Journal of Complexity 77 (2023) 101746

Contents lists available at ScienceDirect

Journal of Complexity

journal homepage: www.elsevier.com/locate/jco

Lower bounds for artificial neural network approximations: A proof that shallow neural networks fail to overcome the curse of dimensionality ☆

Philipp Grohs [a,b], Shokhrukh Ibragimov [c], Arnulf Jentzen [d,c], Sarah Koppensteiner [e]

[a] Faculty of Mathematics and Research Platform Data Science, University of Vienna, Austria
[b] Johann Radon Institute of Computational and Applied Mathematics, Austrian Academy of Sciences, Austria
[c] Applied Mathematics: Institute for Analysis and Numerics, University of Münster, Germany
[d] School of Data Science and Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen, China
[e] Faculty of Mathematics, University of Vienna, Austria

ARTICLE INFO

Article history:
Received 25 May 2021
Received in revised form 3 August 2022
Accepted 21 February 2023
Available online 1 March 2023

Keywords:
Artificial neural networks
Artificial neural network approximations
Curse of dimensionality
Overcoming the curse of dimensionality
Lower bounds

ABSTRACT

Artificial neural networks (ANNs) have become a very powerful tool in the approximation of high-dimensional functions. Especially, deep ANNs, consisting of a large number of hidden layers, have been very successfully used in a series of practical relevant computational problems involving high-dimensional input data ranging from classification tasks in supervised learning to optimal decision problems in reinforcement learning. There are also a number of mathematical results in the scientific literature which study the approximation capacities of ANNs in the context of high-dimensional target functions. In particular, there are a series of mathematical results in the scientific literature which show that sufficiently deep ANNs have the capacity to overcome the curse of dimensionality in the approximation of certain target function classes in the sense that the number of parameters of the approximating ANNs grows at most polynomially in the dimension $d \in \mathbb{N}$ of the target functions under considerations. In the proofs of several of such high-dimensional approximation results it is

☆ Communicated by E. Novak.

E-mail addresses: philipp.grohs@univie.ac.at, philipp.grohs@oeaw.ac.at (P. Grohs), sibragim@uni-muenster.de (S. Ibragimov), ajentzen@cuhk.edu.cn, ajentzen@uni-muenster.de (A. Jentzen), sarah.koppensteiner@univie.ac.at (S. Koppensteiner).

https://doi.org/10.1016/j.jco.2023.101746

crucial that the involved ANNs are sufficiently deep and consist a sufficiently large number of hidden layers which grows in the dimension of the considered target functions. It is the topic of this work to look a bit more detailed to the deepness of the involved ANNs in the approximation of high-dimensional target functions. In particular, the main result of this work proves that there exists a concretely specified sequence of functions which can be approximated without the curse of dimensionality by sufficiently deep ANNs but which cannot be approximated without the curse of dimensionality if the involved ANNs are shallow or not deep enough.

Contents

1. Introduction

Artificial neural networks (ANNs) have become a very powerful tool in the approximation of high-dimensional functions. Especially, deep ANNs, consisting of a large number of hidden layers, have been very successfully used in a series of practical relevant computational problems involving high-dimensional input data ranging from classification tasks in supervised learning to optimal decision problems in reinforcement learning.

There are also a large number of mathematical results in the scientific literature which study the approximation capacities of ANNs; see, e.g., Cybenko [13], Funahashi [20], Hornik et al. [31,32], Leshno et al. [46], Bianchini & Scarselli [10], Mhaskar et al. [49,50], Guliyev & Ismailov [29], Elbrächter et al. [18], and the references mentioned therein. Moreover, in the recent years a series of articles have appeared in the scientific literature which study the approximation capacities of ANNs in the context of high-dimensional target functions. In particular, the results in such articles show that deep ANNs have the capacity to overcome *the curse of dimensionality*[1] (cf., e.g., Bellman [7] and Novak & Woźniakowski [51, Chapter 1]) in the approximation of certain target function classes in the sense that the number of parameters of the approximating ANNs grows at most polynomially in the dimension $d \in \mathbb{N}$ of the target functions under considerations. For example, we refer to Elbrächter et al. [17], Jentzen et al. [35], Gonon et al. [22,23], Grohs et al. [24,25,27], Kutyniok et al. [45], Reisinger & Zhang [55], Beneventano et al. [8], Berner et al. [9], Hornung et al. [33], Hutzenthaler et al. [34], and the overview articles Beck et al. [5] and E et al. [16] for such high-dimensional ANN approximation results in the numerical approximation of solutions of PDEs and we refer to Barron [2–4], Jones [36], Girosi & Anzellotti [21], Donahue et al. [15], Gurvits & Koiran [30], Kůrková et al. [41–44], Kainen et al. [37,38], Klusowski & Barron [40], Li et al. [47], and Cheridito et al. [12] for such high-dimensional ANN approximation results in the numerical approximation of certain specific target function classes independent of solutions of PDEs (cf., e.g., also Maiorov & Pinkus [48], Pinkus [54], Guliyev & Ismailov [28], Petersen & Voigtlaender [53], and Bölcskei et al. [11] for related results). In the proofs of several of the above named high-dimensional approximation results it is crucial that the involved ANNs are sufficiently deep and consist a sufficiently large number of hidden layers which grows in the dimension of the considered target functions.

It is the key topic of this work to look a bit more detailed to the deepness of the involved ANNs in the approximation of high-dimensional target functions. More specifically, Theorem 6.1 in Section 6 below, which is the main result of this work, proves that there exists a concretely specified sequence of high-dimensional functions which can be approximated without the curse of dimensionality by sufficiently deep ANNs but which cannot be approximated without the curse of dimensionality if the involved ANNs are shallow or not deep enough. In the scientific literature related ANN approximation results can also be found in Daniely [14], Eldan & Shamir [19], and Safran & Shamir [57]. One of the differences between the results in the above named references and the results in this work is, roughly speaking, that the considered target functions in the above named references can be approximated by ANNs with two hidden layers without the curse of dimensionality but not with ANNs with one hidden layer while in this work the considered target functions can only be approximated without the curse of dimensionality if the number of the hidden layers of the approximating ANN grows like the dimensions of the target functions.

To illustrate the findings of this work in more detail, we now present in the following result, Theorem 1.1 below, a special case of Theorem 6.1. Below Theorem 1.1 we also add some explanatory comments regarding the mathematical objects appearing in Theorem 1.1 and regarding the statement of Theorem 1.1.

[1] In the literature one usually says that ANN approximations suffer under the curse of dimensionality if the number of parameters of the approximating ANN grows at least exponentially in the dimension of the target functions under considerations (cf., e.g., Bellman [7] and Novak & Woźniakowski [51, Chapter 1]) and one usually speaks of polynomial tractability if the number of ANN parameters grows at most polynomially in the dimension of the target functions under considerations and the inverse of the prescribed approximation accuracy (cf., e.g., Novak & Woźniakowski [51, Chapter 1] and Novak & Woźniakowski [52, Section 9.1]).

Theorem 1.1. *Let $\varphi\colon (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to \mathbb{R}$ and $\mathfrak{R}\colon (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to (\cup_{d\in\mathbb{N}}\mathbb{R}^d)$ satisfy for all $d \in \mathbb{N}$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that $\varphi(x) = (2\pi)^{-d/2}\exp(-\frac{1}{2}(\sum_{j=1}^d |x_j|^2))$ and $\mathfrak{R}(x) = (\max\{x_1, 0\}, \dots, \max\{x_d, 0\})$, let $\mathbf{N} = \cup_{L\in\mathbb{N}} \cup_{l_0,l_1,\dots,l_L\in\mathbb{N}} (\times_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, and let $\mathcal{R}\colon \mathbf{N} \to (\cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$, $\mathcal{H}\colon \mathbf{N} \to \mathbb{N}_0$, $\mathcal{P}\colon \mathbf{N} \to \mathbb{N}$, and $|||\cdot|||\colon \mathbf{N} \to \mathbb{R}$ satisfy for all $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $v_0 \in \mathbb{R}^{l_0}, v_1 \in \mathbb{R}^{l_1}, \dots, v_L \in \mathbb{R}^{l_L}$, $\Phi = ((W_1, B_1), \dots, (W_L, B_L)) = (((W_{1,i,j})_{(i,j)\in\{1,\dots,l_1\}\times\{1,\dots,l_0\}}, (B_{1,i})_{i\in\{1,\dots,l_1\}}), \dots, ((W_{L,i,j})_{(i,j)\in\{1,\dots,l_L\}\times\{1,\dots,l_{L-1}\}}, (B_{L,i})_{i\in\{1,\dots,l_L\}})) \in (\times_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$ with $\forall k \in \{1, 2, \dots, L\}\colon v_k = \mathfrak{R}(W_k v_{k-1} + B_k)$ that $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$, $(\mathcal{R}(\Phi))(v_0) = W_L v_{L-1} + B_L$, $\mathcal{H}(\Phi) = L - 1$, $\mathcal{P}(\Phi) = \sum_{k=1}^L l_k(l_{k-1} + 1)$, and $|||\Phi||| = \max_{1\le n\le L}\max_{1\le i\le l_n}\max_{1\le j\le l_{n-1}}\max\{|W_{n,i,j}|, |B_{n,i}|\}$. Then there exist continuously differentiable $\mathfrak{f}_d\colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, such that for all $\delta \in (0, 1]$, $\varepsilon \in (0, 1/2]$ there exists $\mathfrak{C} \in (0, \infty)$ such that*

(i) it holds for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $d \in \mathbb{N}$ that

$$\min\left\{p \in \mathbb{N}\colon \begin{bmatrix} \exists\, \Phi \in \mathbf{N}\colon\ p = \mathcal{P}(\Phi),\ |||\Phi||| \le \mathfrak{c}d^{\mathfrak{c}}, \\ d \le \mathcal{H}(\Phi) \le \mathfrak{c}d,\ \mathcal{R}(\Phi) \in C(\mathbb{R}^d, \mathbb{R}), \\ [\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{f}_d(x)|^2 \varphi(x)\,dx]^{1/2} \le \varepsilon \end{bmatrix}\right\} \le \mathfrak{c}d^3 \tag{1.1}$$

and

(ii) it holds for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $d \in \mathbb{N}$ that

$$\min\left\{p \in \mathbb{N}\colon \begin{bmatrix} \exists\, \Phi \in \mathbf{N}\colon\ p = \mathcal{P}(\Phi),\ |||\Phi||| \le \mathfrak{c}d^{\mathfrak{c}}, \\ \mathcal{H}(\Phi) \le \mathfrak{c}d^{1-\delta},\ \mathcal{R}(\Phi) \in C(\mathbb{R}^d, \mathbb{R}), \\ [\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{f}_d(x)|^2 \varphi(x)\,dx]^{1/2} \le \varepsilon \end{bmatrix}\right\} \ge (1 + \mathfrak{c}^{-3})^{(d^\delta)}. \tag{1.2}$$

Theorem 1.1 is an immediate consequence of Theorem 6.1, respectively Corollary 6.2, the main results of this paper. They combine the results of Corollary 5.12 and Corollary 4.9, which establish items (i) and (ii), respectively.

In the following we provide some explanatory comments regarding the statement of and mathematical objects appearing in Theorem 1.1. In Theorem 1.1 we measure the error between the target function and the realization of the approximating ANN in the L^2-sense on the whole $\mathbb{R}^d$, $d \in \mathbb{N}$, with respect to standard normal distribution. In particular, we observe that the function $\varphi\colon (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to \mathbb{R}$ in Theorem 1.1 appears in the L^2-errors in items (i) and (ii) in Theorem 1.1 and describes the densities of the standard normal distribution. More formally, note that for all $d \in \mathbb{N}$ it holds that the function $\mathbb{R}^d \ni x \mapsto \varphi(x) = (2\pi)^{-d/2}\exp(-\frac{1}{2}(\sum_{j=1}^d |x_j|^2)) \in \mathbb{R}$ is nothing else but the density of the d-dimensional standard normal distribution.

Theorem 1.1 is an approximation result for ANNs with the rectifier function as the activation function and the function $\mathfrak{R}\colon (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to (\cup_{d\in\mathbb{N}}\mathbb{R}^d)$ in Theorem 1.1 describes multidimensional versions of the rectifier function. More specifically, observe that for all $d \in \mathbb{N}$ it holds that the function $\mathbb{R}^d \ni x \mapsto \mathfrak{R}(x) = (\max\{x_1, 0\}, \dots, \max\{x_d, 0\}) \in \mathbb{R}^d$ is the d-dimensional version of the rectifier activation function $\mathbb{R} \ni x \mapsto \max\{x, 0\} \in \mathbb{R}$.

The set $\mathbf{N} = \cup_{L\in\mathbb{N}} \cup_{l_0,l_1,\dots,l_L\in\mathbb{N}} (\times_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$ in Theorem 1.1 represents the set of all ANNs and the function $\mathcal{R}\colon \mathbf{N} \to (\cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$ in Theorem 1.1 assigns to each ANN in $\mathbf{N}$ its realization function. More formally, note that for every ANN $\Phi \in \mathbf{N}$ it holds that the function $\mathcal{R}(\Phi) \in (\cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$ is the realization function associated to the ANN Φ.

The function $\mathcal{H}\colon \mathbf{N} \to \mathbb{N}_0$ in Theorem 1.1 describes the number of hidden layers of the considered ANN, the function $\mathcal{P}\colon \mathbf{N} \to \mathbb{N}$ in Theorem 1.1 counts the number of parameters (the number of weights and biases) used to describe the considered ANN, and the function $|||\cdot|||\colon \mathbf{N} \to \mathbb{R}$ in Theorem 1.1 specifies the size of the absolute values of the parameters of the considered ANN. More specifically, observe that for every ANN $\Phi \in \mathbf{N}$ it holds that $\mathcal{H}(\Phi)$ is the number of hidden layers of the ANN Φ, that $\mathcal{P}(\Phi)$ is the number of real parameters used to describe the ANN Φ, and that $|||\Phi|||$ is the maximum of the absolute values of the real parameters used to describe the ANN Φ.

It is well-known that shallow ANNs with the rectifier function as activation function can approximate any continuous function uniformly on compacta, see, e.g., Pinkus [54]. But this is a qualitative result and gives no information on the required number of ANN parameters for a given approximation error. Roughly speaking, Theorem 1.1 asserts that there exists a sequence of continuously differentiable target functions $\mathfrak{f}_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, such that for every arbitrarily small prescribed approximation accuracy $\varepsilon \in (0, 1/2]$ it holds that the class of all sufficiently deep ANNs can approximate the target functions $\mathfrak{f}_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, without the curse of dimensionality (with the number of ANN parameters growing at most cubically in the dimension $d \in \mathbb{N}$; see (1.1) in item (i) in Theorem 1.1) and that the class of all shallow ANNs (with the absolute values of the parameters of the considered ANN growing at most polynomially in the dimension $d \in \mathbb{N}$) can only approximate the target functions $\mathfrak{f}_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, with the curse of dimensionality (with the number of ANN parameters growing at least exponentially in the dimension $d \in \mathbb{N}$; see (1.2) in item (ii) in Theorem 1.1). In that sense Theorem 1.1 shows for a specific class of target functions that sufficiently deep ANNs can overcome the curse of dimensionality but shallow ANNs fail to do so. Let us also point out that for this specific class of target functions Theorem 1.1 does not only show that shallow ANNs fail to overcome the curse of dimensionality in the sense of (1.2) but also that all deep ANNs with the number of hidden layers growing less or equal than $\mathfrak{c}d^{1-\delta}$, with an arbitrarily small $\delta \in (0, 1)$ and a sufficiently large $\mathfrak{c} \in (0, \infty)$, fail to overcome the curse of dimensionality; see item (ii) in Theorem 1.1 for details.

The target functions $\mathfrak{f}_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, mentioned in Theorem 1.1 will be constructed explicitly in Corollary 6.2 and can be chosen as $\mathfrak{f}_d(x) = f_d(x)[\int_{\mathbb{R}^d} |f_d(y)|^2 \varphi_d(y)\, dy]^{-1/2}$, where $f_d(x) = \sum_{j=1}^{d} [\max\{|x_j| - \sqrt{2d}, 0\}]^2$. The proof of Theorem 1.1 exploits the fact that the values of the functions $\mathfrak{f}_d$ grow at least exponentially in $d \in \mathbb{N}$ (cf. Lemma 5.9 below for details) as well as the fact that the efficient and accurate approximation of such functions by non-sufficiently deep ANNs requires the weights of the approximating ANNs to grow at least exponentially in $d \in \mathbb{N}$. A weakness of the statement of Theorem 1.1 is that there is no growth restriction on the values of the target functions $\mathfrak{f}_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, appearing in Theorem 1.1. In particular, beyond Theorem 1.1, an interesting topic of future research is to prove or disprove a statement of the type of Theorem 1.1 but for target functions which do not grow at least exponentially but which may only grow at most polynomially in the dimension $d \in \mathbb{N}$ instead.

Let us give some intuition why ANNs with the number of hidden layers growing at most like $d^{1-\delta}$, with an arbitrarily small $\delta \in (0, 1)$, can not approximate the target functions $\mathfrak{f}_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, without the curse of dimensionality but ANNs with the number of hidden layers growing at least like d can approximate the target functions $\mathfrak{f}_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, without the curse of dimensionality. Very roughly speaking, only through the deepness of ANNs an exponential growth can be created. More specifically, if the parameters of the ANNs are assumed to be at most polynomially growing as in item (ii) (and item (i)) in Theorem 1.1, then an exponential growth in the approximating realization functions of the ANNs can only be generated through multiple compositions coming from deep ANNs (cf. Lemma 5.10). The absolute values of our target functions grow like e^d (cf. Lemma 5.9) and, as a consequence of this, only ANNs whose number of compositions (number of layers) growing of order d can approximate functions growing like e^d without the curse of dimensionality but those ANNs having polynomially strictly less than d compositions can not approximate functions growing like e^d without the curse of dimensionality.

The statement of Theorem 1.1 holds subject to the constraint that absolute values of the parameters of the considered ANNs grow at most polynomially in the dimension $d \in \mathbb{N}$. The question whether Theorem 1.1 still holds without this constraint is open and might require different tools than used in the present paper.

The remainder of this article is organized as follows: In Section 2 we briefly recall a few general concepts and results for ANNs. We proceed to collect some technical results on weighted and unweighted Gaussian tails in Section 3. The main work for proving item (ii) of Theorem 1.1 is done in Section 4 culminating in Corollary 4.9. Section 5 essentially establishes item (i) of Theorem 1.1 with Corollary 5.12. In Section 6 we combine those two results and obtain Theorem 6.1, respectively Corollary 6.2, the main ANN approximation results of this work with Theorem 1.1 as an immediate consequence.

2. Basics on artificial neural networks (ANNs)

In this section we briefly recall a few general well-known concepts and well-known results for artificial neural networks (ANNs), which can mostly be found, e.g., in [26, Section 2] and [27, Section 3]. For the vectorized description of ANNs, i.e. Definition 2.22, we refer, e.g., to [6, Definition 2.11].

2.1. Structured description of ANNs

Definition 2.1. We denote by $\mathfrak{R}\colon (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to (\cup_{d\in\mathbb{N}}\mathbb{R}^d)$ the function which satisfies for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $\mathfrak{R}(x) = (\max\{x_1, 0\}, \max\{x_2, 0\}, \ldots, \max\{x_d, 0\})$.

Definition 2.2. We denote by $\mathbf{N}$ the set given by

$$\mathbf{N} = \bigcup_{L\in\mathbb{N}} \bigcup_{l_0,l_1,\ldots,l_L\in\mathbb{N}} \left(\times_{k=1}^{L} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\right) \tag{2.1}$$

and we denote by $\mathcal{R}\colon \mathbf{N} \to (\cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$, $\mathcal{P}\colon \mathbf{N} \to \mathbb{N}$, $\mathcal{L}\colon \mathbf{N} \to \mathbb{N}$, $\mathcal{I}\colon \mathbf{N} \to \mathbb{N}$, $\mathcal{O}\colon \mathbf{N} \to \mathbb{N}$, $\mathcal{H}\colon \mathbf{N} \to \mathbb{N}_0$, $\mathcal{D}\colon \mathbf{N} \to (\cup_{L=2}^{\infty}\mathbb{N}^L)$, and $\mathbb{D}_n\colon \mathbf{N} \to \mathbb{N}_0$, $n \in \mathbb{N}_0$, the functions which satisfy for all $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), (W_2, B_2), \ldots, (W_L, B_L)) \in (\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $v_0 \in \mathbb{R}^{l_0}$, $v_1 \in \mathbb{R}^{l_1}, \ldots, v_L \in \mathbb{R}^{l_L}$, $n \in \mathbb{N}_0$ with $\forall k \in \{1, 2, \ldots, L\}\colon v_k = \mathfrak{R}(W_k v_{k-1} + B_k)$ that $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$, $(\mathcal{R}(\Phi))(v_0) = W_L v_{L-1} + B_L$, $\mathcal{P}(\Phi) = \sum_{k=1}^{L} l_k(l_{k-1} + 1)$, $\mathcal{L}(\Phi) = L$, $\mathcal{I}(\Phi) = l_0$, $\mathcal{O}(\Phi) = l_L$, $\mathcal{H}(\Phi) = L - 1$, $\mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_L)$, and

$$\mathbb{D}_n(\Phi) = \begin{cases} l_n & : n \le L \\ 0 & : n > L \end{cases} \tag{2.2}$$

(cf. Definition 2.1).

