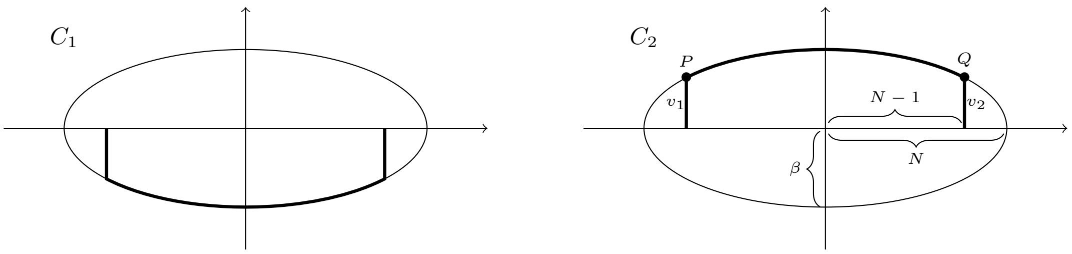

Figure A.1: Paths of integration C_1 and C_2 in (A.5).

Proof of Theorem 4.5. With Θ from Lemma 4.4, denote $F(\boldsymbol{x}) := \Theta(f)(\boldsymbol{x}) = f(2^{1/2}\boldsymbol{x})\,\mathrm{e}^{-\|\boldsymbol{x}\|_2^2}\,\pi^{-d/4}$. Then, since f satisfies Assumption 4.1, for every $\boldsymbol{\beta} \in (0, \boldsymbol{\tau})$ and every $\boldsymbol{x} + \mathrm{i}\boldsymbol{y} \in S_{\boldsymbol{\beta}}$ holds

$$
\begin{aligned}
|F(\boldsymbol{x}+\mathrm{i}\boldsymbol{y})| &= |f(2^{1/2}\boldsymbol{x} + \mathrm{i}2^{1/2}\boldsymbol{y})| \frac{\exp(-\frac{\sum_{j=1}^d x_j^2}{2})}{\pi^{\frac{d}{4}}} \\
&\le B(\boldsymbol{\beta}) \exp\left(\sum_{j=1}^d \left(\frac{(2^{1/2}x_j)^2}{4} - 2^{-1/2}|2^{1/2}x_j|(\beta_j^2 - \frac{1}{2}(2^{1/2}y_j)^2)\right)\right) \frac{\exp(-\frac{\sum_{j=1}^d x_j^2}{2})}{\pi^{\frac{d}{4}}} \\
&\le \frac{B(\boldsymbol{\beta})}{\pi^{\frac{d}{4}}} \exp\left(\sum_{j=1}^d -|x_j|(\beta_j^2 - y_j)^2\right),
\end{aligned}
$$

so that F satisfies the assumption of Corollary 4.3 with the constant $B(\boldsymbol{\beta})/\pi^{\frac{d}{4}}$. The first item thus follows by Corollary 4.3 and Lemma 4.4.

To show the second item, we assume in the following (4.13), which implies by Lemma 4.6 with $\delta(\boldsymbol{\beta})$ as in (4.11)

$$
|\Lambda_\varepsilon| \le 2^d \frac{\log(\varepsilon)^{2d}}{\prod_{j=1}^d \beta_j^2} \;\Rightarrow\; |\Lambda_\varepsilon|^{\frac{1}{2d}} \le 2^{\frac{1}{2}} \frac{|\log(\varepsilon)|}{\delta(\boldsymbol{\beta})}. \tag{B.1}
$$

It suffices to prove the theorem under the constraint (4.13), since $\varepsilon \in (\exp(-\max_{j\in\{1,\dots,d\}} \beta_j), 1)$ only corresponds to finitely many sets Λ_ε.