Definition 2.3 *(Neural network).* We say that Φ is a neural network if and only if it holds that $\Phi \in \mathbf{N}$ (cf. Definition 2.2).

2.2. Compositions of ANNs

Definition 2.4 *(Compositions of ANNs).* We denote by $(\cdot) \bullet (\cdot)\colon \{(\Phi_1, \Phi_2) \in \mathbf{N} \times \mathbf{N}\colon \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)\} \to \mathbf{N}$ the function which satisfies for all $L, \mathfrak{L} \in \mathbb{N}$, $l_0, l_1, \ldots, l_L, \mathfrak{l}_0, \mathfrak{l}_1, \ldots, \mathfrak{l}_{\mathfrak{L}} \in \mathbb{N}$, $\Phi_1 = ((W_1, B_1), (W_2, B_2), \ldots, (W_L, B_L)) \in (\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $\Phi_2 = ((\mathfrak{W}_1, \mathfrak{B}_1), (\mathfrak{W}_2, \mathfrak{B}_2), \ldots, (\mathfrak{W}_{\mathfrak{L}}, \mathfrak{B}_{\mathfrak{L}})) \in (\times_{k=1}^{\mathfrak{L}}(\mathbb{R}^{\mathfrak{l}_k\times \mathfrak{l}_{k-1}} \times \mathbb{R}^{\mathfrak{l}_k}))$ with $l_0 = \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2) = \mathfrak{l}_{\mathfrak{L}}$ that

$$\Phi_1 \bullet \Phi_2 = \begin{cases} \begin{array}{r}\big((\mathfrak{W}_1, \mathfrak{B}_1), (\mathfrak{W}_2, \mathfrak{B}_2), \ldots, (\mathfrak{W}_{\mathfrak{L}-1}, \mathfrak{B}_{\mathfrak{L}-1}), (W_1\mathfrak{W}_{\mathfrak{L}}, W_1\mathfrak{B}_{\mathfrak{L}} + B_1),\\ (W_2, B_2), (W_3, B_3), \ldots, (W_L, B_L)\big)\end{array} & : L > 1 < \mathfrak{L} \\ \big((W_1\mathfrak{W}_1, W_1\mathfrak{B}_1 + B_1), (W_2, B_2), (W_3, B_3), \ldots, (W_L, B_L)\big) & : L > 1 = \mathfrak{L} \\ \big((\mathfrak{W}_1, \mathfrak{B}_1), (\mathfrak{W}_2, \mathfrak{B}_2), \ldots, (\mathfrak{W}_{\mathfrak{L}-1}, \mathfrak{B}_{\mathfrak{L}-1}), (W_1\mathfrak{W}_{\mathfrak{L}}, W_1\mathfrak{B}_{\mathfrak{L}} + B_1)\big) & : L = 1 < \mathfrak{L} \\ \big((W_1\mathfrak{W}_1, W_1\mathfrak{B}_1 + B_1)\big) & : L = 1 = \mathfrak{L} \end{cases} \tag{2.3}$$

(cf. Definition 2.2).

Proposition 2.5. *Let $\Phi_1, \Phi_2 \in \mathbf{N}$ satisfy $\mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)$ (cf. Definition 2.2). Then*

(i) it holds that

$$\mathcal{D}(\Phi_1 \bullet \Phi_2) = (\mathbb{D}_0(\Phi_2), \mathbb{D}_1(\Phi_2), \ldots, \mathbb{D}_{\mathcal{H}(\Phi_2)}(\Phi_2), \mathbb{D}_1(\Phi_1), \mathbb{D}_2(\Phi_1), \ldots, \mathbb{D}_{\mathcal{L}(\Phi_1)}(\Phi_1)), \tag{2.4}$$

(ii) it holds that $\mathcal{H}(\Phi_1 \bullet \Phi_2) = \mathcal{H}(\Phi_1) + \mathcal{H}(\Phi_2)$,
(iii) it holds that $\mathcal{R}(\Phi_1 \bullet \Phi_2) \in C(\mathbb{R}^{\mathcal{I}(\Phi_2)}, \mathbb{R}^{\mathcal{O}(\Phi_1)})$, and
(iv) it holds that $\mathcal{R}(\Phi_1 \bullet \Phi_2) = [\mathcal{R}(\Phi_1)] \circ [\mathcal{R}(\Phi_2)]$

(cf. Definition 2.4).

A proof of Proposition 2.5 can be found, e.g., in [26, Proposition 2.6].

Lemma 2.6. *Let $\Phi_1, \Phi_2, \Phi_3 \in \mathbf{N}$ satisfy $\mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)$ and $\mathcal{I}(\Phi_2) = \mathcal{O}(\Phi_3)$ (cf. Definition 2.2). Then $(\Phi_1 \bullet \Phi_2) \bullet \Phi_3 = \Phi_1 \bullet (\Phi_2 \bullet \Phi_3)$ (cf. Definition 2.4).*

A proof of Lemma 2.6 can be found, e.g., in [26, Lemma 2.8].

2.3. Powers of ANNs

Definition 2.7. Let $n \in \mathbb{N}$. Then we denote by $\mathrm{I}_n \in \mathbb{R}^{n \times n}$ the identity matrix in $\mathbb{R}^{n \times n}$.

Definition 2.8. We denote by $(\cdot)^{\bullet n} : \{\Phi \in \mathbf{N} : \mathcal{I}(\Phi) = \mathcal{O}(\Phi)\} \to \mathbf{N}$, $n \in \mathbb{N}_0$, the functions which satisfy for all $n \in \mathbb{N}_0$, $\Phi \in \mathbf{N}$ with $\mathcal{I}(\Phi) = \mathcal{O}(\Phi)$ that

$$\Phi^{\bullet n} = \begin{cases} \big(\mathrm{I}_{\mathcal{O}(\Phi)}, (0, 0, \dots, 0)\big) \in \mathbb{R}^{\mathcal{O}(\Phi) \times \mathcal{O}(\Phi)} \times \mathbb{R}^{\mathcal{O}(\Phi)} & : n = 0 \\ \Phi \bullet (\Phi^{\bullet (n-1)}) & : n \in \mathbb{N} \end{cases} \tag{2.5}$$

(cf. Definitions 2.2, 2.4, and 2.7).

Lemma 2.9. *Let $d, \mathfrak{i} \in \mathbb{N}$, $\Psi \in \mathbf{N}$ satisfy $\mathcal{D}(\Psi) = (d, \mathfrak{i}, d)$ (cf. Definition 2.2). Then it holds for all $n \in \mathbb{N}_0$ that $\mathcal{H}(\Psi^{\bullet n}) = n$, $\mathcal{D}(\Psi^{\bullet n}) \in \mathbb{N}^{n+2}$, and*

$$\mathcal{D}(\Psi^{\bullet n}) = \begin{cases} (d, d) & : n = 0 \\ (d, \mathfrak{i}, \mathfrak{i}, \dots, \mathfrak{i}, d) & : n \in \mathbb{N} \end{cases} \tag{2.6}$$

(cf. Definition 2.8).

A proof of Lemma 2.9 can be found, e.g., in [26, Lemma 2.13].

2.4. Parallelizations of ANNs

Definition 2.10 *(Parallelization of ANNs with the same length).* Let $n \in \mathbb{N}$. Then we denote by

$$\mathbf{P}_n \colon \big\{(\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n : \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \dots = \mathcal{L}(\Phi_n)\big\} \to \mathbf{N} \tag{2.7}$$

the function which satisfies for all $L \in \mathbb{N}$, $(l_{1,0}, l_{1,1}, \dots, l_{1,L}), (l_{2,0}, l_{2,1}, \dots, l_{2,L}), \dots, (l_{n,0}, l_{n,1}, \dots, l_{n,L}) \in \mathbb{N}^{L+1}$, $\Phi_1 = ((W_{1,1}, B_{1,1}), (W_{1,2}, B_{1,2}), \dots, (W_{1,L}, B_{1,L})) \in (\times_{k=1}^{L} (\mathbb{R}^{l_{1,k} \times l_{1,k-1}} \times \mathbb{R}^{l_{1,k}}))$, $\Phi_2 = ((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \dots, (W_{2,L}, B_{2,L})) \in (\times_{k=1}^{L} (\mathbb{R}^{l_{2,k} \times l_{2,k-1}} \times \mathbb{R}^{l_{2,k}}))$, ..., $\Phi_n = ((W_{n,1}, B_{n,1}), (W_{n,2}, B_{n,2}), \dots, (W_{n,L}, B_{n,L})) \in (\times_{k=1}^{L} (\mathbb{R}^{l_{n,k} \times l_{n,k-1}} \times \mathbb{R}^{l_{n,k}}))$ that

$$\mathbf{P}_n(\Phi_1,\Phi_2,\dots,\Phi_n)=\left(\left(\begin{pmatrix} W_{1,1} & 0 & 0 & \cdots & 0\\ 0 & W_{2,1} & 0 & \cdots & 0\\ 0 & 0 & W_{3,1} & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & W_{n,1}\end{pmatrix},\begin{pmatrix} B_{1,1}\\ B_{2,1}\\ B_{3,1}\\ \vdots\\ B_{n,1}\end{pmatrix}\right),\right.$$
$$\left(\begin{pmatrix} W_{1,2} & 0 & 0 & \cdots & 0\\ 0 & W_{2,2} & 0 & \cdots & 0\\ 0 & 0 & W_{3,2} & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & W_{n,2}\end{pmatrix},\begin{pmatrix} B_{1,2}\\ B_{2,2}\\ B_{3,2}\\ \vdots\\ B_{n,2}\end{pmatrix}\right),\dots, \tag{2.8}$$
$$\left.\left(\begin{pmatrix} W_{1,L} & 0 & 0 & \cdots & 0\\ 0 & W_{2,L} & 0 & \cdots & 0\\ 0 & 0 & W_{3,L} & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & W_{n,L}\end{pmatrix},\begin{pmatrix} B_{1,L}\\ B_{2,L}\\ B_{3,L}\\ \vdots\\ B_{n,L}\end{pmatrix}\right)\right)$$

(cf. Definition 2.2).

Proposition 2.11. *Let $n\in\mathbb{N}$, $\Phi=(\Phi_1,\Phi_2,\dots,\Phi_n)\in\mathbf{N}^n$ satisfy $\mathcal{L}(\Phi_1)=\mathcal{L}(\Phi_2)=\ldots=\mathcal{L}(\Phi_n)$ (cf. Definition 2.2). Then*

(i) it holds that $\mathcal{R}(\mathbf{P}_n(\Phi))\in C(\mathbb{R}^{[\sum_{j=1}^n \mathcal{I}(\Phi_j)]},\mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]})$ and
(ii) it holds for all $x_1\in\mathbb{R}^{\mathcal{I}(\Phi_1)}, x_2\in\mathbb{R}^{\mathcal{I}(\Phi_2)},\dots,x_n\in\mathbb{R}^{\mathcal{I}(\Phi_n)}$ that

$$\begin{aligned}&\big(\mathcal{R}\big(\mathbf{P}_n(\Phi)\big)\big)(x_1,x_2,\dots,x_n)\\ &=\big((\mathcal{R}(\Phi_1))(x_1),(\mathcal{R}(\Phi_2))(x_2),\dots,(\mathcal{R}(\Phi_n))(x_n)\big)\in\mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]}\end{aligned}\tag{2.9}$$

(cf. Definition 2.10).

A proof of Proposition 2.11 can be found, e.g., in [26, Proposition 2.19].

Proposition 2.12. *Let $n\in\mathbb{N}$, $\Phi_1,\Phi_2,\dots,\Phi_n\in\mathbf{N}$ satisfy $\mathcal{L}(\Phi_1)=\mathcal{L}(\Phi_2)=\ldots=\mathcal{L}(\Phi_n)$ (cf. Definition 2.2). Then*

$$\mathcal{D}\big(\mathbf{P}_n(\Phi_1,\Phi_2,\dots,\Phi_n)\big)=\big(\textstyle\sum_{j=1}^n \mathbb{D}_0(\Phi_j),\sum_{j=1}^n \mathbb{D}_1(\Phi_j),\dots,\sum_{j=1}^n \mathbb{D}_L(\Phi_j)\big)\tag{2.10}$$

(cf. Definition 2.10).

A proof of Proposition 2.12 can be found, e.g., in [26, Proposition 2.20].

Definition 2.13. We denote by $\mathfrak{I}=(\mathfrak{I}_d)_{d\in\mathbb{N}}\colon\mathbb{N}\to\mathbf{N}$ the function which satisfies for all $d\in\mathbb{N}$ that

$$\mathfrak{I}_1=\left(\left(\begin{pmatrix}1\\-1\end{pmatrix},\begin{pmatrix}0\\0\end{pmatrix}\right),\left(\begin{pmatrix}1 & -1\end{pmatrix},0\right)\right)\in\left(\left(\mathbb{R}^{2\times 1}\times\mathbb{R}^2\right)\times\left(\mathbb{R}^{1\times 2}\times\mathbb{R}^1\right)\right)\tag{2.11}$$

and

$$\mathfrak{I}_d=\mathbf{P}_d(\mathfrak{I}_1,\mathfrak{I}_1,\dots,\mathfrak{I}_1)\tag{2.12}$$

(cf. Definitions 2.2 and 2.10).

Lemma 2.14. *Let $d \in \mathbb{N}$. Then*

(i) it holds that $\mathcal{D}(\mathfrak{I}_d) = (d, 2d, d) \in \mathbb{N}^3$,
(ii) it holds that $\mathcal{R}(\mathfrak{I}_d) \in C(\mathbb{R}^d, \mathbb{R}^d)$, and
(iii) it holds for all $x \in \mathbb{R}^d$ that $(\mathcal{R}(\mathfrak{I}_d))(x) = x$

(cf. Definitions 2.2 and 2.13).

A proof of Lemma 2.14 can be found, e.g., in [27, Lemma 3.16].

2.5. Linear transformations as ANNs

Definition 2.15 *(Affine linear transformation NN).* Let $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^m$. Then we denote by $\mathbf{A}_{W,B} \in (\mathbb{R}^{m\times n} \times \mathbb{R}^m) \subseteq \mathbf{N}$ the neural network given by $\mathbf{A}_{W,B} = (W, B)$ (cf. Definitions 2.2 and 2.3).

Lemma 2.16. *Let $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^m$. Then*

(i) it holds that $\mathcal{D}(\mathbf{A}_{W,B}) = (n, m) \in \mathbb{N}^2$,
(ii) it holds that $\mathcal{R}(\mathbf{A}_{W,B}) \in C(\mathbb{R}^n, \mathbb{R}^m)$, and
(iii) it holds for all $x \in \mathbb{R}^n$ that $(\mathcal{R}(\mathbf{A}_{W,B}))(x) = Wx + B$

(cf. Definitions 2.2 and 2.15).

The proof of Lemma 2.16 is clear and therefore is omitted.

2.6. Scalar multiplications of ANNs

Definition 2.17 *(Scalar multiplications of ANNs).* We denote by $(\cdot) \circledast (\cdot) : \mathbb{R} \times \mathbf{N} \to \mathbf{N}$ the function which satisfies for all $\lambda \in \mathbb{R}$, $\Phi \in \mathbf{N}$ that $\lambda \circledast \Phi = \mathbf{A}_{\lambda \mathrm{I}_{\mathcal{O}(\Phi)},0} \bullet \Phi$ (cf. Definitions 2.2, 2.4, 2.7, and 2.15).

Lemma 2.18. *Let $\lambda \in \mathbb{R}$, $\Phi \in \mathbf{N}$ (cf. Definition 2.2). Then*

(i) it holds that $\mathcal{D}(\lambda \circledast \Phi) = \mathcal{D}(\Phi)$,
(ii) it holds that $\mathcal{R}(\lambda \circledast \Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})$, and
(iii) it holds for all $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that $(\mathcal{R}(\lambda \circledast \Phi))(x) = \lambda\big((\mathcal{R}(\Phi))(x)\big)$

(cf. Definition 2.17).

A proof of Lemma 2.18 can be found, e.g., in [27, Lemma 3.14].

2.7. Sums of ANNs

Definition 2.19. Let $m, n \in \mathbb{N}$. Then we denote by $\mathfrak{S}_{m,n} \in (\mathbb{R}^{m\times(mn)} \times \mathbb{R}^m)$ the neural network given by $\mathfrak{S}_{m,n} = \mathbf{A}_{(\mathrm{I}_m\ \mathrm{I}_m\ \dots\ \mathrm{I}_m),0}$ (cf. Definitions 2.3, 2.7, and 2.15).

Lemma 2.20. *Let $m, n \in \mathbb{N}$. Then*

(i) it holds that $\mathcal{D}(\mathfrak{S}_{m,n}) = (mn, m) \in \mathbb{N}^2$,
(ii) it holds that $\mathcal{R}(\mathfrak{S}_{m,n}) \in C(\mathbb{R}^{mn}, \mathbb{R}^m)$, and
(iii) it holds for all $x_1, x_2, \dots, x_n \in \mathbb{R}^m$ that $(\mathcal{R}(\mathfrak{S}_{m,n}))(x_1, x_2, \dots, x_n) = \sum_{k=1}^n x_k$

(cf. Definitions 2.2 and 2.19).

A proof of Lemma 2.20 can be found, e.g., in [27, Lemma 3.18].

2.8. On the connection to the vectorized description of ANNs

Definition 2.21 *(p-norm).* We denote by $\|\cdot\|_p\colon \big(\bigcup_{d=1}^{\infty}\mathbb{R}^d\big) \to [0,\infty)$, $p\in[1,\infty]$, the functions which satisfy for all $p\in[1,\infty)$, $d\in\mathbb{N}$, $\theta=(\theta_1,\theta_2,\dots,\theta_d)\in\mathbb{R}^d$ that $\|\theta\|_p = \big(\sum_{i=1}^{d}|\theta_i|^p\big)^{1/p}$ and $\|\theta\|_\infty = \max_{i\in\{1,2,\dots,d\}}|\theta_i|$.

Definition 2.22. We denote by $\mathcal{T}\colon \mathbf{N} \to \big(\bigcup_{d\in\mathbb{N}}\mathbb{R}^d\big)$ the function which satisfies for all $L,d\in\mathbb{N}$, $l_0,l_1,\dots,l_L\in\mathbb{N}$, $\Phi=((W_1,B_1),(W_2,B_2),\dots,(W_L,B_L))\in(\times_{m=1}^{L}(\mathbb{R}^{l_m\times l_{m-1}}\times\mathbb{R}^{l_m}))$, $\theta=(\theta_1,\theta_2,\dots,\theta_d)\in\mathbb{R}^d$, $k\in\{1,2,\dots,L\}$ with $\mathcal{T}(\Phi)=\theta$ that

$$d=\mathcal{P}(\Phi),\qquad B_k=\begin{pmatrix}\theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+l_kl_{k-1}+1}\\ \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+l_kl_{k-1}+2}\\ \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+l_kl_{k-1}+3}\\ \vdots\\ \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+l_kl_{k-1}+l_k}\end{pmatrix},\qquad\text{and}$$

$$W_k=\begin{pmatrix}\theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+1} & \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+2} & \cdots & \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+l_{k-1}}\\ \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+l_{k-1}+1} & \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+l_{k-1}+2} & \cdots & \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+2l_{k-1}}\\ \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+2l_{k-1}+1} & \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+2l_{k-1}+2} & \cdots & \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+3l_{k-1}}\\ \vdots & \vdots & \ddots & \vdots\\ \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+(l_k-1)l_{k-1}+1} & \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+(l_k-1)l_{k-1}+2} & \cdots & \theta_{(\sum_{i=1}^{k-1}l_i(l_{i-1}+1))+l_kl_{k-1}}\end{pmatrix}\tag{2.13}$$

(cf. Definition 2.2).

Lemma 2.23. *Let* $L,\mathfrak{L}\in\mathbb{N}$, $l_0,l_1,\dots,l_L,\mathfrak{l}_0,\mathfrak{l}_1,\dots,\mathfrak{l}_{\mathfrak{L}}\in\mathbb{N}$, $\Phi_1=((W_1,B_1),(W_2,B_2),\dots,(W_L,B_L))\in(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k}))$, $\Phi_2=((\mathfrak{W}_1,\mathfrak{B}_1),(\mathfrak{W}_2,\mathfrak{B}_2),\dots,(\mathfrak{W}_{\mathfrak{L}},\mathfrak{B}_{\mathfrak{L}}))\in(\times_{k=1}^{\mathfrak{L}}(\mathbb{R}^{\mathfrak{l}_k\times\mathfrak{l}_{k-1}}\times\mathbb{R}^{\mathfrak{l}_k}))$. *Then*

$$\|\mathcal{T}(\Phi_1\bullet\Phi_2)\|_\infty\le\max\big\{\|\mathcal{T}(\Phi_1)\|_\infty,\|\mathcal{T}(\Phi_2)\|_\infty,\|\mathcal{T}\big(((W_1\mathfrak{W}_{\mathfrak{L}},W_1\mathfrak{B}_{\mathfrak{L}}+B_1))\big)\|_\infty\big\}\tag{2.14}$$

(cf. Definitions 2.4, 2.21, and 2.22).

Proof of Lemma 2.23. Observe that (2.3) and (2.13) establish (2.14). The proof of Lemma 2.23 is thus complete. □

3. Upper bounds for weighted Gaussian tails

In this section we establish suitable upper bounds for certain weighted Gaussian tail integrals in Lemma 3.9. Furthermore, we collect some upper and lower bounds for the Gamma function and unweighted Gaussian tail estimates needed in the subsequent sections.

3.1. Lower and upper bounds for evaluations of the Gamma function

Lemma 3.1. *Let* $\Gamma\colon(0,\infty)\to(0,\infty)$ *satisfy for all* $x\in(0,\infty)$ *that* $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}\,dt$. *Then*

(i) it holds that $\Gamma(1)=1$ *and*
(ii) it holds for all $d\in\mathbb{N}$ *that*

$$\Gamma\Big(\frac{d}{2}\Big)=\begin{cases}\big(\frac{d}{2}-1\big)! & :\frac{d}{2}\in\mathbb{N}\\ \frac{(d-1)!\sqrt{\pi}}{(\frac{d-1}{2})!\,2^{d-1}} & :\frac{d}{2}\notin\mathbb{N}.\end{cases}\tag{3.1}$$

Lemma 3.1 collects some basic well-known results of the Gamma function found, e.g., in Whittaker & Watson [58, Chapter XII] or Andrews et al. [1, Chapter 1].

Lemma 3.2. *Let* $\mathfrak{n} \in \mathbb{N}$. *Then*

$$\sqrt{2\pi\mathfrak{n}}\left[\frac{\mathfrak{n}}{e}\right]^{\mathfrak{n}} e^{\frac{1}{12\mathfrak{n}+1}} < \mathfrak{n}! < \sqrt{2\pi\mathfrak{n}}\left[\frac{\mathfrak{n}}{e}\right]^{\mathfrak{n}} e^{\frac{1}{12\mathfrak{n}}}. \tag{3.2}$$

The well-known Stirling inequalities of Lemma 3.2 are proved, e.g., in Robbins [56].

Corollary 3.3. *Let* $\mathfrak{m} \in \mathbb{N} \cap [2, \infty)$. *Then*

(i) it holds that

$$\sqrt{2\pi(\mathfrak{m}-1)}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} \le (\mathfrak{m}-1)! \le \sqrt{3\pi(\mathfrak{m}-1)}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} \tag{3.3}$$

and

(ii) it holds that

$$\sqrt{\pi}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} \le \frac{(2\mathfrak{m}-2)!\sqrt{\pi}}{4^{\mathfrak{m}-1}(\mathfrak{m}-1)!} \le \sqrt{2\pi}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1}. \tag{3.4}$$

Proof of Corollary 3.3. Note that Lemma 3.2 (applied with $\mathfrak{n} \curvearrowleft \mathfrak{m}-1$ in the notation of Lemma 3.2) implies that

$$\sqrt{2\pi(\mathfrak{m}-1)}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} e^{\frac{1}{12\mathfrak{m}-11}} \le (\mathfrak{m}-1)! \le \sqrt{2\pi(\mathfrak{m}-1)}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} e^{\frac{1}{12\mathfrak{m}-12}}. \tag{3.5}$$

The fact that $e \le \left(\frac{3}{2}\right)^{(6\mathfrak{m}-6)}$ therefore assures that

$$\begin{aligned}\sqrt{2\pi(\mathfrak{m}-1)}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} &\le \sqrt{2\pi(\mathfrak{m}-1)}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} e^{\frac{1}{12\mathfrak{m}-11}} \le (\mathfrak{m}-1)! \\ &\le \sqrt{2\pi(\mathfrak{m}-1)}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} e^{\frac{1}{12\mathfrak{m}-12}} \le \sqrt{3\pi(\mathfrak{m}-1)}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1}.\end{aligned} \tag{3.6}$$

This establishes item (i). Moreover, observe that Lemma 3.2 (applied with $\mathfrak{n} \curvearrowleft 2\mathfrak{m}-2$ in the notation of Lemma 3.2) ensures that

$$\sqrt{2\pi(2\mathfrak{m}-2)}\left[\frac{2\mathfrak{m}-2}{e}\right]^{2\mathfrak{m}-2} e^{\frac{1}{24\mathfrak{m}-23}} \le (2\mathfrak{m}-2)! \le \sqrt{2\pi(2\mathfrak{m}-2)}\left[\frac{2\mathfrak{m}-2}{e}\right]^{2\mathfrak{m}-2} e^{\frac{1}{24\mathfrak{m}-24}}. \tag{3.7}$$

Combining this with (3.5) demonstrates that

$$\sqrt{2\pi}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} e^{\frac{-12\mathfrak{m}+11}{(24\mathfrak{m}-23)(12\mathfrak{m}-12)}} \le \frac{(2\mathfrak{m}-2)!\sqrt{\pi}}{4^{\mathfrak{m}-1}(\mathfrak{m}-1)!} \le \sqrt{2\pi}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} e^{\frac{-12\mathfrak{m}+13}{(24\mathfrak{m}-24)(12\mathfrak{m}-11)}}. \tag{3.8}$$

The fact that $e^{(11-12\mathfrak{m})} \ge 2^{(24\mathfrak{m}-23)(6-6\mathfrak{m})}$ and the fact that for all $x \in [1, \infty)$ it holds that $x^{(-12m+13)} \le 1$ hence ensure that

$$\sqrt{\pi}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} \le \sqrt{2\pi}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} e^{\frac{-12\mathfrak{m}+11}{(24\mathfrak{m}-23)(12\mathfrak{m}-12)}} \le \frac{(2\mathfrak{m}-2)!\sqrt{\pi}}{4^{\mathfrak{m}-1}(\mathfrak{m}-1)!}$$
$$\le \sqrt{2\pi}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} e^{\frac{-12\mathfrak{m}+13}{(24\mathfrak{m}-24)(12\mathfrak{m}-11)}} \le \sqrt{2\pi}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1}. \tag{3.9}$$

This establishes item (ii). The proof of Corollary 3.3 is thus complete. □

Corollary 3.4. *Let $\Gamma\colon (0,\infty) \to (0,\infty)$ satisfy for all $x \in (0,\infty)$ that $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$. Then*

(i) it holds for all $\mathfrak{m} \in \mathbb{N} \cap [2,\infty)$ that

$$\sqrt{2\pi(\mathfrak{m}-1)}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} \le \Gamma(\mathfrak{m}) \le \sqrt{3\pi(\mathfrak{m}-1)}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} \tag{3.10}$$

and

(ii) it holds for all $\mathfrak{m} \in \mathbb{N} \cap [2,\infty)$ that

$$\sqrt{\pi}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1} \le \Gamma\!\left(\mathfrak{m}-\frac{1}{2}\right) \le \sqrt{2\pi}\left[\frac{\mathfrak{m}-1}{e}\right]^{\mathfrak{m}-1}. \tag{3.11}$$

Proof of Corollary 3.4. Note that Corollary 3.3 and item (ii) in Lemma 3.1 establish items (i) and (ii). The proof of Corollary 3.4 is thus complete. □

3.2. Lower and upper bounds for Gaussian tails

Lemma 3.5. *Let $\sigma, s \in (0,\infty)$. Then*

(i) it holds that $\int_0^\infty e^{-\sigma x^2}\,dx = \frac{\sqrt{\pi}}{2\sqrt{\sigma}}$,

(ii) it holds that

$$\int_s^\infty e^{-\sigma x^2}\,dx \le \left[\frac{\sqrt{\pi}}{2\sqrt{\sigma}}\right] e^{-\sigma s^2}, \tag{3.12}$$

and

(iii) it holds that

$$\int_0^s e^{-\sigma x^2}\,dx \ge \left[\frac{\sqrt{\pi}}{2\sqrt{\sigma}}\right]\left(1 - e^{-\sigma s^2}\right). \tag{3.13}$$

Proof of Lemma 3.5. Observe that the integral transformation theorem shows that

$$\int_0^\infty e^{-\sigma x^2}\,dx = \frac{1}{\sqrt{2\sigma}}\int_0^\infty \exp(-\tfrac{x^2}{2})\,dx = \frac{\sqrt{\pi}}{\sqrt{\sigma}}\int_0^\infty \frac{\exp(-\frac{x^2}{2})}{\sqrt{2\pi}}\,dx = \frac{\sqrt{\pi}}{2\sqrt{\sigma}}. \tag{3.14}$$

This establishes item (i). Next note that the integral transformation theorem and (3.14) ensure that

$$\int_s^\infty e^{-\sigma x^2}\,dx = \int_0^\infty e^{-\sigma(x+s)^2}\,dx = \int_0^\infty \left(e^{-\sigma x^2 - 2\sigma s x - \sigma s^2}\right)dx$$
$$= e^{-\sigma s^2}\left[\int_0^\infty e^{-\sigma x^2 - 2\sigma s x}\,dx\right] \le e^{-\sigma s^2}\left[\int_0^\infty e^{-\sigma x^2}\,dx\right] = \left[\frac{\sqrt{\pi}}{2\sqrt{\sigma}}\right]e^{-\sigma s^2}. \tag{3.15}$$

This establishes item (ii). Next we combine (3.14) and (3.15) to obtain that

$$\int_0^s e^{-\sigma x^2}\,dx = \int_0^\infty e^{-\sigma x^2}\,dx - \int_s^\infty e^{-\sigma x^2}\,dx = \frac{\sqrt{\pi}}{2\sqrt{\sigma}} - \int_s^\infty e^{-\sigma x^2}\,dx \geq \left[\frac{\sqrt{\pi}}{2\sqrt{\sigma}}\right]\big(1 - e^{-\sigma s^2}\big). \quad (3.16)$$

This establishes item (iii). The proof of Lemma 3.5 is thus complete. □

Corollary 3.6. *Let $d \in \mathbb{N}$, $\sigma, s \in (0,\infty)$. Then*

(i) it holds that

$$\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\leq s\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\,dx \geq \left[1 - e^{-\sigma s^2/d}\right]^d \quad (3.17)$$

and

(ii) it holds that

$$\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\geq s\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\,dx \leq d e^{-\sigma s^2/d} \quad (3.18)$$

(cf. Definition 2.21).

Proof of Corollary 3.6. Observe that item (i) in Lemma 3.5 implies that

$$\begin{aligned}\int_{\mathbb{R}^d} e^{-\sigma\|x\|_2^2}\,dx &= \int_{\mathbb{R}}\int_{\mathbb{R}}\ldots\int_{\mathbb{R}} e^{-\sigma(|x_1|^2+|x_2|^2+\ldots+|x_d|^2)}\,dx_d\ldots dx_2 dx_1 \\ &= \left[\int_{\mathbb{R}} e^{-\sigma x^2}\,dx\right]^d = \left[2\int_0^\infty e^{-\sigma x^2}\,dx\right]^d = \left[\frac{\pi}{\sigma}\right]^{d/2}\end{aligned} \quad (3.19)$$

(cf. Definition 2.21). Next note that item (iii) in Lemma 3.5 (applied with $\sigma \curvearrowleft \sigma$, $s \curvearrowleft d^{-1/2}s$ in the notation of Lemma 3.5) and the fact that

$$\{y = (y_1, y_2, \ldots, y_d) \in \mathbb{R}^d\colon (\forall\, j \in \{1,2,\ldots,d\}\colon |y_j| \leq d^{-1/2}s)\} \subseteq \{y \in \mathbb{R}^d\colon \|y\|_2 \leq s\} \quad (3.20)$$

ensure that

$$\begin{aligned}\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\leq s\}} e^{-\sigma\|x\|_2^2}\,dx &\geq \prod_{j=1}^d \left[\int_{-d^{-1/2}s}^{d^{-1/2}s} e^{-\sigma|x_j|^2}\,dx_j\right] \\ &= \left[2\int_0^{d^{-1/2}s} e^{-\sigma|x|^2}\,dx\right]^d \geq \left[\frac{\pi}{\sigma}\right]^{d/2}\left[1 - e^{-\sigma s^2/d}\right]^d.\end{aligned} \quad (3.21)$$

This immediately establishes item (i). For item (ii), we combine (3.19) and (3.21) with Bernoulli's inequality, i.e., the well-known fact that for all $\alpha \in \mathbb{R}\backslash(0,1)$ and for all $x \in (-1,\infty)$ it holds that $(1+x)^\alpha \geq 1+\alpha x$, to obtain

$$\int_{\{y\in\mathbb{R}^d\colon\,\|y\|_2\geq s\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\,dx = 1 - \int_{\{y\in\mathbb{R}^d\colon\,\|y\|_2\leq s\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\,dx \\ \leq 1 - \left[1 - e^{-\sigma s^2/d}\right]^d \leq 1 - \left[1 - de^{-\sigma s^2/d}\right] = de^{-\sigma s^2/d}. \tag{3.22}$$

The proof of Corollary 3.6 is thus complete. □

Lemma 3.7. *Let $d \in \mathbb{N}$, $\sigma \in (0,\infty)$, $\alpha, s \in [0,\infty)$ and let $\Gamma\colon (0,\infty) \to (0,\infty)$ satisfy for all $x \in (0,\infty)$ that $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$. Then*

$$\int_{\{y\in\mathbb{R}^d\colon\,\|y\|_2\geq s\}} \left[\frac{\sigma}{\pi}\right]^{d/2} \|x\|_2^\alpha\, e^{-\sigma\|x\|_2^2}\,dx = \frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)}\left[\int_s^\infty e^{-\sigma r^2} r^{\alpha+d-1}\,dr\right] \tag{3.23}$$

(cf. Definition 2.21).

Lemma 3.7 is a direct consequence of the integral transformation theorem.

Lemma 3.8. *Let $d \in \mathbb{N} \cap [3,\infty)$, $\beta, \sigma \in (0,\infty)$. Then*

$$\int_{\left\{y\in\mathbb{R}^d\colon\,\frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}\leq\|y\|_2\leq\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\,dx \leq d\left[\frac{1+\beta}{e^\beta}\right]^{d/2} \tag{3.24}$$

(cf. Definition 2.21).

Proof of Lemma 3.8. Throughout this proof let $\Gamma\colon (0,\infty) \to (0,\infty)$ satisfy for all $x \in (0,\infty)$ that $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$. Observe that Lemma 3.7 (applied with $d \curvearrowleft d$, $\sigma \curvearrowleft \sigma$, $\alpha \curvearrowleft 0$, $s \curvearrowleft (2\sigma)^{-1/2}(d(1+\beta))^{1/2}$ in the notation of Lemma 3.7) implies that

$$\int_{\left\{y\in\mathbb{R}^d\colon\,\|y\|_2\geq\frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\,dx = \frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int_{\frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}}^\infty e^{-\sigma r^2} r^{d-1}\,dr \tag{3.25}$$

(cf. Definition 2.21). Next note that Lemma 3.7 (applied with $d \curvearrowleft d$, $\sigma \curvearrowleft \sigma$, $\alpha \curvearrowleft 0$, $s \curvearrowleft (2\sigma)^{-1/2}d(1+\beta)^{1/2}$ in the notation of Lemma 3.7) shows that

$$\int_{\left\{y\in\mathbb{R}^d\colon\,\|y\|_2\geq\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\,dx = \frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int_{\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}}^\infty e^{-\sigma r^2} r^{d-1}\,dr. \tag{3.26}$$

Combining this with (3.25) ensures that

$$\int_{\left\{y\in\mathbb{R}^d\colon\,\frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}\leq\|y\|_2\leq\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\,dx \\ = \int_{\left\{y\in\mathbb{R}^d\colon\,\|y\|_2\geq\frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\,dx - \int_{\left\{y\in\mathbb{R}^d\colon\,\|y\|_2\geq\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\,dx \tag{3.27}$$

$$= \frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int_{\frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}}^{\infty} e^{-\sigma r^2} r^{d-1}\, dr - \frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int_{\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}}^{\infty} e^{-\sigma r^2} r^{d-1}\, dr$$

$$= \frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int_{\frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}}^{\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}} e^{-\sigma r^2} r^{d-1}\, dr.$$

Next observe that the chain rule ensures that for all $x \in [(2\sigma)^{-1/2} d^{1/2}, \infty)$ it holds that

$$\big[e^{-\sigma x^2} x^{d-1}\big]' = e^{-\sigma x^2} x^{d-2}(d-1-2\sigma x^2) \le e^{-\sigma x^2} x^{d-2}(d-1-d) < 0. \tag{3.28}$$

This ensures that the function $[(2\sigma)^{-1/2} d^{1/2}, \infty) \ni x \mapsto e^{-\sigma x^2} x^{d-1} \in \mathbb{R}$ is strictly decreasing. Hence, we obtain that

$$\begin{aligned} &\frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int_{\frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}}^{\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}} e^{-\sigma r^2} r^{d-1}\, dr \le \frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int_{\frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}}^{\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}} \left[e^{-\frac{d(1+\beta)}{2}}\right]\left[\frac{d(1+\beta)}{2\sigma}\right]^{\frac{d-1}{2}} dr \\ &= \frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \left[e^{-\frac{d(1+\beta)}{2}}\right]\left[\frac{d(1+\beta)}{2\sigma}\right]^{\frac{d-1}{2}} \left[\frac{(d-\sqrt{d})\sqrt{1+\beta}}{\sqrt{2\sigma}}\right] \\ &\le \frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \left[e^{-\frac{d(1+\beta)}{2}}\right]\left[\frac{d(1+\beta)}{2\sigma}\right]^{\frac{d-1}{2}} \left[d\left[\frac{1+\beta}{2\sigma}\right]^{1/2}\right] = \frac{2}{\Gamma\left(\frac{d}{2}\right)} \left[e^{-\frac{d(1+\beta)}{2}}\right]\left[d^{\frac{d+1}{2}}\right]\left[\frac{1+\beta}{2}\right]^{d/2}. \end{aligned} \tag{3.29}$$

Next note that item (i) in Corollary 3.4 and the fact that for all $m \in \mathbb{N} \cap [2, \infty)$ it holds that

$$\left[\left[1+\frac{1}{m-1}\right]^{m-1}\right]^{\frac{2m-1}{2m-2}} \le e\left[\left[1+\frac{1}{m-1}\right]^{m-1}\right]^{\frac{1}{2m-2}} = e\left[1+\frac{1}{m-1}\right]^{1/2} \le e\big[2^{1/2}\big] \le \frac{3e}{2} \tag{3.30}$$

assure that for all $k, m \in \mathbb{N}$ with $k = 2m \ge 4$ it holds that

$$\begin{aligned} &\frac{2}{\Gamma\left(\frac{k}{2}\right)} \left[e^{-\frac{k(1+\beta)}{2}}\right]\left[k^{\frac{k+1}{2}}\right]\left[\frac{1+\beta}{2}\right]^{k/2} = \frac{2}{\Gamma(m)} \left[e^{-m(1+\beta)}\right]\left[(2m)^{m+\frac{1}{2}}\right]\left[\frac{1+\beta}{2}\right]^{m} \\ &\le \frac{2}{\sqrt{2\pi(m-1)}} \left[\frac{e}{m-1}\right]^{m-1} \left[e^{-m(1+\beta)}\right]\left[(2m)^{m+\frac{1}{2}}\right]\left[\frac{1+\beta}{2}\right]^{m} \\ &= \frac{2m}{\sqrt{\pi}} \left[e^{-1-m\beta}\right]\left[\left[1+\frac{1}{m-1}\right]^{m-1}\right]^{\frac{2m-1}{2m-2}} (1+\beta)^m \le \frac{2m}{\sqrt{\pi}} \left[e^{-1-m\beta}\right]\left[\frac{3e}{2}\right](1+\beta)^m \\ &= \frac{3k}{2\sqrt{\pi}} \left[\frac{1+\beta}{e^{\beta}}\right]^{k/2} \le k\left[\frac{1+\beta}{e^{\beta}}\right]^{k/2}. \end{aligned} \tag{3.31}$$

Next observe that item (ii) in Corollary 3.4 and the fact that for all $m \in \mathbb{N} \cap [2, \infty)$ it holds that $(1 + (2m-2)^{-1})^{2m-2} \le e$ show that for all $k, m \in \mathbb{N}$ with $k = 2m - 1 \ge 3$ it holds that

$$
\begin{aligned}
&\left[\frac{2}{\Gamma\left(\frac{k}{2}\right)}\right]\left[e^{-\frac{k(1+\beta)}{2}} k^{\frac{k+1}{2}}\right]\left[\frac{1+\beta}{2}\right]^{k/2}=\left[\frac{2}{\Gamma\left(m-\frac{1}{2}\right)}\right]\left[e^{(-m+\frac{1}{2})(1+\beta)}(2m-1)^m\right]\left[\frac{1+\beta}{2}\right]^{m-\frac{1}{2}} \\
&\leq \frac{2}{\sqrt{\pi}}\left[\frac{e}{m-1}\right]^{m-1}\left[e^{(-m+\frac{1}{2})(1+\beta)}(2m-1)^m\right]\left[\frac{1+\beta}{2}\right]^{m-\frac{1}{2}} \\
&=\left[\frac{2}{\pi}\right]^{1/2}\left[(2m-1)e^{-m\beta+\frac{\beta}{2}-\frac{1}{2}}\right]\left[\left[1+\frac{1}{2m-2}\right]^{2m-2}\right]^{1/2}(1+\beta)^{m-\frac{1}{2}} \\
&\leq \left[\frac{2}{\pi}\right]^{1/2}(2m-1)\, e^{-m\beta+\frac{\beta}{2}-\frac{1}{2}}\, e^{\frac{1}{2}}(1+\beta)^{m-\frac{1}{2}}=\left[\frac{2}{\pi}\right]^{1/2} k\left[\frac{1+\beta}{e^{\beta}}\right]^{k/2} \leq k\left[\frac{1+\beta}{e^{\beta}}\right]^{k/2}.
\end{aligned}
\tag{3.32}
$$

Combining this with (3.29) and (3.31) assures that

$$
\frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int_{\frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}}^{\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}} e^{-\sigma r^2} r^{d-1}\, dr \leq d\left[\frac{1+\beta}{e^{\beta}}\right]^{d/2}. \tag{3.33}
$$

This and (3.27) imply that

$$
\int_{\left\{y\in\mathbb{R}^d\colon \frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}} \leq \|y\|_2 \leq \frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\, dx \leq d\left[\frac{1+\beta}{e^{\beta}}\right]^{d/2}. \tag{3.34}
$$

The proof of Lemma 3.8 is thus complete. □

3.3. Upper bounds for weighted Gaussian tails

Lemma 3.9. *Let $d \in \mathbb{N} \cap [3, \infty)$, $\beta, \sigma \in (0, \infty)$, $k \in \mathbb{N}_0$ and let $\Gamma\colon (0, \infty) \to (0, \infty)$ satisfy for all $x \in (0, \infty)$ that $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\, dt$. Then*

$$
\begin{aligned}
&\int_{\left\{y\in\mathbb{R}^d\colon \|y\|_2 \geq \frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} \|x\|_2^k\, e^{-\sigma\|x\|_2^2}\, dx \\
&\leq d^{1+k}\left[\frac{1+\beta}{2\sigma}\right]^{k/2}\left[\frac{1+\beta}{e^{\beta}}\right]^{d/2} + \frac{\Gamma\left(\frac{d+k}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\left[\frac{d+k}{\sigma^{k/2}}\right] e^{-\frac{d^2(1+\beta)}{2(d+k)}}
\end{aligned}
\tag{3.35}
$$

(cf. Definition 2.21).