Denote $\chi_j := (\log(\varepsilon))^2/\beta_j^2$. Then with $K(\boldsymbol{\beta}) := \prod_{j=1}^d K(\beta_j)$, using that $\boldsymbol{\nu} \in \Lambda_\varepsilon$ iff $\nu_j \le \chi_j$ for all j (cp. (4.10)),

$$
\begin{aligned}
\left\| f - \sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} \langle f, H_{\boldsymbol{\nu}}\rangle H_{\boldsymbol{\nu}} \right\|^2_{L^2(\mathbb{R}^d,\gamma_d)} &= \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^d\setminus\Lambda_\varepsilon} |\langle f, H_{\boldsymbol{\nu}}\rangle|^2 \\
&\le K(\boldsymbol{\beta})^2 B(\boldsymbol{\beta})^2 \sum_{j=1}^d \sum_{n>\chi_j} \exp(-2\beta_j\sqrt{2n+1}) \sum_{\{(\nu_i)_{i\neq j}\,:\,\nu_i\in\mathbb{N}_0\}} \prod_{i\neq j} \exp(-2\beta_i\sqrt{2\nu_i+1}).
\end{aligned}
$$

By Lemma B.1, with $a_j := \exp(-2^{3/2}\beta_j)$ and $C_j := \frac{2}{a_j(1-a_j)}$,

$$
\sum_{n>\chi_j} \exp(-2\beta_j\sqrt{2n+1}) \le C_j(\sqrt{\chi_j+2}+4)\exp(-2^{3/2}\beta_j\sqrt{\chi_j}) \le \tilde{C}_j \exp(-2\beta_j\sqrt{\chi_j}) = \tilde{C}_j\varepsilon^2,
$$

for some $\tilde{C}_j$ depending on β_j. Furthermore (e.g. by Lemma B.1) we have $D(\beta_j) := \sum_{n\in\mathbb{N}_0}\exp(-2\beta_j\sqrt{2n+1}) < \infty$, and thus

$$
\sum_{\{(\nu_i)_{i\neq j}\,:\,\nu_i\in\mathbb{N}_0\}} \prod_{i\neq j} \exp(-2\beta_i\sqrt{2\nu_i+1}) = \prod_{i\neq j}\sum_{n\in\mathbb{N}_0} \exp(-2\beta_i\sqrt{2n+1}) \le (\max_{j\le d} D(\beta_j))^{d-1} =: C_{\max}.
$$

Hence

$$\text{(B.2)} \qquad \left\| f - \sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} \langle f, H_{\boldsymbol{\nu}}\rangle H_{\boldsymbol{\nu}} \right\|^2_{L^2(\mathbb{R}^d,\gamma_d)} \le K(\boldsymbol{\beta})^2 B(\boldsymbol{\beta})^2 C_{\max} \sum_{j=1}^d \tilde{C}_j \varepsilon^2 = C\varepsilon^2,$$

with the $\boldsymbol{\beta}$ and d-dependent constant $C := K(\boldsymbol{\beta})^2 B(\boldsymbol{\beta})^2 C_{\max} \sum_{j=1}^d \tilde{C}_j$.

By (B.1)

$$\varepsilon \le \exp\left(-2^{-\frac{1}{2}}\delta(\boldsymbol{\beta})|\Lambda_\varepsilon|^{\frac{1}{2d}}\right)$$

so that together with (B.2)

$$\left\| f - \sum_{\boldsymbol{\nu}\in\Lambda_\varepsilon} \langle f, H_{\boldsymbol{\nu}}\rangle H_{\boldsymbol{\nu}} \right\|_{L^2(\mathbb{R}^d,\gamma_d)} \le C\exp(-2^{-\frac{1}{2}}\delta(\boldsymbol{\beta})|\Lambda_\varepsilon|^{\frac{1}{2d}}).$$

■

REFERENCES

[1] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 45(3):1005–1034, 2007.

[2] M. Bachmayr, A. Cohen, R. DeVore, and G. Migliorati. Sparse polynomial approximation of parametric elliptic PDEs. Part II: Lognormal coefficients. *ESAIM Math. Model. Numer. Anal.*, 51(1):341–363, 2017.