Proof of Lemma 3.9. Note that Lemma 3.8 (applied with $d \curvearrowleft d$, $\beta \curvearrowleft \beta$, $\sigma \curvearrowleft \sigma$ in the notation of Lemma 3.8) ensures that

$$
\int_{\left\{y\in\mathbb{R}^d\colon \frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}} \leq \|y\|_2 \leq \frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} \|x\|_2^k\, e^{-\sigma\|x\|_2^2}\, dx
$$

$$\le d^k \left[\frac{1+\beta}{2\sigma}\right]^{k/2} \int\limits_{\left\{y\in\mathbb{R}^d:\, \frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}} \le \|y\|_2 \le \frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\, dx \le d^{1+k}\left[\frac{1+\beta}{2\sigma}\right]^{k/2}\left[\frac{1+\beta}{e^\beta}\right]^{d/2} \tag{3.36}$$

(cf. Definition 2.21). Moreover, observe that Lemma 3.7 (applied with $d \curvearrowleft d+k$, $\sigma \curvearrowleft \sigma$, $\alpha \curvearrowleft 0$, $s \curvearrowleft (2\sigma)^{-1/2} d(1+\beta)^{1/2}$ in the notation of Lemma 3.7) and item (ii) in Corollary 3.6 (applied with $d \curvearrowleft d+k$, $\sigma \curvearrowleft \sigma$, $s \curvearrowleft (2\sigma)^{-1/2} d(1+\beta)^{1/2}$ in the notation of Corollary 3.6) assure that

$$\frac{2\sigma^{\frac{d+k}{2}}}{\Gamma\left(\frac{d+k}{2}\right)} \int_{\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}}^{\infty} e^{-\sigma r^2} r^{d+k-1}\, dr = \int\limits_{\left\{y\in\mathbb{R}^{d+k}:\, \|y\|_2 \ge \frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{\frac{d+k}{2}} e^{-\sigma\|x\|_2^2}\, dx \le (d+k) e^{-\frac{d^2(1+\beta)}{2(d+k)}}. \tag{3.37}$$

Lemma 3.7 (applied with $d \curvearrowleft d$, $\sigma \curvearrowleft \sigma$, $\alpha \curvearrowleft k$, $s \curvearrowleft (2\sigma)^{-1/2} d(1+\beta)^{1/2}$ in the notation of Lemma 3.7) hence shows that

$$\begin{aligned} &\int\limits_{\left\{y\in\mathbb{R}^d:\, \|y\|_2 \ge \frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} \|x\|_2^k\, e^{-\sigma\|x\|_2^2}\, dx = \frac{2\sigma^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int_{\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}}^{\infty} e^{-\sigma r^2} r^{d+k-1}\, dr \\ &= \frac{\Gamma\left(\frac{d+k}{2}\right)}{\Gamma\left(\frac{d}{2}\right)\sigma^{k/2}} \left[\frac{2\sigma^{\frac{d+k}{2}}}{\Gamma\left(\frac{d+k}{2}\right)} \int_{\frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}}^{\infty} e^{-\sigma r^2} r^{d+k-1}\, dr\right] \le \frac{\Gamma\left(\frac{d+k}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \left[\frac{d+k}{\sigma^{k/2}}\right] e^{-\frac{d^2(1+\beta)}{2(d+k)}}. \end{aligned} \tag{3.38}$$

Combining this with (3.36) demonstrates that

$$\begin{aligned} &\int\limits_{\left\{y\in\mathbb{R}^d:\, \|y\|_2 \ge \frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} \|x\|_2^k\, e^{-\sigma\|x\|_2^2}\, dx \\ &= \int\limits_{\left\{y\in\mathbb{R}^d:\, \frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}} \le \|y\|_2 \le \frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} \|x\|_2^k\, e^{-\sigma\|x\|_2^2}\, dx \\ &\quad + \int\limits_{\left\{y\in\mathbb{R}^d:\, \|y\|_2 \ge \frac{d\sqrt{1+\beta}}{\sqrt{2\sigma}}\right\}} \left[\frac{\sigma}{\pi}\right]^{d/2} \|x\|_2^k\, e^{-\sigma\|x\|_2^2}\, dx \\ &\le d^{1+k}\left[\frac{1+\beta}{2\sigma}\right]^{k/2}\left[\frac{1+\beta}{e^\beta}\right]^{d/2} + \frac{\Gamma\left(\frac{d+k}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \left[\frac{d+k}{\sigma^{k/2}}\right] e^{-\frac{d^2(1+\beta)}{2(d+k)}}. \end{aligned} \tag{3.39}$$

The proof of Lemma 3.9 is thus complete. □

4. Lower bounds for the number of ANN parameters in the approximation of high-dimensional functions

This section compiles the main work for proving item (ii) in Theorem 1.1. The key result is Theorem 4.1, which establishes a lower bound for the depth, number and size of parameters of the approximating ANN in terms of the approximation error and the concentration of the target function around the origin.

Theorem 4.1. *Let $d \in \mathbb{N} \cap [4,\infty)$, $\beta, \sigma \in (0,\infty)$, $\Phi \in \mathbf{N}$ satisfy $\mathcal{I}(\Phi) = d$ and $\mathcal{O}(\Phi) = 1$, let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$, $g\colon \mathbb{R}^d \to \mathbb{R}$, and $\mathfrak{g}\colon \mathbb{R}^d \to \mathbb{R}$ be measurable, and assume for all $x \in \mathbb{R}^d$ that $\varphi(x) = (\sigma/\pi)^{d/2} \exp\big(-\sigma \|x\|_2^2\big)$, $\int_{\mathbb{R}^d} |g(y)|^2 \varphi(y)\,dy \in (0,\infty)$, $\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(y)|\,dy > 0$, and $\mathfrak{g}(x) = \big[\int_{\mathbb{R}^d} |g(y)|^2 \varphi(y)\,dy\big]^{-1/2} g(x)$ (cf. Definitions 2.2 and 2.21). Then*

$$\mathcal{L}(\Phi)\left[\frac{\mathcal{P}(\Phi)\max\{1,\|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)}\right]^{\mathcal{L}(\Phi)} \geq \left[\frac{e^\beta}{1+\beta}\right]^{d/6}\left[\frac{2\sqrt{\sigma}\,\big[\int_{\mathbb{R}^d}|(\mathcal{R}(\Phi))(x)|^2\varphi(x)\,dx\big]^{1/2}}{d^{3/2}(6+4\beta+\sigma)^{1/2}}\right] \cdot\left[1-\left[\int_{\left\{y\in\mathbb{R}^d\colon \|y\|_2\leq\frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}\right\}}|\mathfrak{g}(x)|^2\varphi(x)\,dx\right]^{1/2} - \int_{\mathbb{R}^d}|(\mathcal{R}(\Phi))(x)-\mathfrak{g}(x)|^2\varphi(x)\,dx\right] \tag{4.1}$$

(cf. Definition 2.22).

The idea of this theorem is as follows: Given a measurable target function $\mathfrak{g}\colon \mathbb{R}^d \to \mathbb{R}$ with $\int_{\mathbb{R}^d} |\mathfrak{g}(y)|^2 \varphi(y)\,dy = 1$, the approximation error can be estimated by

$$\begin{aligned}\int_{\mathbb{R}^d}|(\mathcal{R}(\Phi))(x)-\mathfrak{g}(x)|^2\varphi(x)\,dx &\geq 1 - C_\Phi\int_{\mathbb{R}^d}|(\mathcal{R}(\Phi))(x)|\,|\mathfrak{g}(x)|\,\varphi(x)\,dx\\ &= 1 - C_\Phi\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\leq\mathscr{R}\}}|(\mathcal{R}(\Phi))(x)|\,|\mathfrak{g}(x)|\,\varphi(x)\,dx - C_\Phi\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\geq\mathscr{R}\}}|(\mathcal{R}(\Phi))(x)|\,|\mathfrak{g}(x)|\,\varphi(x)\,dx,\end{aligned} \tag{4.2}$$

where $C_\Phi = \big[\int_{\mathbb{R}^d}|(\mathcal{R}(\Phi))(x)|^2\varphi(x)\,dx\big]^{-1/2}$ appears as a normalization term. Inequality (4.2) is already quite similar to (4.1) up to some rearranging, that is bringing the approximation error and the integral over $\{y \in \mathbb{R}^d\colon \|y\|_2 \geq \mathscr{R}\}$ to their respective other side.

The main work of this section will be to bound the integral over $\{y \in \mathbb{R}^d\colon \|y\|_2 \geq \mathscr{R}\}$ in terms of ANN parameters. This will be done in Subsection 4.1 and Subsection 4.2. In Subsection 4.3 we verify the first inequality of (4.2). We then have all the technical tools to give a formal proof of Theorem 4.1 in Subsection 4.4. Finally in Subsection 4.5, we apply Theorem 4.1 to the functions $\mathfrak{g}_d\colon \mathbb{R}^d \to \mathbb{R}$ defined by taking $g_d\colon \mathbb{R}^d \to \mathbb{R}$ with $g_d(x) = \sum_{j=1}^d [\max\{|x_j| - \sqrt{2d}, 0\}]^2$ and setting $\mathfrak{g}_d(x) = [\int_{\mathbb{R}^d}|g_d(y)|^2\varphi(y)\,dy]^{-1/2} g_d(x)$. The result is Corollary 4.9, which establishes item (ii) of Theorem 1.1.

4.1. Upper bounds for realizations of ANNs

Lemma 4.2. *Let $m, n \in \mathbb{N}$, $A = (A_{i,j})_{(i,j)\in\{1,2,\dots,m\}\times\{1,2,\dots,n\}} \in \mathbb{R}^{m\times n}$, $B = (B_1, B_2, \dots, B_m) \in \mathbb{R}^m$, $x \in \mathbb{R}^n$. Then*

(i) it holds that

$$\|Ax + B\|_\infty \leq \sqrt{n}\Big[\max_{i\in\{1,2,\dots,m\}}\max_{j\in\{1,2,\dots,n\}}|A_{i,j}|\Big]\|x\|_2 + \|B\|_\infty \tag{4.3}$$

and

(ii) it holds that

$$\|Ax+B\|_\infty \le n\Big[\max_{i\in\{1,2,\dots,m\}}\max_{j\in\{1,2,\dots,n\}}|A_{i,j}|\Big]\|x\|_\infty+\|B\|_\infty \tag{4.4}$$

(cf. Definition 2.21).

Proof of Lemma 4.2. Throughout this proof let $\alpha\in\mathbb{R}$ satisfy $\alpha=\max_{i\in\{1,2,\dots,m\}}\max_{j\in\{1,2,\dots,n\}}|A_{i,j}|$ and let $\beta\in\mathbb{R}$ satisfy $\beta=\|B\|_\infty$ (cf. Definition 2.21). Note that the triangle inequality and the fact that for all $v=(v_1,v_2,\dots,v_n)\in\mathbb{R}^n$ it holds that $\sum_{j=1}^n|v_j|\le\sqrt{n}\|v\|_2$ ensure that for all $v=(v_1,v_2,\dots,v_n)\in\mathbb{R}^n$ it holds that

$$\begin{aligned}\|Av+B\|_\infty&=\max_{i\in\{1,2,\dots,m\}}\Big|B_i+\sum_{j=1}^n A_{i,j}v_j\Big|\le\max_{i\in\{1,2,\dots,m\}}\Big(|B_i|+\sum_{j=1}^n|A_{i,j}v_j|\Big)\\&\le\beta+\alpha\sum_{j=1}^n|v_j|\le\beta+\alpha\sqrt{n}\|v\|_2.\end{aligned}\tag{4.5}$$

This establishes item (i). Moreover, observe that the fact that for all $v\in\mathbb{R}^n$ it holds that $\sqrt{n}\|v\|_2\le n\|v\|_\infty$ and item (i) demonstrate that for all $v\in\mathbb{R}^n$ it holds that

$$\|Av+B\|_\infty\le\beta+\alpha\sqrt{n}\|v\|_2\le\beta+\alpha n\|v\|_\infty.\tag{4.6}$$

This establishes item (ii). The proof of Lemma 4.2 is thus complete. □

Lemma 4.3. *Let $L\in\mathbb{N}\cap[2,\infty)$, $l_0,l_1,\dots,l_L\in\mathbb{N}$, $\Phi=((W_1,B_1),(W_2,B_2),\dots,(W_L,B_L))\in\big(\times_{k=1}^L(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k})\big)$, $x_0\in\mathbb{R}^{l_0},x_1\in\mathbb{R}^{l_1},\dots,x_L\in\mathbb{R}^{l_L}$ satisfy for all $k\in\{1,2,\dots,L\}$ that $x_k=\mathfrak{R}(W_kx_{k-1}+B_k)$ (cf. Definition 2.1). Then*

(i) it holds for all $k\in\{1,2,\dots,L\}$, $j\in\{1,2,\dots,k\}$ that

$$\|x_k\|_\infty\le l_{k-1}l_{k-2}\cdots l_{k-j}\,(\max\{1,\|\mathcal{T}(\Phi)\|_\infty\})^j(\|x_{k-j}\|_\infty+j)\tag{4.7}$$

and

(ii) it holds that

$$\|(\mathcal{R}(\Phi))(x_0)\|_\infty\le l_{L-1}l_{L-2}\cdots l_1(\max\{1,\|\mathcal{T}(\Phi)\|_\infty\})^{L-1}(\|x_1\|_\infty+L-1)\tag{4.8}$$

(cf. Definitions 2.2, 2.21, and 2.22).

Proof of Lemma 4.3. Throughout this proof let $\alpha=\max\{1,\|\mathcal{T}(\Phi)\|_\infty\}$ (cf. Definitions 2.21 and 2.22). Note that the fact that for all $x\in\mathbb{R}$ it holds that $|\max\{x,0\}|\le|x|$ and item (ii) in Lemma 4.2 (applied for every $k\in\{1,2,\dots,L\}$ with $m\curvearrowleft l_k$, $n\curvearrowleft l_{k-1}$, $A\curvearrowleft W_k$, $B\curvearrowleft B_k$, $x\curvearrowleft x_{k-1}$ in the notation of Lemma 4.2) imply that for all $k\in\{1,2,\dots,L\}$ it holds that

$$\begin{aligned}\|x_k\|_\infty&=\|\mathfrak{R}(W_kx_{k-1}+B_k)\|_\infty\le\|W_kx_{k-1}+B_k\|_\infty\\&\le\alpha\,l_{k-1}\|x_{k-1}\|_\infty+\alpha\le\alpha\,l_{k-1}(\|x_{k-1}\|_\infty+1).\end{aligned}\tag{4.9}$$

This demonstrates that for all $k\in\{2,3,\dots,L\}$, $i\in\{1,2,\dots,k-1\}$ with $\|x_k\|_\infty\le l_{k-1}l_{k-2}\cdots l_{k-i}\,\alpha^i(\|x_{k-i}\|_\infty+i)$ it holds that

$$\begin{aligned}\|x_k\|_\infty&\le l_{k-1}l_{k-2}\cdots l_{k-i}\alpha^i(\|x_{k-i}\|_\infty+i)\\&\le l_{k-1}l_{k-2}\cdots l_{k-i}\alpha^i(\alpha l_{k-i-1}(\|x_{k-i-1}\|_\infty+1)+i)\\&\le l_{k-1}l_{k-2}\cdots l_{k-i}l_{k-i-1}\alpha^{i+1}(\|x_{k-i-1}\|_\infty+i+1).\end{aligned}\tag{4.10}$$

This, (4.9), and induction show that for all $k \in \{1, 2, \dots, L\}$, $j \in \{1, 2, \dots, k\}$ it holds that

$$\|x_k\|_\infty \le l_{k-1} l_{k-2} \cdots l_{k-j} \, (\max\{1, \|\mathcal{T}(\Phi)\|_\infty\})^j (\|x_{k-j}\|_\infty + j). \tag{4.11}$$

This establishes item (i). Next observe that item (ii) in Lemma 4.2 (applied with $m \curvearrowleft l_L$, $n \curvearrowleft l_{L-1}$, $A \curvearrowleft W_L$, $B \curvearrowleft B_L$, $x \curvearrowleft x_{L-1}$ in the notation of Lemma 4.2) ensures that

$$\|(\mathcal{R}(\Phi))(x_0)\|_\infty = \|W_L x_{L-1} + B_L\|_\infty \le \alpha l_{L-1} \|x_{L-1}\|_\infty + \alpha \le \alpha l_{L-1} (\|x_{L-1}\|_\infty + 1) \tag{4.12}$$

(cf. Definition 2.2). This and item (i) demonstrate that

$$\begin{aligned} \|(\mathcal{R}(\Phi))(x_0)\|_\infty &\le \alpha l_{L-1} (\|x_{L-1}\|_\infty + 1) \\ &\le \alpha l_{L-1} ([l_{L-2} l_{L-3} \cdots l_1 \alpha^{L-2} (\|x_1\|_\infty + L - 2)] + 1) \\ &\le l_{L-1} l_{L-2} \cdots l_1 \alpha^{L-1} (\|x_1\|_\infty + L - 1). \end{aligned} \tag{4.13}$$

This establishes item (ii). The proof of Lemma 4.3 is thus complete. □

Corollary 4.4. *It holds for all $\Phi \in \mathbf{N}$, $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that*

$$\|(\mathcal{R}(\Phi))(x)\|_\infty \le \left[\frac{\mathcal{P}(\Phi) \max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)} \right]^{\mathcal{L}(\Phi)} (\|x\|_2 + \mathcal{L}(\Phi)) \tag{4.14}$$

(cf. Definitions 2.2, 2.21, and 2.22).

Proof of Corollary 4.4. Throughout this proof let $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), (W_2, B_2), \dots, (W_L, B_L)) \in (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$, $\alpha = \max\{1, \|\mathcal{T}(\Phi)\|_\infty\}$, $x_0 \in \mathbb{R}^{l_0}$, $x_1 \in \mathbb{R}^{l_1}$ satisfy $x_1 = \mathfrak{R}(W_1 x_0 + B_1)$ (cf. Definitions 2.1, 2.21, and 2.22). Note that item (i) in Lemma 4.2 (applied with $m \curvearrowleft l_1$, $n \curvearrowleft l_0$, $A \curvearrowleft W_1$, $B \curvearrowleft B_1$, $x \curvearrowleft x_0$ in the notation of Lemma 4.2) ensures that

$$\|W_1 x_0 + B_1\|_\infty \le \alpha \sqrt{l_0} \|x_0\|_2 + \alpha \le \alpha \sqrt{l_0} (\|x_0\|_2 + 1). \tag{4.15}$$

In the following we distinguish between the case $\mathcal{L}(\Phi) = 1$ and the case $\mathcal{L}(\Phi) > 1$. We first prove (4.14) in the case $\mathcal{L}(\Phi) = 1$. Observe that (4.15) demonstrates that

$$\begin{aligned} \|(\mathcal{R}(\Phi))(x_0)\|_\infty &= \|W_1 x_0 + B_1\|_\infty \le \alpha \sqrt{l_0} (\|x_0\|_2 + 1) \\ &\le \frac{(l_0 + 1)\alpha}{2} (\|x_0\|_2 + 1) \le \frac{l_1 (l_0 + 1)\alpha}{2} (\|x_0\|_2 + 1) \\ &= \left[\frac{\mathcal{P}(\Phi) \max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)} \right]^{\mathcal{L}(\Phi)} (\|x_0\|_2 + \mathcal{L}(\Phi)). \end{aligned} \tag{4.16}$$

This proves (4.14) in case $\mathcal{L}(\Phi) = 1$. We now prove (4.14) in the case $\mathcal{L}(\Phi) > 1$. Note that (4.15) and the fact that for all $x \in \mathbb{R}$ it holds that $|\max\{x, 0\}| \le |x|$ show that

$$\|x_1\|_\infty = \|\mathfrak{R}(W_1 x_0 + B_1)\|_\infty \le \|W_1 x_0 + B_1\|_\infty \le \alpha \sqrt{l_0} (\|x_0\|_2 + 1). \tag{4.17}$$

This and item (ii) in Lemma 4.3 (applied with $L \curvearrowleft L$, $l_0 \curvearrowleft l_0$, $l_1 \curvearrowleft l_1$, …, $l_L \curvearrowleft l_L$, $\Phi \curvearrowleft \Phi$, $x_0 \curvearrowleft x_0$, $x_1 \curvearrowleft x_1$ in the notation of Lemma 4.3) ensure that

$$\begin{aligned} \|(\mathcal{R}(\Phi))(x_0)\|_\infty &\le l_{L-1} l_{L-2} \cdots l_1 \alpha^{L-1} (\|x_1\|_\infty + L - 1) \\ &\le l_{L-1} l_{L-2} \cdots l_1 \alpha^{L-1} (\alpha \sqrt{l_0} (\|x_0\|_2 + 1) + L - 1) \\ &\le l_{L-1} l_{L-2} \cdots l_1 \sqrt{l_0} \, \alpha^L (\|x_0\|_2 + L). \end{aligned} \tag{4.18}$$

In the next step observe that the inequality of arithmetic and geometric means assures that

$$\begin{aligned}\mathcal{P}(\Phi) &= \sum_{k=1}^{L} l_k(l_{k-1}+1) = l_1 + l_2 + \ldots + l_L + l_0 l_1 + l_1 l_2 + \ldots + l_{L-1} l_L \\ &\ge 2L[(l_1 l_2 \cdots l_L)(l_0 l_1 l_1 l_2 \cdots l_{L-1} l_L)]^{1/2L} = 2L\Big[l_0 (l_1)^3 (l_2)^3 \cdots (l_{L-1})^3 (l_L)^2\Big]^{1/2L} \\ &\ge 2L\Big[l_0 (l_1)^2 (l_2)^2 \cdots (l_{L-1})^2\Big]^{1/2L}.\end{aligned} \tag{4.19}$$

Hence, we obtain that

$$l_{L-1} l_{L-2} \cdots l_1 \sqrt{l_0} \le \left[\frac{\mathcal{P}(\Phi)}{2L}\right]^L. \tag{4.20}$$

Combining this and (4.18) shows that

$$\begin{aligned}\|(\mathcal{R}(\Phi))(x_0)\|_\infty &\le l_{L-1} l_{L-2} \cdots l_1 \sqrt{l_0}\, \alpha^L (\|x_0\|_2 + L) \\ &\le \left[\frac{\mathcal{P}(\Phi)\alpha}{2L}\right]^L (\|x_0\|_2 + L) \\ &= \left[\frac{\mathcal{P}(\Phi) \max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)}\right]^{\mathcal{L}(\Phi)} (\|x_0\|_2 + \mathcal{L}(\Phi)).\end{aligned} \tag{4.21}$$

This proves (4.14) in the case $\mathcal{L}(\Phi) > 1$. The proof of Corollary 4.4 is thus complete. □

Lemma 4.5. *Let $d \in \mathbb{N} \cap [4, \infty)$, $\beta, \sigma \in (0, \infty)$, $\Phi \in \mathbf{N}$ satisfy $\mathcal{I}(\Phi) = d$ and $\mathcal{O}(\Phi) = 1$ and let $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$ that $\varphi(x) = (\sigma/\pi)^{d/2} \exp(-\sigma \|x\|_2^2)$ (cf. Definitions 2.2 and 2.21). Then*

$$\begin{aligned}&\int_{\left\{y\in\mathbb{R}^d \colon \|y\|_2 \ge \frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}\right\}} |(\mathcal{R}(\Phi))(x)|^2 \varphi(x)\, dx \\ &\quad \le |\mathcal{L}(\Phi)|^2 \left[\frac{\mathcal{P}(\Phi) \max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)}\right]^{2\mathcal{L}(\Phi)} \left[\frac{1+\beta}{e^\beta}\right]^{d/3} \left[\frac{d^3(6+4\beta+\sigma)}{4\sigma}\right]\end{aligned} \tag{4.22}$$

(cf. Definition 2.22).

Proof of Lemma 4.5. Throughout this proof let $\mathscr{R} \in \mathbb{R}$ satisfy $\sqrt{2\sigma}\mathscr{R} = \sqrt{d(1+\beta)}$. Note that the fact that for all $\mathfrak{a}, \mathfrak{b} \in \mathbb{R}$ it holds that $(\mathfrak{a}+\mathfrak{b})^2 \le 2(\mathfrak{a}^2 + \mathfrak{b}^2)$ and Corollary 4.4 imply that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}|(\mathcal{R}(\Phi))(x)|^2 = \|(\mathcal{R}(\Phi))(x)\|_\infty^2 &\le \left[\frac{\mathcal{P}(\Phi) \max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)}\right]^{2\mathcal{L}(\Phi)} (\|x\|_2 + \mathcal{L}(\Phi))^2 \\ &\le 2\left[\frac{\mathcal{P}(\Phi) \max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)}\right]^{2\mathcal{L}(\Phi)} (\|x\|_2^2 + |\mathcal{L}(\Phi)|^2)\end{aligned} \tag{4.23}$$

(cf. Definition 2.22). Observe that Lemma 3.1 ensures that $\Gamma(d/2+1) = [d/2]\Gamma(d/2)$. Combining this with Lemma 3.9 (applied with $d \curvearrowleft d$, $\beta \curvearrowleft \beta$, $\sigma \curvearrowleft \sigma$, $k \curvearrowleft 0$, $k \curvearrowleft 2$ in the notation of Lemma 3.9) assures that

$$\begin{aligned}&\int_{\{y\in\mathbb{R}^d \colon \|y\|_2 \ge \mathscr{R}\}} \big[|\mathcal{L}(\Phi)|^2 + \|x\|_2^2\big]\varphi(x)\, dx \\ &= |\mathcal{L}(\Phi)|^2 \int_{\{y\in\mathbb{R}^d \colon \|y\|_2 \ge \mathscr{R}\}} \varphi(x)\, dx + \int_{\{y\in\mathbb{R}^d \colon \|y\|_2 \ge \mathscr{R}\}} \|x\|_2^2\, \varphi(x)\, dx\end{aligned}$$

$$
\begin{aligned}
&= |\mathcal{L}(\Phi)|^2 \int_{\{y\in\mathbb{R}^d :\, \|y\|_2 \geq \mathscr{R}\}} \left[\frac{\sigma}{\pi}\right]^{d/2} e^{-\sigma\|x\|_2^2}\, dx + \int_{\{y\in\mathbb{R}^d :\, \|y\|_2 \geq \mathscr{R}\}} \left[\frac{\sigma}{\pi}\right]^{d/2} \|x\|_2^2\, e^{-\sigma\|x\|_2^2}\, dx \\
&\leq |\mathcal{L}(\Phi)|^2 \left[d\left[\frac{1+\beta}{e^\beta}\right]^{d/2} + d e^{-\frac{d(1+\beta)}{2}} \right] \qquad (4.24)\\
&\quad + \left[\frac{d^3(1+\beta)}{2\sigma}\left[\frac{1+\beta}{e^\beta}\right]^{d/2} + \frac{\Gamma\left(\frac{d}{2}+1\right)}{\Gamma\left(\frac{d}{2}\right)} \left[\frac{d+2}{\sigma}\right] e^{-\frac{d^2(1+\beta)}{2(d+2)}} \right] \\
&= |\mathcal{L}(\Phi)|^2 \left[d\left[\frac{1+\beta}{e^\beta}\right]^{d/2} + d e^{-\frac{d(1+\beta)}{2}} \right] + \left[\frac{d^3(1+\beta)}{2\sigma}\left[\frac{1+\beta}{e^\beta}\right]^{d/2} + \left[\frac{d(d+2)}{2\sigma}\right] e^{-\frac{d^2(1+\beta)}{2(d+2)}} \right] \\
&\leq |\mathcal{L}(\Phi)|^2 \left[d\left[\frac{1+\beta}{e^\beta}\right]^{d/2} + d e^{-\frac{d(1+\beta)}{2}} + \left[\frac{d^3(1+\beta)}{2\sigma}\right]\left[\frac{1+\beta}{e^\beta}\right]^{d/2} + \left[\frac{d(d+2)}{2\sigma}\right] e^{-\frac{d^2(1+\beta)}{2(d+2)}} \right].
\end{aligned}
$$

This, the fact that $d^3(1+\beta) + d(d+2) + 4d\sigma \leq d^3\left(\frac{3}{2} + \beta + \frac{\sigma}{4}\right)$, and the fact that

$$
\max\left\{ e^{\frac{-d(1+\beta)}{2}}, e^{-\frac{d^2(1+\beta)}{2(d+2)}}, \left[\frac{1+\beta}{e^\beta}\right]^{d/2} \right\} \leq \left[\frac{1+\beta}{e^\beta}\right]^{d/3} \qquad (4.25)
$$

imply that

$$
\begin{aligned}
\int_{\{y\in\mathbb{R}^d :\, \|y\|_2 \geq \mathscr{R}\}} \left[|\mathcal{L}(\Phi)|^2 + \|x\|_2^2\right]\varphi(x)\, dx &\leq |\mathcal{L}(\Phi)|^2 \left[\frac{1+\beta}{e^\beta}\right]^{d/3} \left[\frac{d^3(1+\beta) + d(d+2) + 4d\sigma}{2\sigma}\right] \\
&\leq |\mathcal{L}(\Phi)|^2 \left[\frac{1+\beta}{e^\beta}\right]^{d/3} \left[\frac{d^3(6+4\beta+\sigma)}{8\sigma}\right].
\end{aligned} \qquad (4.26)
$$

Combining this with (4.23) demonstrates that

$$
\begin{aligned}
\int_{\left\{y\in\mathbb{R}^d :\, \|y\|_2 \geq \frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}\right\}} |(\mathcal{R}(\Phi))(x)|^2 \varphi(x)\, dx &= \int_{\{y\in\mathbb{R}^d :\, \|y\|_2 \geq \mathscr{R}\}} |(\mathcal{R}(\Phi))(x)|^2 \varphi(x)\, dx \\
&\leq 2\left[\frac{\mathcal{P}(\Phi)\max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)}\right]^{2\mathcal{L}(\Phi)} \left[\int_{\{y\in\mathbb{R}^d :\, \|y\|_2 \geq \mathscr{R}\}} \left[|\mathcal{L}(\Phi)|^2 + \|x\|_2^2\right]\varphi(x)\, dx \right] \qquad (4.27)\\
&\leq |\mathcal{L}(\Phi)|^2 \left[\frac{\mathcal{P}(\Phi)\max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)}\right]^{2\mathcal{L}(\Phi)} \left[\frac{1+\beta}{e^\beta}\right]^{d/3} \left[\frac{d^3(6+4\beta+\sigma)}{4\sigma}\right].
\end{aligned}
$$

The proof of Lemma 4.5 is thus complete. □

4.2. Upper bounds for scalar products involving realizations of ANNs

Lemma 4.6. *Let $d \in \mathbb{N} \cap [4, \infty)$, $\beta, \sigma \in (0, \infty)$, $\Phi \in \mathbf{N}$ satisfy $\mathcal{I}(\Phi) = d$ and $\mathcal{O}(\Phi) = 1$, let $\varphi \colon \mathbb{R}^d \to \mathbb{R}$, $\mathfrak{f} \colon \mathbb{R}^d \to \mathbb{R}$, and $\mathfrak{g} \colon \mathbb{R}^d \to \mathbb{R}$ be measurable, and assume for all $x \in \mathbb{R}^d$ that $\varphi(x) = (\sigma/\pi)^{d/2} \exp(-\sigma\|x\|_2^2)$, $\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(y)|\, dy > 0$, $\int_{\mathbb{R}^d} |\mathfrak{g}(y)|^2\, dy = 1$, and*

$$
\mathfrak{f}(x) = \left[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(y)|^2 \varphi(y)\, dy\right]^{-1/2} (\mathcal{R}(\Phi))(x)\, [\varphi(x)]^{1/2} \qquad (4.28)
$$

(cf. Definitions 2.2 and 2.21). Then

$$\int_{\mathbb{R}^d} |\mathfrak{f}(x)\mathfrak{g}(x)|\,dx \le \left[\int_{\left\{y\in\mathbb{R}^d\colon \|y\|_2 \le \frac{\sqrt{d(1+\beta)}}{\sqrt{2\sigma}}\right\}} |\mathfrak{g}(x)|^2\,dx\right]^{1/2} + \mathcal{L}(\Phi)\left[\frac{\mathcal{P}(\Phi)\max\{1,\|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)}\right]^{\mathcal{L}(\Phi)}\left[\frac{1+\beta}{e^\beta}\right]^{d/6}\left[\frac{d^{3/2}(6+4\beta+\sigma)^{1/2}}{2\sqrt{\sigma}\left[\int_{\mathbb{R}^d}|(\mathcal{R}(\Phi))(y)|^2\varphi(y)\,dy\right]^{1/2}}\right] \tag{4.29}$$

(cf. Definition 2.22).

Proof of Lemma 4.6. Throughout this proof let $\Gamma\colon(0,\infty)\to(0,\infty)$ satisfy for all $x\in(0,\infty)$ that $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}\,dt$, let $\mathfrak{a}\in\mathbb{R}$ satisfy $\mathfrak{a}=\left[\int_{\mathbb{R}^d}|(\mathcal{R}(\Phi))(y)|^2\varphi(y)\,dy\right]^{1/2}$, and let $\mathscr{R}\in\mathbb{R}$ satisfy $\sqrt{2\sigma}\mathscr{R}=\sqrt{d(1+\beta)}$. Note that $\mathfrak{a}\in(0,\infty)$ and

$$\int_{\mathbb{R}^d}|\mathfrak{f}(x)|^2\,dx = \left[\int_{\mathbb{R}^d}|(\mathcal{R}(\Phi))(y)|^2\varphi(y)\,dy\right]^{-1}\int_{\mathbb{R}^d}|(\mathcal{R}(\Phi))(x)|^2\varphi(x)\,dx = 1. \tag{4.30}$$

Combining this with the Hölder inequality shows that

$$\begin{aligned}
\int_{\mathbb{R}^d}|\mathfrak{f}(x)\mathfrak{g}(x)|\,dx &= \int_{\{y\in\mathbb{R}^d\colon \|y\|_2\le\mathscr{R}\}}|\mathfrak{f}(x)\mathfrak{g}(x)|\,dx + \int_{\{y\in\mathbb{R}^d\colon \|y\|_2\ge\mathscr{R}\}}|\mathfrak{f}(x)\mathfrak{g}(x)|\,dx\\
&\le \left[\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\le\mathscr{R}\}}|\mathfrak{f}(x)|^2\,dx\right]^{1/2}\left[\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\le\mathscr{R}\}}|\mathfrak{g}(x)|^2\,dx\right]^{1/2}\\
&\quad+\left[\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\ge\mathscr{R}\}}|\mathfrak{f}(x)|^2\,dx\right]^{1/2}\left[\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\ge\mathscr{R}\}}|\mathfrak{g}(x)|^2\,dx\right]^{1/2}\\
&\le \left[\int_{\mathbb{R}^d}|\mathfrak{f}(x)|^2\,dx\right]^{1/2}\left[\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\le\mathscr{R}\}}|\mathfrak{g}(x)|^2\,dx\right]^{1/2}\\
&\quad+\left[\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\ge\mathscr{R}\}}|\mathfrak{f}(x)|^2\,dx\right]^{1/2}\left[\int_{\mathbb{R}^d}|\mathfrak{g}(x)|^2\,dx\right]^{1/2}\\
&= \left[\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\le\mathscr{R}\}}|\mathfrak{g}(x)|^2\,dx\right]^{1/2} + \left[\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\ge\mathscr{R}\}}|\mathfrak{f}(x)|^2\,dx\right]^{1/2}.
\end{aligned} \tag{4.31}$$

Next we obtain that Lemma 4.5 (applied with $d \curvearrowleft d$, $\beta\curvearrowleft\beta$, $\sigma\curvearrowleft\sigma$, $\Phi\curvearrowleft\Phi$, $\varphi\curvearrowleft\varphi$ in the notation of Lemma 4.5) implies that

$$
\begin{aligned}
\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\geq\mathscr{R}\}} |\mathfrak{f}(x)|^2\,dx &= \mathfrak{a}^{-2}\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\geq\mathscr{R}\}} |(\mathcal{R}(\Phi))(x)|^2\varphi(x)\,dx \\
&\leq \mathfrak{a}^{-2}|\mathcal{L}(\Phi)|^2\left[\frac{\mathcal{P}(\Phi)\max\{1,\|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)}\right]^{2\mathcal{L}(\Phi)}\left[\frac{1+\beta}{e^\beta}\right]^{d/3}\left[\frac{d^3(6+4\beta+\sigma)}{4\sigma}\right].
\end{aligned}
\tag{4.32}
$$

This and (4.31) imply that

$$
\begin{aligned}
\int_{\mathbb{R}^d}|\mathfrak{f}(x)\mathfrak{g}(x)|\,dx &\leq \left[\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\leq\mathscr{R}\}}|\mathfrak{g}(x)|^2\,dx\right]^{1/2} + \left[\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\geq\mathscr{R}\}}|f(x)|^2\,dx\right]^{1/2} \\
&\leq \left[\int_{\{y\in\mathbb{R}^d\colon \|y\|_2\leq\mathscr{R}\}}|\mathfrak{g}(x)|^2\,dx\right]^{1/2} \\
&\quad + \mathfrak{a}^{-1}\mathcal{L}(\Phi)\left[\frac{\mathcal{P}(\Phi)\max\{1,\|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)}\right]^{\mathcal{L}(\Phi)}\left[\frac{1+\beta}{e^\beta}\right]^{d/6}\left[\frac{d^{3/2}(6+4\beta+\sigma)^{1/2}}{2\sqrt{\sigma}}\right].
\end{aligned}
\tag{4.33}
$$

The proof of Lemma 4.6 is thus complete. □

4.3. *On the connection of distances and scalar products*

Lemma 4.7. *Let $d\in\mathbb{N}$, $\alpha\in\mathbb{R}$, let $\mathfrak{f}\colon\mathbb{R}^d\to\mathbb{R}$ and $\mathfrak{g}\colon\mathbb{R}^d\to\mathbb{R}$ be measurable, and assume $\int_{\mathbb{R}^d}|\mathfrak{f}(x)|^2\,dx = \int_{\mathbb{R}^d}|\mathfrak{g}(x)|^2\,dx = 1$. Then*

$$
\int_{\mathbb{R}^d}|\alpha\mathfrak{f}(x)-\mathfrak{g}(x)|^2\,dx \geq 1-\int_{\mathbb{R}^d}|\mathfrak{f}(x)\mathfrak{g}(x)|\,dx. \tag{4.34}
$$

Proof of Lemma 4.7. Observe that the Hölder inequality implies that

$$
\int_{\mathbb{R}^d}|\mathfrak{f}(x)\mathfrak{g}(x)|\,dx \leq \left[\int_{\mathbb{R}^d}|\mathfrak{f}(x)|^2\,dx\right]^{1/2}\left[\int_{\mathbb{R}^d}|\mathfrak{g}(x)|^2\,dx\right]^{1/2} = 1. \tag{4.35}
$$

Next note that

$$
\begin{aligned}
\int_{\mathbb{R}^d}|\alpha\mathfrak{f}(x)-\mathfrak{g}(x)|^2\,dx &= \alpha^2+1-2\alpha\int_{\mathbb{R}^d}\mathfrak{f}(x)\mathfrak{g}(x)\,dx \\
&= \left[\alpha-\int_{\mathbb{R}^d}\mathfrak{f}(x)\mathfrak{g}(x)\,dx\right]^2 + 1 - \left[\int_{\mathbb{R}^d}\mathfrak{f}(x)\mathfrak{g}(x)\,dx\right]^2 \\
&\geq 1-\left[\int_{\mathbb{R}^d}\mathfrak{f}(x)\mathfrak{g}(x)\,dx\right]^2 \geq 1-\left[\int_{\mathbb{R}^d}|\mathfrak{f}(x)\mathfrak{g}(x)|\,dx\right]^2.
\end{aligned}
\tag{4.36}
$$

This and (4.35) ensure that

$$
\int_{\mathbb{R}^d}|\alpha\mathfrak{f}(x)-\mathfrak{g}(x)|^2\,dx \geq 1-\left[\int_{\mathbb{R}^d}|\mathfrak{f}(x)\mathfrak{g}(x)|\,dx\right]^2 \geq 1-\int_{\mathbb{R}^d}|\mathfrak{f}(x)\mathfrak{g}(x)|\,dx. \tag{4.37}
$$