[3] V. I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.

[4] R. H. Cameron and W. T. Martin. The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. *Ann. of Math. (2)*, 48:385–392, 1947.

[5] D. Dung, V. K. Nguyen, C. Schwab, and J. Zech. Analyticity and sparsity in forward and inverse UQ for PDEs with gaussian random field inputs. Technical report, Seminar for Applied Mathematics, ETH Zürich, 2021. in preparation.

[6] D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei. Deep neural network approximation theory. *IEEE Trans. Inform. Theory*, 67(5):2581–2623, 2021.

[7] E. Hille. Contributions to the theory of Hermitian series. II. The representation problem. *Trans. Amer. Math. Soc.*, 47:80–94, 1940.

[8] V. H. Hoang and C. Schwab. N-term Wiener chaos approximation rate for elliptic PDEs with lognormal Gaussian random inputs. *Math. Models Methods Appl. Sci.*, 24(4):797–826, 2014.

[9] J. Indritz. An inequality for Hermite polynomials. *Proc. Amer. Math. Soc.*, 12:981–983, 1961.

[10] S. Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[11] C. Jerez-Hanckes, C. Schwab, and J. Zech. Electromagnetic wave scattering by random surfaces: Shape holomorphy. *Mathematical Models and Methods in Applied Sciences*, 27(12):2229–2259, 2017.

[12] A. Lang and C. Schwab. Isotropic gaussian random fields on the sphere: regularity, fast simulation, and stochastic partial differential equations. *Ann. Appl. Probability*, 25(6):3047–3094, 2015.

[13] M. A. Lifshits. *Gaussian random functions*, volume 322 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 1995.

[14] H. Luschgy and G. Pagès. Expansions for Gaussian processes and Parseval frames. *Electron. J. Probab.*, 14:no. 42, 1198–1221, 2009.

[15] L. Mattner. Complex differentiation under the integral. *Nieuw Arch. Wiskd. (5)*, 2(1):32–35, 2001.

[16] J. A. A. Opschoor, Christoph Schwab, and J. Zech. Exponential ReLU DNN expression of holomorphic maps in high dimension. *Constructive Approximation*, 2021.

[17] J. A. A. Opschoor, P. C. Petersen, and C. Schwab. Deep ReLU networks and high-order finite element methods. *Anal. Appl. (Singap.)*, 18(5):715–770, 2020.

[18] P. Petersen and F. Voigtlaender. Optimal approximation of piecewise smooth functions using deep relu neural networks. *Neural Networks*, 108:296–330, 2018.

[19] Ch. Schwab and J. Zech. Deep learning in high dimension: Neural network expression rates for generalized polynomial chaos expansions in UQ. *Anal. Appl. (Singap.)*, 17(1):19–55, 2019.

[20] A. M. Stuart and A. L. Teckentrup. Posterior consistency for Gaussian process approximations of Bayesian posterior distributions. *Math. Comp.*, 87(310):721–753, 2018.

[21] G. Szegő. *Orthogonal polynomials*. American Mathematical Society, Providence, R.I., fourth edition, 1975. American Mathematical Society, Colloquium Publications, Vol. XXIII.
[22] N. Wiener. The Homogeneous Chaos. *Amer. J. Math.*, 60(4):897–936, 1938.
[23] D. Yarotsky. Error bounds for approximations with deep ReLU networks. *Neural Networks*, 94:103–114, 2017.
[24] D. Yarotsky. Optimal approximation of continuous functions by very deep ReLU networks. In S. Bubeck, V. Perchet, and P. Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 639–649. PMLR, 06–09 Jul 2018.
[25] J. Zech and C. Schwab. Convergence rates of high dimensional Smolyak quadrature. *ESAIM Math. Model. Numer. Anal.*, 54(4):1259–1307, 2020.