The proof of Lemma 4.7 is thus complete. □

4.4. ANN approximations for a class of general high-dimensional functions

Proof of Theorem 4.1. Throughout this proof let $\mathbf{f}\colon \mathbb{R}^d \to \mathbb{R}$ and $\mathbf{g}\colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$ that $\mathbf{g}(x) = \mathfrak{g}(x)[\varphi(x)]^{1/2}$ and $\mathbf{f}(x) = \big[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(y)|^2 \varphi(y)\,dy\big]^{-1/2} (\mathcal{R}(\Phi))(x)[\varphi(x)]^{1/2}$ and let $\mathscr{R} \in \mathbb{R}$ satisfy $\sqrt{2\sigma}\mathscr{R} = \sqrt{d(1+\beta)}$ (cf. Definition 2.22). Observe that $\int_{\mathbb{R}^d} |\mathbf{f}(x)|^2\,dx = \int_{\mathbb{R}^d} |\mathbf{g}(x)|^2\,dx = 1$. Lemma 4.7 (applied with $d \curvearrowleft d$, $\alpha \curvearrowleft \big[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(y)|^2 \varphi(y)\,dy\big]^{1/2}$, $\mathfrak{f} \curvearrowleft \mathbf{f}$, $\mathfrak{g} \curvearrowleft \mathbf{g}$ in the notation of Lemma 4.7) hence ensures that

$$\begin{aligned} \int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\,dx &= \int_{\mathbb{R}^d} \left| \mathbf{f}(x) \left[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(y)|^2 \varphi(y)\,dy \right]^{1/2} - \mathbf{g}(x) \right|^2 dx \\ &\geq 1 - \int_{\mathbb{R}^d} |\mathbf{f}(x)\mathbf{g}(x)|\,dx. \end{aligned} \tag{4.38}$$

Combining this with Lemma 4.6 (applied with $d \curvearrowleft d$, $\beta \curvearrowleft \beta$, $\sigma \curvearrowleft \sigma$, $\Phi \curvearrowleft \Phi$, $\varphi \curvearrowleft \varphi$, $\mathfrak{f} \curvearrowleft \mathbf{f}$, $\mathfrak{g} \curvearrowleft \mathbf{g}$ in the notation of Lemma 4.6) demonstrates that

$$\begin{aligned} &\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\,dx \geq 1 - \int_{\mathbb{R}^d} |\mathbf{f}(x)\mathbf{g}(x)|\,dx \geq 1 - \left[\int_{\{y \in \mathbb{R}^d \colon \|y\|_2 \leq \mathscr{R}\}} |\mathbf{g}(x)|^2\,dx \right]^{1/2} \\ &\quad - \mathcal{L}(\Phi) \left[\frac{\mathcal{P}(\Phi) \max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)} \right]^{\mathcal{L}(\Phi)} \left[\frac{1+\beta}{e^\beta} \right]^{d/6} \left[\frac{d^{3/2}(6+4\beta+\sigma)^{1/2}}{2\sqrt{\sigma}\,\big[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(y)|^2 \varphi(y)\,dy\big]^{1/2}} \right] \\ &= 1 - \left[\int_{\{y \in \mathbb{R}^d \colon \|y\|_2 \leq \mathscr{R}\}} |\mathfrak{g}(x)|^2 \varphi(x)\,dx \right]^{1/2} \\ &\quad - \mathcal{L}(\Phi) \left[\frac{\mathcal{P}(\Phi) \max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)} \right]^{\mathcal{L}(\Phi)} \left[\frac{1+\beta}{e^\beta} \right]^{d/6} \left[\frac{d^{3/2}(6+4\beta+\sigma)^{1/2}}{2\sqrt{\sigma}\,\big[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(y)|^2 \varphi(y)\,dy\big]^{1/2}} \right]. \end{aligned} \tag{4.39}$$

This implies that

$$\begin{aligned} \mathcal{L}(\Phi) \left[\frac{\mathcal{P}(\Phi) \max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)} \right]^{\mathcal{L}(\Phi)} &\geq \left[\frac{e^\beta}{1+\beta} \right]^{d/6} \left[\frac{2\sqrt{\sigma}\,\big[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x)|^2 \varphi(x)\,dx\big]^{1/2}}{d^{3/2}(6+4\beta+\sigma)^{1/2}} \right] \\ &\quad \cdot \left[1 - \left[\int_{\{y \in \mathbb{R}^d \colon \|y\|_2 \leq \mathscr{R}\}} |\mathfrak{g}(x)|^2 \varphi(x)\,dx \right]^{1/2} - \int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\,dx \right]. \end{aligned} \tag{4.40}$$

The proof of Theorem 4.1 is thus complete. □

4.5. ANN approximations for certain specific high-dimensional functions

Corollary 4.8. *Let $d \in \mathbb{N} \cap [4, \infty)$, $\varepsilon \in (0, 1/4]$, let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ and $g\colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ that $\varphi(x) = (2\pi)^{-d/2} \exp(-\frac{1}{2}\|x\|_2^2)$ and $g(x) = \sum_{j=1}^d [\max\{|x_j| - \sqrt{2d}, 0\}]^2$, let*

$\mathfrak{g}\colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$ that $\mathfrak{g}(x) = [\int_{\mathbb{R}^d} |g(y)|^2 \varphi(y)\,dy]^{-1/2} g(x)$, and let $\Phi \in \mathbf{N}$ satisfy $\mathcal{I}(\Phi) = d$, $\mathcal{O}(\Phi) = 1$, and $\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\,dx \le \varepsilon$ (cf. Definitions 2.2 and 2.21). Then

$$\mathcal{P}(\Phi)\max\{1, \|\mathcal{T}(\Phi)\|_\infty\} \ge \left[\tfrac{2}{7}\right] d^{-3/2} \exp\!\left(\tfrac{d}{20\mathcal{L}(\Phi)}\right) \tag{4.41}$$

(cf. Definition 2.22).

Proof of Corollary 4.8. Note that the triangle inequality ensures that

$$\begin{aligned}
&\left[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x)|^2 \varphi(x)\,dx\right]^{1/2} \\
&\ge \left[\int_{\mathbb{R}^d} |\mathfrak{g}(x)|^2 \varphi(x)\,dx\right]^{1/2} - \left[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\,dx\right]^{1/2} \\
&= 1 - \left[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\,dx\right]^{1/2} \ge 1 - \varepsilon^{1/2} \\
&\ge 1 - 4^{-1/2} = \frac{1}{2} > 0.
\end{aligned} \tag{4.42}$$

Hence, we obtain that $\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x)|\,dx > 0$. Next observe that for all $x = (x_1, x_2, \ldots, x_d) \in \{y \in \mathbb{R}^d \colon \|y\|_2 \le \sqrt{2d}\}$, $j \in \{1, 2, \ldots, d\}$ it holds that $|x_j| \le \|x\|_2 \le \sqrt{2d}$. This ensures that for all $x = (x_1, x_2, \ldots, x_d) \in \{y \in \mathbb{R}^d \colon \|y\|_2 \le \sqrt{2d}\}$ it holds that $\mathfrak{g}(x) = g(x) = 0$. Combining Theorem 4.1 (applied with $d \curvearrowleft d$, $\beta \curvearrowleft 1$, $\sigma \curvearrowleft 1/2$, $\Phi \curvearrowleft \Phi$, $\varphi \curvearrowleft \varphi$, $g \curvearrowleft g$, $\mathfrak{g} \curvearrowleft \mathfrak{g}$ in the notation of Theorem 4.1), the fact that $e/2 \ge e^{3/10}$, the fact that $\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x)|\,dx > 0$, and (4.42) therefore implies that

$$\begin{aligned}
&\mathcal{L}(\Phi)\left[\frac{\mathcal{P}(\Phi)\max\{1, \|\mathcal{T}(\Phi)\|_\infty\}}{2\mathcal{L}(\Phi)}\right]^{\mathcal{L}(\Phi)} \ge \left[\frac{e}{2}\right]^{d/6} \left[\frac{\sqrt{2}\,[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x)|^2 \varphi(x)\,dx]^{1/2}}{d^{3/2}(6 + 4 + 1/2)^{1/2}}\right] \\
&\quad \cdot \left[1 - \left[\int_{\{y \in \mathbb{R}^d \colon \|y\|_2 \le \sqrt{2d}\}} |\mathfrak{g}(x)|^2 \varphi(x)\,dx\right]^{1/2} - \int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\,dx\right] \\
&= \left[\frac{e}{2}\right]^{d/6} \left[\frac{\sqrt{2}\,[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x)|^2 \varphi(x)\,dx]^{1/2}}{d^{3/2}(6 + 4 + 1/2)^{1/2}}\right] \left[1 - \int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\,dx\right] \\
&\ge [(21)^{-1/2}]\left[\frac{e}{2}\right]^{d/6} d^{-3/2}(1 - \varepsilon) \ge \frac{e^{d/20}}{7d^{3/2}}
\end{aligned} \tag{4.43}$$

(cf. Definition 2.22). Hence, we obtain that

$$\mathcal{P}(\Phi)\max\{1, \|\mathcal{T}(\Phi)\|_\infty\} \ge 2\mathcal{L}(\Phi)\left[\frac{e^{d/20}}{7d^{3/2}\mathcal{L}(\Phi)}\right]^{1/\mathcal{L}(\Phi)} \ge \left[\frac{2}{7}\right] d^{-3/2} \exp\!\left(\frac{d}{20\mathcal{L}(\Phi)}\right). \tag{4.44}$$

The proof of Corollary 4.8 is thus complete. □

Corollary 4.9. *Let $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and $g_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $\varphi_d(x) = (2\pi)^{-d/2} \exp(-\frac{1}{2}(\sum_{j=1}^d |x_j|^2))$ and $g_d(x) = \sum_{j=1}^d [\max\{|x_j| - \sqrt{2d}, 0\}]^2$, let $\mathfrak{g}_d \colon \mathbb{R}^d \to \mathbb{R}$,*

$d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $\mathfrak{g}_d(x) = [\int_{\mathbb{R}^d} |g_d(y)|^2 \varphi_d(y)\, dy]^{-1/2} g_d(x)$, and let $\delta \in (0,1]$, $\mathfrak{C} \in [100(\delta \ln(1.03))^{-2}, \infty)$ satisfy $2\mathfrak{C}^{5/\delta} \le (1.03)^{\sqrt{\mathfrak{C}}}$. Then it holds for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1/2]$, $\Phi \in \mathbf{N}$ with $\mathcal{I}(\Phi) = d$, $\mathcal{O}(\Phi) = 1$, $\mathcal{H}(\Phi) \le \mathfrak{c} d^{1-\delta}$, $\|\mathcal{T}(\Phi)\|_\infty \le \mathfrak{c} d^{\mathfrak{c}}$, and $\big[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}_d(x)|^2 \varphi_d(x)\, dx\big]^{1/2} \le \varepsilon$ that $\mathcal{P}(\Phi) \ge (1+\mathfrak{c}^{-3})^{(d^\delta)}$ (cf. Definitions 2.2, 2.21, and 2.22).

Proof of Corollary 4.9. Note that the assumption that $\mathfrak{C} \in [100(\delta \ln(1.03))^{-2}, \infty)$ and the chain rule ensure that for all $x \in [\mathfrak{C}, \infty)$ it holds that

$$\begin{aligned}
\big[2^{-1}(1.03)^{\sqrt{x}} x^{-5/\delta}\big]' &= (1.03)^{\sqrt{x}} \ln(1.03) \left[\frac{1}{4\sqrt{x}}\right] x^{-5/\delta} - (1.03)^{\sqrt{x}} \left[\frac{5}{2\delta x}\right] x^{-5/\delta} \\
&= (1.03)^{\sqrt{x}} \left[\frac{x^{-5/\delta}}{4x}\right] \ln(1.03) \big[\sqrt{x} - 10(\delta \ln(1.03))^{-1}\big] \\
&\ge (1.03)^{\sqrt{x}} \left[\frac{x^{-5/\delta}}{4x}\right] \ln(1.03) \big[\sqrt{\mathfrak{C}} - 10(\delta \ln(1.03))^{-1}\big] \ge 0.
\end{aligned} \tag{4.45}$$

This implies that the function $[\mathfrak{C}, \infty) \ni x \mapsto 2^{-1}(1.03)^{\sqrt{x}} x^{-5/\delta} \in \mathbb{R}$ is non-decreasing. The assumption that $\mathfrak{C} \in [100(\delta \ln(1.03))^{-2}, \infty)$ and the assumption that $2\mathfrak{C}^{5/\delta} \le (1.03)^{\sqrt{\mathfrak{C}}}$ therefore assure that for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$ it holds that $\mathfrak{c} \ge 100(\delta \ln(1.03))^{-2}$ and

$$2^{-1}(1.03)^{\sqrt{\mathfrak{c}}} \mathfrak{c}^{-5/\delta} \ge 2^{-1}(1.03)^{\sqrt{\mathfrak{C}}} \mathfrak{C}^{-5/\delta} \ge 1. \tag{4.46}$$

The fact that for all $x \in (0, \infty)$ it holds that $(1 + x^{-1})^x \le e$ hence ensures that for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $d \in \mathbb{N}$, $\Phi \in \mathbf{N}$ with $d \le \mathfrak{c}^{5/(2\delta)}$ it holds that

$$(1+\mathfrak{c}^{-3})^{(d^\delta)} \le (1+\mathfrak{c}^{-3})^{(\mathfrak{c}^{5/2})} = \big[(1+\mathfrak{c}^{-3})^{(\mathfrak{c}^3)}\big]^{\frac{1}{\sqrt{\mathfrak{c}}}} \le e^{\frac{1}{\sqrt{\mathfrak{c}}}} \le 2 \le \mathcal{P}(\Phi) \tag{4.47}$$

(cf. Definition 2.2). Moreover, observe that the chain rule and (4.46) show that for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $x \in [\mathfrak{c}^{5/(2\delta)}, \infty)$ it holds that

$$\begin{aligned}
\big[(1.03)^{(x^\delta)/\mathfrak{c}} x^{-2\mathfrak{c}}\big]' &= (1.03)^{(x^\delta)/\mathfrak{c}} \ln(1.03) \left[\frac{\delta}{\mathfrak{c}}\right] x^{-2\mathfrak{c}-1+\delta} - 2\mathfrak{c}(1.03)^{(x^\delta)/\mathfrak{c}} x^{-2\mathfrak{c}-1} \\
&= (1.03)^{(x^\delta)/\mathfrak{c}} x^{-2\mathfrak{c}-1} \left[\frac{\delta}{\mathfrak{c}}\right] \ln(1.03) [x^\delta - 2\mathfrak{c}^2(\delta \ln(1.03))^{-1}] \\
&\ge (1.03)^{(x^\delta)/\mathfrak{c}} x^{-2\mathfrak{c}-1} \left[\frac{\delta}{\mathfrak{c}}\right] \ln(1.03) [\mathfrak{c}^{5/2} - 2\mathfrak{c}^2(\delta \ln(1.03))^{-1}] \\
&= (1.03)^{(x^\delta)/\mathfrak{c}} x^{-2\mathfrak{c}-1} \delta \mathfrak{c} \ln(1.03) [\sqrt{\mathfrak{c}} - 2(\delta \ln(1.03))^{-1}] \\
&\ge (1.03)^{(x^\delta)/\mathfrak{c}} x^{-2\mathfrak{c}-1} 8\mathfrak{c} > 0.
\end{aligned} \tag{4.48}$$

This implies for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$ that the function $[\mathfrak{c}^{5/(2\delta)}, \infty) \ni x \mapsto (1.03)^{(x^\delta)/\mathfrak{c}} x^{-2\mathfrak{c}} \in \mathbb{R}$ is strictly increasing. The fact that $e^{1/30} \ge 1.03$, (4.46), and the fact that for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$ it holds that $2^{\mathfrak{c}+1} \ge 7\mathfrak{c}$ therefore demonstrate that for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $d \in \mathbb{N}$ with $d \ge \mathfrak{c}^{5/(2\delta)}$ it holds that

$$e^{(d^\delta)/(30\mathfrak{c})} d^{-2\mathfrak{c}} \ge (1.03)^{(d^\delta)/\mathfrak{c}} d^{-2\mathfrak{c}} \ge (1.03)^{(\mathfrak{c}^{3/2})} \mathfrak{c}^{-5\mathfrak{c}/\delta} = \big[(1.03)^{\sqrt{\mathfrak{c}}} \mathfrak{c}^{-5/\delta}\big]^{\mathfrak{c}} \ge 2^{\mathfrak{c}} \ge \left[\frac{7}{2}\right] \mathfrak{c}. \tag{4.49}$$

The fact that for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$ it holds that $(25\mathfrak{c})^{-1} \ge (30\mathfrak{c})^{-1} + \mathfrak{c}^{-3}$, the fact that for all $x \in \mathbb{R}$ it holds that $e^x \ge 1 + x$, and Corollary 4.8 hence ensure that for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1/2]$, $\Phi \in \mathbf{N}$ with $d \ge \mathfrak{c}^{5/(2\delta)}$, $\mathcal{I}(\Phi) = d$, $\mathcal{O}(\Phi) = 1$, $\mathcal{H}(\Phi) \le \mathfrak{c} d^{1-\delta}$, $\|\mathcal{T}(\Phi)\|_\infty \le \mathfrak{c} d^{\mathfrak{c}}$, and $\big[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}_d(x)|^2 \varphi_d(x)\, dx\big]^{1/2} \le \varepsilon$ it holds that

$$\begin{aligned}\mathcal{P}(\Phi) &\geq (\max\{1, \|\mathcal{T}(\Phi)\|_\infty\})^{-1}\left[\tfrac{2}{7}\right]d^{-3/2}\exp\left(\tfrac{d}{20\mathcal{L}(\Phi)}\right) \geq \left[\tfrac{2}{7}\right]\exp\left(\tfrac{d^\delta}{25\mathfrak{c}}\right)d^{-2\mathfrak{c}}\mathfrak{c}^{-1} \\ &\geq \left[\tfrac{2}{7}\right]\exp\left(\tfrac{d^\delta}{30\mathfrak{c}}\right)d^{-2\mathfrak{c}}\mathfrak{c}^{-1}\exp\left(\tfrac{d^\delta}{\mathfrak{c}^3}\right) \geq \exp\left(\tfrac{d^\delta}{\mathfrak{c}^3}\right) \geq (1+\mathfrak{c}^{-3})^{(d^\delta)}\end{aligned} \tag{4.50}$$

(cf. Definitions 2.21 and 2.22). Combining this with (4.47) assures that for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1/2]$, $\Phi \in \mathbf{N}$ with $\mathcal{I}(\Phi) = d$, $\mathcal{O}(\Phi) = 1$, $\mathcal{H}(\Phi) \leq \mathfrak{c}d^{1-\delta}$, $\|\mathcal{T}(\Phi)\|_\infty \leq \mathfrak{c}d^{\mathfrak{c}}$, and $\left[\int_{\mathbb{R}^d}|(\mathcal{R}(\Phi))(x) - \mathfrak{g}_d(x)|^2\varphi_d(x)\,dx\right]^{1/2} \leq \varepsilon$ it holds that $\mathcal{P}(\Phi) \geq (1+\mathfrak{c}^{-3})^{(d^\delta)}$. The proof of Corollary 4.9 is thus complete. □

5. Upper bounds for the number of ANN parameters in the approximation of high-dimensional functions

Recall the sequence of functions $\mathfrak{g}_d\colon \mathbb{R}^d \to \mathbb{R}$ satisfying Corollary 4.9, i.e., item (ii) of Theorem 1.1: Let $g_d\colon \mathbb{R}^d \to \mathbb{R}$ be defined by $g_d(x) = \sum_{j=1}^d [\max\{|x_j| - \sqrt{2d}, 0\}]^2$, then we set $\mathfrak{g}_d(x) = [\int_{\mathbb{R}^d}|g_d(y)|^2\varphi(y)\,dy]^{-1/2}g_d(x)$.

The goal of this section is to approximate $\mathfrak{g}_d$ by a suitably deep ANN satisfying the conditions in item (i) of Theorem 1.1. This approximation is done in Theorem 5.11 employing (i) the elementary ANN approximation result for shifted squared rectifier functions established throughout Subsections 5.1, 5.2, and 5.3, (ii) the lower and upper bounds for Gaussian integrals presented in Subsection 5.4, and (iii) the elementary ANN representation result for multiplications with powers of real numbers established in Subsection 5.5. We conclude this section with Corollary 5.12, which establishes item (i) of Theorem 1.1.

5.1. ANN approximations for the square function

The approximations of the square function presented in Lemma 5.1 and Lemma 5.2 are based on the well-known result in Yarotsky [59, Proposition 2]. In the current form Lemma 5.1 and Lemma 5.2 are slight extensions of, e.g., [17, Lemma 6.1] and [26, Proposition 3.3].

Lemma 5.1. *Let $(A_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{4\times4}$, $\mathbb{B} \in \mathbb{R}^{4\times1}$, $(c_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$ satisfy for all $k \in \mathbb{N}$ that*

$$A_k = \begin{pmatrix} 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ -c_k & 2c_k & -c_k & 1 \end{pmatrix}, \qquad \mathbb{B} = \begin{pmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{pmatrix}, \qquad \text{and} \qquad c_k = 2^{1-2k}, \tag{5.1}$$

let $g_n\colon \mathbb{R} \to [0,1]$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$, $x \in \mathbb{R}$ that

$$g_1(x) = \begin{cases} 2x & : x \in [0, \frac{1}{2}) \\ 2-2x & : x \in [\frac{1}{2}, 1] \\ 0 & : x \in \mathbb{R}\backslash[0,1] \end{cases} \tag{5.2}$$

and $g_{n+1}(x) = g_1(g_n(x))$, let $f_n\colon [0,1] \to [0,1]$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$, $k \in \{0, 1, \ldots, 2^n - 1\}$, $x \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)$ that $f_n(1) = 1$ and

$$f_n(x) = \left[\tfrac{2k+1}{2^n}\right]x - \tfrac{(k^2+k)}{2^{2n}}, \tag{5.3}$$

and let $r_k = (r_{k,1}, r_{k,2}, r_{k,3}, r_{k,4})\colon \mathbb{R} \to \mathbb{R}^4$, $k \in \mathbb{N}$, satisfy for all $k \in \mathbb{N}$, $x \in \mathbb{R}$ that $r_1(x) = \mathfrak{R}(x, x-\frac{1}{2}, x-1, x)$ and $r_{k+1}(x) = \mathfrak{R}(A_k r_k(x) + \mathbb{B})$ (cf. Definition 2.1). Then

(i) it holds for all $k \in \mathbb{N}$, $x \in \mathbb{R}$ that

$$2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) = g_k(x) \tag{5.4}$$

and

(ii) it holds for all $k \in \mathbb{N}$, $x \in \mathbb{R}$ that

$$r_{k,4}(x) = \begin{cases} f_{k-1}(x) & : x \in [0,1] \\ \max\{x,0\} & : x \in \mathbb{R}\backslash[0,1]. \end{cases} \tag{5.5}$$

Proof of Lemma 5.1. We prove (5.4) and (5.5) by induction on $k \in \mathbb{N}$. Note that (5.2) and the assumption that for all $x \in \mathbb{R}$ it holds that $r_1(x) = \mathfrak{R}\big(x, x - \frac{1}{2}, x-1, x\big)$ show that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} 2r_{1,1}(x) - 4r_{1,2}(x) + 2r_{1,3}(x) &= 2\mathfrak{R}(x) - 4\mathfrak{R}(x - \tfrac{1}{2}) + 2\mathfrak{R}(x-1) \\ &= 2\max\{x,0\} - 4\max\{x - \tfrac{1}{2}, 0\} + 2\max\{x-1,0\} = g_1(x). \end{aligned} \tag{5.6}$$

Furthermore, observe that the assumption that for all $x \in \mathbb{R}$ it holds that $r_1(x) = \mathfrak{R}(x, x - \frac{1}{2}, x-1, x)$ and the fact that for all $x \in [0,1]$ it holds that $f_0(x) = x = \max\{x,0\}$ imply that for all $x \in \mathbb{R}$ it holds that

$$r_{1,4}(x) = \max\{x,0\} = \begin{cases} f_0(x) & : x \in [0,1] \\ \max\{x,0\} & : x \in \mathbb{R}\backslash[0,1]. \end{cases} \tag{5.7}$$

Combining this with (5.6) proves (5.4) and (5.5) in the base case $k=1$. For the induction step let $k \in \mathbb{N}$ satisfy for all $x \in \mathbb{R}$ that

$$2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) = g_k(x) \tag{5.8}$$

and

$$r_{k,4}(x) = \begin{cases} f_{k-1}(x) & : x \in [0,1] \\ \max\{x,0\} & : x \in \mathbb{R}\backslash[0,1]. \end{cases} \tag{5.9}$$

Note that (5.1), (5.6), (5.8), and the assumption that for all $n \in \mathbb{N}$, $x \in \mathbb{R}$ it holds that $r_{n+1}(x) = \mathfrak{R}(A_n r_n(x) + \mathbb{B})$ ensure that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} g_{k+1}(x) &= g_1(g_k(x)) = g_1(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x)) \\ &= 2\mathfrak{R}\big(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x)\big) \\ &\quad - 4\mathfrak{R}\big(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) - \tfrac{1}{2}\big) \\ &\quad + 2\mathfrak{R}\big(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) - 1\big) \\ &= 2r_{k+1,1}(x) - 4r_{k+1,2}(x) + 2r_{k+1,3}(x). \end{aligned} \tag{5.10}$$

In addition, observe that (5.1), (5.8), and the assumption that for all $n \in \mathbb{N}$, $x \in \mathbb{R}$ it holds that $r_{n+1}(x) = \mathfrak{R}(A_n r_n(x) + \mathbb{B})$ demonstrate that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} r_{k+1,4}(x) &= \mathfrak{R}(-c_k r_{k,1}(x) + 2c_k r_{k,2}(x) - c_k r_{k,3}(x) + r_{k,4}(x)) \\ &= \mathfrak{R}(-[2^{1-2k}]r_{k,1}(x) + [2^{2-2k}]r_{k,2}(x) - [2^{1-2k}]r_{k,3}(x) + r_{k,4}(x)) \\ &= \mathfrak{R}(-[2^{-2k}][2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x)] + r_{k,4}(x)) \\ &= \mathfrak{R}(-[2^{-2k}]g_k(x) + r_{k,4}(x)). \end{aligned} \tag{5.11}$$

Combining this with (5.9), [26, Lemma 3.2], and the fact that for all $x \in [0,1]$ it holds that $f_k(x) \geq 0$ shows that for all $x \in [0,1]$ it holds that

$$\begin{aligned} r_{k+1,4}(x) &= \mathfrak{R}(-[2^{-2k}]g_k(x) + r_{k,4}(x)) = \mathfrak{R}(-[2^{-2k}g_k(x)] + f_{k-1}(x)) \\ &= \mathfrak{R}\big(-[2^{-2k}g_k(x)] + x - \big[\textstyle\sum_{j=1}^{k-1}[2^{-2j}g_j(x)]\big]\big) \\ &= \mathfrak{R}\big(x - \big[\textstyle\sum_{j=1}^{k}[2^{-2j}g_j(x)]\big]\big) = \mathfrak{R}(f_k(x)) = f_k(x). \end{aligned} \tag{5.12}$$

Next note that (5.9), (5.11), and the fact that for all $x \in \mathbb{R}\backslash[0,1]$ it holds that $g_k(x) = 0$ prove that for all $x \in \mathbb{R}\backslash[0,1]$ it holds that

$$r_{k+1,4}(x) = \mathfrak{R}(-[2^{-2k}]g_k(x) + r_{k,4}(x)) = \mathfrak{R}(r_{k,4}(x)) = \mathfrak{R}(\max\{x,0\}) = \max\{x,0\}. \tag{5.13}$$

Combining (5.10) and (5.12) hence proves (5.4) and (5.5) in the case $k+1$. Induction thus establishes items (i) and (ii). The proof of Lemma 5.1 is thus complete. □

Lemma 5.2. *Let $M \in \mathbb{N}$, $(A_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{4\times 4}$, $\mathbb{A}, \mathbb{B} \in \mathbb{R}^{4\times 1}$, $(C_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{1\times 4}$, $(c_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$ satisfy for all $k \in \mathbb{N}$ that*

$$A_k = \begin{pmatrix} 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ -c_k & 2c_k & -c_k & 1 \end{pmatrix}, \quad \mathbb{A} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbb{B} = \begin{pmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{pmatrix}, \quad C_k = \begin{pmatrix} -c_k & 2c_k & -c_k & 1 \end{pmatrix}, \tag{5.14}$$

and $c_k = 2^{1-2k}$ and let $\Phi \in \mathbf{N}$ satisfy

$$\Phi = \begin{cases} ((\mathbb{A},\mathbb{B}),(C_1,0)) & : M = 1 \\ ((\mathbb{A},\mathbb{B}),(A_1,\mathbb{B}),(A_2,\mathbb{B}),\ldots,(A_{M-1},\mathbb{B}),(C_M,0)) & : M > 1 \end{cases} \tag{5.15}$$

(cf. Definition 2.2). Then

(i) it holds that $\mathcal{R}(\Phi) \in C(\mathbb{R},\mathbb{R})$,
(ii) it holds for all $x \in [0,1]$ that $|x^2 - (\mathcal{R}(\Phi))(x)| \le 4^{-M-1}$,
(iii) it holds for all $x \in \mathbb{R}\backslash[0,1]$ that $(\mathcal{R}(\Phi))(x) = \mathfrak{R}(x)$,
(iv) it holds that $\mathcal{D}(\Phi) = (1,4,4,\ldots,4,1) \in \mathbb{N}^{M+2}$,
(v) it holds that $\|\mathcal{T}(\Phi)\|_\infty \le 4$,
(vi) it holds that $\mathcal{H}(\Phi) = M$, and
(vii) it holds that $\mathcal{P}(\Phi) = 20M - 7$

(cf. Definitions 2.1, 2.21, and 2.22).

Proof of Lemma 5.2. Throughout this proof let $g_n \colon \mathbb{R} \to [0,1]$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$, $x \in \mathbb{R}$ that

$$g_1(x) = \begin{cases} 2x & : x \in [0,\frac{1}{2}) \\ 2-2x & : x \in [\frac{1}{2},1] \\ 0 & : x \in \mathbb{R}\backslash[0,1] \end{cases} \tag{5.16}$$

and $g_{n+1}(x) = g_1(g_n(x))$, let $f_n \colon [0,1] \to [0,1]$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$, $k \in \{0,1,\ldots,2^n-1\}$, $x \in \big[\frac{k}{2^n},\frac{k+1}{2^n}\big)$ that $f_n(1) = 1$ and

$$f_n(x) = \big[\tfrac{2k+1}{2^n}\big]x - \tfrac{(k^2+k)}{2^{2n}}, \tag{5.17}$$

and let $r_k = (r_{k,1}, r_{k,2}, r_{k,3}, r_{k,4}) \colon \mathbb{R} \to \mathbb{R}^4$, $k \in \mathbb{N}$, satisfy for all $k \in \mathbb{N}$, $x \in \mathbb{R}$ that

$$r_1(x) = \mathfrak{R}\big(x, x-\tfrac{1}{2}, x-1, x\big) \tag{5.18}$$

and

$$r_{k+1}(x) = \mathfrak{R}\big(A_k r_k(x) + \mathbb{B}\big) \tag{5.19}$$

(cf. Definition 2.1). Observe that item (i) in Lemma 5.1 (applied with $(A_k)_{k\in\mathbb{N}} \curvearrowleft (A_k)_{k\in\mathbb{N}}$, $\mathbb{B} \curvearrowleft \mathbb{B}$, $(c_k)_{k\in\mathbb{N}} \curvearrowleft (c_k)_{k\in\mathbb{N}}$, $(g_n)_{n\in\mathbb{N}} \curvearrowleft (g_n)_{n\in\mathbb{N}}$, $(f_n)_{n\in\mathbb{N}} \curvearrowleft (f_n)_{n\in\mathbb{N}}$, $(r_k)_{k\in\mathbb{N}} \curvearrowleft (r_k)_{k\in\mathbb{N}}$ in the notation of Lemma 5.1), (5.14), (5.15), (5.18), and (5.19) assure that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}(\Phi))(x) &= -c_M r_{M,1}(x) + 2c_M r_{M,2}(x) - c_M r_{M,3}(x) + r_{M,4}(x)\\ &= -[2^{1-2M}] r_{M,1}(x) + [2^{2-2M}] r_{M,2}(x) - [2^{1-2M}] r_{M,3}(x) + r_{M,4}(x)\\ &= -[2^{-2M}][2r_{M,1}(x) - 4r_{M,2}(x) + 2r_{M,3}(x)] + r_{M,4}(x)\\ &= -[2^{-2M}] g_M(x) + r_{M,4}(x).\end{aligned} \tag{5.20}$$

This establishes item (i). Moreover, note that (5.20), [26, Lemma 3.2], and item (ii) in Lemma 5.1 (applied with $(A_k)_{k\in\mathbb{N}} \curvearrowleft (A_k)_{k\in\mathbb{N}}$, $\mathbb{B} \curvearrowleft \mathbb{B}$, $(c_k)_{k\in\mathbb{N}} \curvearrowleft (c_k)_{k\in\mathbb{N}}$, $(g_n)_{n\in\mathbb{N}} \curvearrowleft (g_n)_{n\in\mathbb{N}}$, $(f_n)_{n\in\mathbb{N}} \curvearrowleft (f_n)_{n\in\mathbb{N}}$, $(r_k)_{k\in\mathbb{N}} \curvearrowleft (r_k)_{k\in\mathbb{N}}$ in the notation of Lemma 5.1) show that for all $x \in [0, 1]$ it holds that

$$\begin{aligned}(\mathcal{R}(\Phi))(x) &= -[2^{-2M}] g_M(x) + r_{M,4}(x) = -[2^{-2M} g_M(x)] + f_{M-1}(x)\\ &= -[2^{-2M} g_M(x)] + x - \left[\textstyle\sum_{j=1}^{M-1} [2^{-2j} g_j(x)]\right]\\ &= x - \left[\textstyle\sum_{j=1}^{M} [2^{-2j} g_j(x)]\right] = f_M(x).\end{aligned} \tag{5.21}$$

This and [26, Lemma 3.2] imply that for all $x \in [0, 1]$ it holds that

$$|x^2 - (\mathcal{R}(\Phi))(x)| = |x^2 - f_M(x)| \le 2^{-2M-2} = 4^{-M-1}. \tag{5.22}$$

This establishes item (ii). Furthermore, observe that (5.20), the fact that for all $x \in \mathbb{R}\backslash[0, 1]$ it holds that $g_M(x) = 0$, and item (ii) in Lemma 5.1 (applied with $(A_k)_{k\in\mathbb{N}} \curvearrowleft (A_k)_{k\in\mathbb{N}}$, $\mathbb{B} \curvearrowleft \mathbb{B}$, $(c_k)_{k\in\mathbb{N}} \curvearrowleft (c_k)_{k\in\mathbb{N}}$, $(g_n)_{n\in\mathbb{N}} \curvearrowleft (g_n)_{n\in\mathbb{N}}$, $(f_n)_{n\in\mathbb{N}} \curvearrowleft (f_n)_{n\in\mathbb{N}}$, $(r_k)_{k\in\mathbb{N}} \curvearrowleft (r_k)_{k\in\mathbb{N}}$ in the notation of Lemma 5.1) ensure that for all $x \in \mathbb{R}\backslash[0, 1]$ it holds that

$$(\mathcal{R}(\Phi))(x) = -[2^{-2M}] g_M(x) + r_{M,4}(x) = r_{M,4}(x) = \max\{x, 0\} = \mathfrak{R}(x). \tag{5.23}$$

This establishes item (iii). In addition, note that (5.14) and (5.15) imply that $\mathcal{D}(\Phi) = (1, 4, 4, \ldots, 4, 1) \in \mathbb{N}^{M+2}$, $\|\mathcal{T}(\Phi)\|_\infty \le 4$, $\mathcal{H}(\Phi) = M$, and

$$\mathcal{P}(\Phi) = 4(1+1) + \left[\textstyle\sum_{j=2}^{M} 4(4+1)\right] + (4+1) = 8 + 20(M-1) + 5 = 20M - 7. \tag{5.24}$$

This establishes items (iv), (v), (vi), and (vii). The proof of Lemma 5.2 is thus complete. □

5.2. ANN approximations for the squared rectifier function

Corollary 5.3. *Let $M \in \mathbb{N}$, $R \in [1, \infty)$, $q \in (2, \infty)$, $(A_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{4\times 4}$, $\mathbb{A}, \mathbb{B} \in \mathbb{R}^{4\times 1}$, $(C_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{1\times 4}$, $(c_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$ satisfy for all $k \in \mathbb{N}$ that*

$$A_k = \begin{pmatrix} 2 & -4 & 2 & 0\\ 2 & -4 & 2 & 0\\ 2 & -4 & 2 & 0\\ -c_k & 2c_k & -c_k & 1\end{pmatrix}, \quad \mathbb{A} = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \quad \mathbb{B} = \begin{pmatrix}0\\ -\frac{1}{2}\\ -1\\ 0\end{pmatrix}, \quad C_k = \begin{pmatrix}-c_k & 2c_k & -c_k & 1\end{pmatrix}, \tag{5.25}$$

and $c_k = 2^{1-2k}$, and let $\Psi, \Phi \in \mathbf{N}$ satisfy

$$\Psi = \begin{cases} ((\mathbb{A}, \mathbb{B}), (C_1, 0)) & : M = 1\\ ((\mathbb{A}, \mathbb{B}), (A_1, \mathbb{B}), (A_2, \mathbb{B}), \ldots, (A_{M-1}, \mathbb{B}), (C_M, 0)) & : M > 1\end{cases} \tag{5.26}$$

and $\Phi = \mathbf{A}_{R^2,0} \bullet \Psi \bullet \mathbf{A}_{R^{-1},0}$ (cf. Definitions 2.2, 2.4, and 2.15). Then

(i) it holds that $\mathcal{R}(\Phi) \in C(\mathbb{R}, \mathbb{R})$,
(ii) it holds for all $x \in (-\infty, 0]$ that $\left|[\mathfrak{R}(x)]^2 - (\mathcal{R}(\Phi))(x)\right| = 0$,
(iii) it holds for all $x \in [0, R]$ that $\left|[\mathfrak{R}(x)]^2 - (\mathcal{R}(\Phi))(x)\right| \le 4^{-M-1} R^2$,
(iv) it holds for all $x \in [R, \infty)$ that $\left|[\mathfrak{R}(x)]^2 - (\mathcal{R}(\Phi))(x)\right| \le |\mathfrak{R}(x)|^q R^{2-q}$,

(v) it holds that $\mathcal{D}(\Phi) = (1, 4, \ldots, 4, 1) \in \mathbb{N}^{M+2}$,
(vi) it holds that $\mathcal{H}(\Phi) = M$,
(vii) it holds that $\mathcal{P}(\Phi) = 20M - 7$, *and*
(viii) it holds that $\|\mathcal{T}(\Phi)\|_\infty \le \max\{4, R^2\}$

(cf. Definitions 2.1, 2.21, and 2.22).

Proof of Corollary 5.3. Observe that Lemma 5.2 (applied with $M \curvearrowleft M$, $(A_k)_{k\in\mathbb{N}} \curvearrowleft (A_k)_{k\in\mathbb{N}}$, $\mathbb{A} \curvearrowleft \mathbb{A}$, $\mathbb{B} \curvearrowleft \mathbb{B}$, $(C_k)_{k\in\mathbb{N}} \curvearrowleft (C_k)_{k\in\mathbb{N}}$, $(c_k)_{k\in\mathbb{N}} \curvearrowleft (c_k)_{k\in\mathbb{N}}$, $\Phi \curvearrowleft \Psi$ in the notation of Lemma 5.2) assures that

(I) it holds that $\mathcal{R}(\Psi) \in C(\mathbb{R}, \mathbb{R})$,
(II) it holds for all $x \in \mathbb{R}\backslash[0, 1]$ that $(\mathcal{R}(\Psi))(x) = \mathfrak{R}(x)$, and
(III) it holds for all $x \in [0, 1]$ that $|x^2 - (\mathcal{R}(\Psi))(x)| \le 4^{-M-1}$

(cf. Definition 2.1). Next note that Proposition 2.5 and Lemma 2.16 imply that for all $x \in \mathbb{R}$ it holds that $\mathcal{R}(\Phi) \in C(\mathbb{R}, \mathbb{R})$ and

$$\begin{aligned}(\mathcal{R}(\Phi))(x) &= (\mathcal{R}(\mathbf{A}_{R^2,0} \bullet \Psi \bullet \mathbf{A}_{R^{-1},0}))(x) = \big(\mathcal{R}(\mathbf{A}_{R^2,0})\big)\Big(\big(\mathcal{R}(\Psi)\big)\big((\mathcal{R}(\mathbf{A}_{R^{-1},0}))(x)\big)\Big) \\ &= (\mathcal{R}(\mathbf{A}_{R^2,0}))\big((\mathcal{R}(\Psi))\big(R^{-1}x\big)\big) = R^2\big[(\mathcal{R}(\Psi))\big(R^{-1}x\big)\big].\end{aligned} \tag{5.27}$$

This establishes item (i). Moreover, observe that (5.27), item (I), item (II), and the fact that for all $x \in (-\infty, 0]$ it holds that $R^{-1}x \in (-\infty, 0]$ ensure that for all $x \in (-\infty, 0]$ it holds that

$$\begin{aligned}\Big|[\mathfrak{R}(x)]^2 - (\mathcal{R}(\Phi))(x)\Big| &= \Big|[\mathfrak{R}(x)]^2 - R^2\big[(\mathcal{R}(\Psi))\big(R^{-1}x\big)\big]\Big| \\ &= \Big|[\mathfrak{R}(x)]^2 - R^2\mathfrak{R}\big(R^{-1}x\big)\Big| = 0.\end{aligned} \tag{5.28}$$

This establishes item (ii). In the next step we note that item (II), (5.27), and the fact that for all $x \in [R, \infty)$ it holds that $R^{-1}x \in [1, \infty)$ demonstrate that for all $x \in [R, \infty)$ it holds that

$$0 \le (\mathcal{R}(\Phi))(x) = R^2\big[(\mathcal{R}(\Psi))\big(R^{-1}x\big)\big] = R^2\mathfrak{R}\big(R^{-1}x\big) = Rx \le x^2 = |\mathfrak{R}(x)|^2. \tag{5.29}$$

The triangle inequality and the assumption that $q \in (2, \infty)$ therefore ensure that for all $x \in [R, \infty)$ it holds that

$$\begin{aligned}\Big|[\mathfrak{R}(x)]^2 - (\mathcal{R}(\Phi))(x)\Big| &= |\mathfrak{R}(x)|^2 - (\mathcal{R}(\Phi))(x) \le |\mathfrak{R}(x)|^2 \\ &= |x|^2 = |x|^q|x|^{2-q} \le |x|^q R^{2-q} = |\mathfrak{R}(x)|^q R^{2-q}.\end{aligned} \tag{5.30}$$

This establishes item (iv). Next observe that item (III), (5.27), and the fact that for all $x \in [0, R]$ it holds that $R^{-1}x \in [0, 1]$ demonstrate that for all $x \in [0, R]$ it holds that

$$\begin{aligned}\Big|[\mathfrak{R}(x)]^2 - (\mathcal{R}(\Phi))(x)\Big| &= \Big|x^2 - R^2\big[(\mathcal{R}(\Psi))\big(R^{-1}x\big)\big]\Big| \\ &= R^2\Big|[R^{-1}x]^2 - (\mathcal{R}(\Psi))\big(R^{-1}x\big)\Big| \le 4^{-M-1}R^2.\end{aligned} \tag{5.31}$$

This establishes item (iii). Next note that (5.25) and (5.26) show that

$$R^2 C_M = \begin{pmatrix} -2^{1-2M}R^2 & 2^{2-2M}R^2 & -2^{1-2M}R^2 & R^2 \end{pmatrix} \in \mathbb{R}^{1\times 4} \tag{5.32}$$

and

$$\Phi = \mathbf{A}_{R^2,0} \bullet \Psi \bullet \mathbf{A}_{R^{-1},0} = \begin{cases} \big((R^{-1}\mathbb{A}, \mathbb{B}), (R^2C_1, 0)\big) & : M = 1 \\ \big((R^{-1}\mathbb{A}, \mathbb{B}), (A_1, \mathbb{B}), \ldots, (A_{M-1}, \mathbb{B}), (R^2C_M, 0)\big) & : M > 1. \end{cases} \tag{5.33}$$

Combining this with (5.25) implies that $\mathcal{D}(\Phi) = (1, 4, \dots, 4, 1) \in \mathbb{N}^{M+2}$, $\mathcal{H}(\Phi) = M$, $\mathcal{P}(\Phi) = 20M - 7$, and $\|\mathcal{T}(\Phi)\|_\infty \le \max\{4, R^2\}$ (cf. Definitions 2.21 and 2.22). This establishes items (v), (vi), (vii), and (viii). The proof of Corollary 5.3 is thus complete. □

5.3. ANN approximations for shifted squared rectifier functions

Lemma 5.4. *Let $\mathfrak{a} \in \mathbb{R}$, $\mathbb{J}, \Phi, \Psi \in \mathbf{N}$ satisfy*

$$\mathbb{J} = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 1 \end{pmatrix}, (-\mathfrak{a}) \right) \right) \in \left(\left(\mathbb{R}^{2 \times 1} \times \mathbb{R}^2 \right) \times \left(\mathbb{R}^{1 \times 2} \times \mathbb{R}^1 \right) \right), \tag{5.34}$$

$\mathcal{I}(\Phi) = 1$, and $\Psi = \Phi \bullet \mathbb{J}$ (cf. Definitions 2.2 and 2.4). Then

(i) it holds that $\mathcal{D}(\Psi) = \big(1, 2, \mathbb{D}_1(\Phi), \dots, \mathbb{D}_{\mathcal{L}(\Phi)}(\Phi)\big) \in \mathbb{N}^{\mathcal{L}(\Phi)+2}$,
(ii) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}(\mathbb{J}))(x) = |x| - \mathfrak{a}$,
(iii) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}(\Psi))(x) = (\mathcal{R}(\Phi))(|x| - \mathfrak{a})$, and
(iv) it holds that $\|\mathcal{T}(\Psi)\|_\infty \le (|\mathfrak{a}| + 1) \max\{1, \|\mathcal{T}(\Phi)\|_\infty\}$

(cf. Definitions 2.21 and 2.22).

Proof of Lemma 5.4. Throughout this proof let $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$ satisfy $(l_0, l_1, \dots, l_L) = \mathcal{D}(\Phi)$ and let $W_k \in \mathbb{R}^{l_k \times l_{k-1}}$, $k \in \{1, 2, \dots, L\}$, and $B_k \in \mathbb{R}^{l_k}$, $k \in \{1, 2, \dots, L\}$, satisfy $\Phi = ((W_1, B_1), (W_2, B_2), \dots, (W_L, B_L))$. Observe that $\mathcal{D}(\mathbb{J}) = (1, 2, 1) \in \mathbb{N}^3$. Proposition 2.5 therefore ensures that $\mathcal{D}(\Psi) = \mathcal{D}(\Phi \bullet \mathbb{J}) = (1, 2, \mathbb{D}_1(\Phi), \dots, \mathbb{D}_{\mathcal{L}(\Phi)}(\Phi)) \in \mathbb{N}^{\mathcal{L}(\Phi)+2}$. This establishes item (i). Next note that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}(\mathbb{J}))(x) = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} \mathfrak{R}(x + 0) \\ \mathfrak{R}(-x + 0) \end{pmatrix} - \mathfrak{a} = \mathfrak{R}(x) + \mathfrak{R}(-x) - \mathfrak{a} = |x| - \mathfrak{a} \tag{5.35}$$

(cf. Definition 2.1). This establishes item (ii). Moreover, observe that (5.35) and Proposition 2.5 assure that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}(\Psi))(x) = (\mathcal{R}(\Phi \bullet \mathbb{J}))(x) = (\mathcal{R}(\Phi))((\mathcal{R}(\mathbb{J}))(x)) = (\mathcal{R}(\Phi))(|x| - \mathfrak{a}). \tag{5.36}$$

This establishes item (iii). In addition, note that

$$\Psi = \Phi \bullet \mathbb{J} = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right), \left(W_1 \begin{pmatrix} 1 & 1 \end{pmatrix}, W_1(-\mathfrak{a}) + B_1 \right), (W_2, B_2), \dots, (W_L, B_L) \right). \tag{5.37}$$

The fact that for all $\mathfrak{W} = (w_i)_{i \in \{1, 2, \dots, l_1\}} \in \mathbb{R}^{l_1 \times 1}$, $\mathfrak{B} = (b_1, b_2, \dots, b_{l_1}) \in \mathbb{R}^{l_1}$ it holds that

$$\mathfrak{W} \begin{pmatrix} 1 & 1 \end{pmatrix} = \begin{pmatrix} w_1 & w_1 \\ w_2 & w_2 \\ \vdots & \vdots \\ w_{l_1} & w_{l_1} \end{pmatrix} \in \mathbb{R}^{l_1 \times 2} \qquad \text{and} \qquad \mathfrak{W}(-\mathfrak{a}) + \mathfrak{B} = \begin{pmatrix} -\mathfrak{a} w_1 + b_1 \\ -\mathfrak{a} w_2 + b_2 \\ \vdots \\ -\mathfrak{a} w_{l_1} + b_{l_1} \end{pmatrix} \in \mathbb{R}^{l_1} \tag{5.38}$$

hence demonstrates that

$$\begin{aligned} \|\mathcal{T}(\Psi)\|_\infty &\le \max\{1, \|\mathcal{T}(\Phi)\|_\infty, (|\mathfrak{a}| + 1)\|\mathcal{T}(\Phi)\|_\infty\} = \max\{1, (|\mathfrak{a}| + 1)\|\mathcal{T}(\Phi)\|_\infty\} \\ &\le \max\{(|\mathfrak{a}| + 1), (|\mathfrak{a}| + 1)\|\mathcal{T}(\Phi)\|_\infty\} = (|\mathfrak{a}| + 1) \max\{1, \|\mathcal{T}(\Phi)\|_\infty\} \end{aligned} \tag{5.39}$$

(cf. Definitions 2.21 and 2.22). This establishes item (iv). The proof of Lemma 5.4 is thus complete. □

Corollary 5.5. *Let $\mathfrak{a} \in [0,\infty)$, $M \in \mathbb{N} \cap [2,\infty)$, $R \in [1,\infty)$, $q \in (2,\infty)$, $(A_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{4\times 4}$, $\mathbb{A}, \mathbb{B} \in \mathbb{R}^{4\times 1}$, $(C_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{1\times 4}$, $(c_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$ satisfy for all $k \in \mathbb{N}$ that*

$$A_k = \begin{pmatrix} 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ -c_k & 2c_k & -c_k & 1 \end{pmatrix}, \quad \mathbb{A} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbb{B} = \begin{pmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{pmatrix}, \quad C_k = \begin{pmatrix} -c_k & 2c_k & -c_k & 1 \end{pmatrix}, \tag{5.40}$$

and $c_k = 2^{1-2k}$ and let $\Phi, \mathbb{J}, \Psi \in \mathbf{N}$ satisfy

$$\Phi = \big((R^{-1}\mathbb{A}, \mathbb{B}), (A_1, \mathbb{B}), \ldots, (A_{M-1}, \mathbb{B}), (R^2 C_M, 0)\big), \tag{5.41}$$

$$\mathbb{J} = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right), \left(\begin{pmatrix} 1 & 1 \end{pmatrix}, (-\mathfrak{a})\right)\right), \tag{5.42}$$

and $\Psi = \Phi \bullet \mathbb{J}$ (cf. Definitions 2.2 and 2.4). Then

(i) it holds that $\mathcal{R}(\Psi) \in C(\mathbb{R},\mathbb{R})$,
(ii) it holds that $\mathcal{D}(\Psi) = (1,2,4,4,\ldots,4,1) \in \mathbb{N}^{M+3}$,
(iii) it holds that $\mathcal{H}(\Psi) = M+1$,
(iv) it holds that $\mathcal{P}(\Psi) = 20M+1$,
(v) it holds that $\|\mathcal{T}(\Psi)\|_\infty \le (|\mathfrak{a}|+1)\max\{4, R^2\}$,
(vi) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}(\Psi))(x) = (\mathcal{R}(\Psi))(-x)$,
(vii) it holds for all $x \in \mathbb{R}$ with $|x| \le \mathfrak{a}$ that $\big|[\mathfrak{R}(|x|-\mathfrak{a})]^2 - (\mathcal{R}(\Psi))(x)\big| = 0$,
(viii) it holds for all $x \in \mathbb{R}$ with $\mathfrak{a} \le |x| \le R+\mathfrak{a}$ that $\big|[\mathfrak{R}(|x|-\mathfrak{a})]^2 - (\mathcal{R}(\Psi))(x)\big| \le 4^{-M-1}R^2$, and
(ix) it holds for all $x \in \mathbb{R}$ with $|x| \ge R+\mathfrak{a}$ that $\big|[\mathfrak{R}(|x|-\mathfrak{a})]^2 - (\mathcal{R}(\Psi))(x)\big| \le [|x|-\mathfrak{a}]^q R^{2-q}$

(cf. Definitions 2.1, 2.21, and 2.22).

Proof of Corollary 5.5. Observe that Corollary 5.3 (applied with $M \curvearrowleft M$, $R \curvearrowleft R$, $q \curvearrowleft q$, $(A_k)_{k\in\mathbb{N}} \curvearrowleft (A_k)_{k\in\mathbb{N}}$, $\mathbb{A} \curvearrowleft \mathbb{A}$, $\mathbb{B} \curvearrowleft \mathbb{B}$, $(C_k)_{k\in\mathbb{N}} \curvearrowleft (C_k)_{k\in\mathbb{N}}$, $(c_k)_{k\in\mathbb{N}} \curvearrowleft (c_k)_{k\in\mathbb{N}}$, $\Phi \curvearrowleft \Phi$ in the notation of Corollary 5.3) implies that

(I) it holds that $\mathcal{R}(\Phi) \in C(\mathbb{R},\mathbb{R})$,
(II) it holds for all $x \in (-\infty, 0]$ that $\big|[\mathfrak{R}(x)]^2 - (\mathcal{R}(\Phi))(x)\big| = 0$,
(III) it holds for all $x \in [0, R]$ that $\big|[\mathfrak{R}(x)]^2 - (\mathcal{R}(\Phi))(x)\big| \le 4^{-M-1}R^2$,
(IV) it holds for all $x \in [R,\infty)$ that $\big|[\mathfrak{R}(x)]^2 - (\mathcal{R}(\Phi))(x)\big| \le |\mathfrak{R}(x)|^q R^{2-q}$,
(V) it holds that $\mathcal{D}(\Phi) = (1,4,\ldots,4,1) \in \mathbb{N}^{M+2}$, and
(VI) it holds that $\|\mathcal{T}(\Phi)\|_\infty \le \max\{4, R^2\}$

(cf. Definitions 2.1, 2.21, and 2.22). Next note that Lemma 5.4 (applied with $\mathfrak{a} \curvearrowleft \mathfrak{a}$, $\mathbb{J} \curvearrowleft \mathbb{J}$, $\Phi \curvearrowleft \Phi$, $\Psi \curvearrowleft \Psi$ in the notation of Lemma 5.4), item (V), and item (VI) ensure that

$$\mathcal{D}(\Psi) = \big(1, 2, \mathbb{D}_1(\Phi), \ldots, \mathbb{D}_{\mathcal{L}(\Phi)}(\Phi)\big) = (1, 2, \underbrace{4, \ldots, 4}_{M}, 1) \in \mathbb{N}^{M+3}, \qquad \mathcal{H}(\Psi) = M+1, \tag{5.43}$$

$$\mathcal{P}(\Psi) = 2(1+1) + 4(2+1) + \underbrace{4(4+1) + \ldots + 4(4+1)}_{M-1} + 1(4+1) = 20M+1, \tag{5.44}$$

and

$$\|\mathcal{T}(\Psi)\|_\infty \le (|\mathfrak{a}|+1)\max\{1, \|\mathcal{T}(\Phi)\|_\infty\} \le (|\mathfrak{a}|+1)\max\{4, R^2\}. \tag{5.45}$$

This establishes items (i), (ii), (iii), (iv), and (v). Next observe that Lemma 5.4 (applied with $\mathfrak{a} \curvearrowleft \mathfrak{a}$, $\mathbb{J} \curvearrowleft \mathbb{J}$, $\Phi \curvearrowleft \Phi$, $\Psi \curvearrowleft \Psi$ in the notation of Lemma 5.4) assures that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}(\mathbb{J}))(x) = |x| - \mathfrak{a} \tag{5.46}$$

and

$$(\mathcal{R}(\Psi))(x) = (\mathcal{R}(\Phi))(|x| - \mathfrak{a}) = (\mathcal{R}(\Phi))(|-x| - \mathfrak{a}) = (\mathcal{R}(\Psi))(-x). \tag{5.47}$$

This establishes item (vi). Furthermore, note that (5.46) shows that for all $x \in [-\mathfrak{a}, \mathfrak{a}]$ it holds that $(\mathcal{R}(\mathbb{J}))(x) = |x| - \mathfrak{a} \leq 0$. Combining this with item (II) proves that for all $x \in [-\mathfrak{a}, \mathfrak{a}]$ it holds that

$$\left|[\mathfrak{R}(|x| - \mathfrak{a})]^2 - (\mathcal{R}(\Psi))(x)\right| = \left|[\mathfrak{R}((\mathcal{R}(\mathbb{J}))(x))]^2 - (\mathcal{R}(\Phi))((\mathcal{R}(\mathbb{J}))(x))\right| = 0. \tag{5.48}$$

This establishes item (vii). Moreover, observe that (5.46) demonstrates that for all $x \in \mathbb{R}$ with $\mathfrak{a} \leq |x| \leq R + \mathfrak{a}$ it holds that $(\mathcal{R}(\mathbb{J}))(x) = |x| - \mathfrak{a} \in [0, R]$. This and item (III) ensure that for all $x \in \mathbb{R}$ with $\mathfrak{a} \leq |x| \leq R + \mathfrak{a}$ it holds that

$$\left|[\mathfrak{R}(|x| - \mathfrak{a})]^2 - (\mathcal{R}(\Psi))(x)\right| = \left|[\mathfrak{R}((\mathcal{R}(\mathbb{J}))(x))]^2 - (\mathcal{R}(\Phi))((\mathcal{R}(\mathbb{J}))(x))\right| \leq 4^{-M-1} R^2. \tag{5.49}$$

This establishes item (viii). In addition, note that (5.46) proves that for all $x \in \mathbb{R}$ with $|x| \geq R + \mathfrak{a}$ it holds that $(\mathcal{R}(\mathbb{J}))(x) = |x| - \mathfrak{a} \in [R, \infty)$. Item (IV) hence shows that for all $x \in \mathbb{R}$ with $|x| \geq R + \mathfrak{a}$ it holds that

$$\begin{aligned} \left|[\mathfrak{R}(|x| - \mathfrak{a})]^2 - (\mathcal{R}(\Psi))(x)\right| &= \left|[\mathfrak{R}((\mathcal{R}(\mathbb{J}))(x))]^2 - (\mathcal{R}(\Phi))((\mathcal{R}(\mathbb{J}))(x))\right| \\ &\leq |\mathfrak{R}((\mathcal{R}(\mathbb{J}))(x))|^q R^{2-q} = |\mathfrak{R}(|x| - \mathfrak{a})|^q R^{2-q}. \end{aligned} \tag{5.50}$$

This establishes item (ix). The proof of Corollary 5.5 is thus complete. □

5.4. Lower and upper bounds for integrals of certain specific high-dimensional functions

The goal of this subsection is to establish in Lemma 5.9 below lower and upper bounds for the weighted L^2-norm of the function $\mathbb{R}^d \ni (x_1, \ldots, x_d) \mapsto \sum_{j=1}^d [\max\{|x_j| - \sqrt{2d}, 0\}]^2 \in \mathbb{R}$. To this end, we recollect some well-known Gaussian tail estimates in Lemma 5.6 and Lemma 5.7. The first can be found, e.g., in Klenke [39, Lemma 22.2], but for completeness, we include their short proof.

Lemma 5.6. *Let $s \in (0, \infty)$. Then*

$$\int_s^\infty e^{-\frac{1}{2}x^2}\, dx \geq \frac{e^{-\frac{1}{2}s^2}}{s + s^{-1}}. \tag{5.51}$$

Proof of Lemma 5.6. Observe that the integration by parts formula ensures that

$$\begin{aligned} \int_s^\infty e^{-\frac{1}{2}x^2}\, dx &= \int_s^\infty -x^{-1}\left[e^{-\frac{1}{2}x^2}\right]'\, dx = \lim_{T \to \infty}\left(\left[-x^{-1}e^{-\frac{1}{2}x^2}\right]_{x=s}^{x=T}\right) - \int_s^\infty \left[x^{-2}e^{-\frac{1}{2}x^2}\right] dx \\ &= s^{-1}e^{-\frac{1}{2}s^2} - \int_s^\infty \left[x^{-2}e^{-\frac{1}{2}x^2}\right] dx \geq s^{-1}e^{-\frac{1}{2}s^2} - s^{-2}\int_s^\infty e^{-\frac{1}{2}x^2}\, dx. \end{aligned} \tag{5.52}$$

Hence, we obtain that

$$\begin{aligned}\int_s^\infty e^{-\frac{1}{2}x^2}\,dx &= \left[\frac{s^2}{1+s^2}\right]\left[1+\frac{1}{s^2}\right]\left[\int_s^\infty e^{-\frac{1}{2}x^2}\,dx\right]\\ &= \left[\frac{s^2}{1+s^2}\right]\left[\int_s^\infty e^{-\frac{1}{2}x^2}\,dx + \frac{1}{s^2}\int_s^\infty e^{-\frac{1}{2}x^2}\,dx\right]\\ &\geq \left[\frac{s^2}{1+s^2}\right]\left[\frac{e^{-\frac{1}{2}s^2}}{s}\right] = \frac{e^{-\frac{1}{2}s^2}}{s+s^{-1}}.\end{aligned} \tag{5.53}$$

The proof of Lemma 5.6 is thus complete. □

Lemma 5.7. *Let $\sigma, s \in (0,\infty)$. Then*

$$\int_s^\infty e^{-\sigma x^2}\,dx \geq \frac{e^{-\sigma s^2}}{s^{-1}+2\sigma s}. \tag{5.54}$$

Proof of Lemma 5.7. Note that the integral transformation theorem and Lemma 5.6 (applied with $s \curvearrowleft s\sqrt{2\sigma}$ in the notation of Lemma 5.6) ensure that

$$\int_s^\infty e^{-\sigma x^2}\,dx = \frac{1}{\sqrt{2\sigma}}\int_{s\sqrt{2\sigma}}^\infty e^{-\frac{1}{2}x^2}\,dx \geq \frac{1}{\sqrt{2\sigma}}\left[\frac{e^{-\frac{1}{2}(s\sqrt{2\sigma})^2}}{s\sqrt{2\sigma}+(s\sqrt{2\sigma})^{-1}}\right] = \frac{e^{-\sigma s^2}}{s^{-1}+2\sigma s}. \tag{5.55}$$

The proof of Lemma 5.7 is thus complete. □

Lemma 5.8. *Let $d \in \mathbb{N}$. Then*

$$\frac{\sqrt{2d}(2d+1)}{4d^2(4d^2+6d+1)}\left[\frac{2}{\pi}\right]^{1/2} e^{-1-\frac{1}{4d}} \geq 50^{-1}d^{-5/2}. \tag{5.56}$$

Proof of Lemma 5.8. Observe that $48d^2 - 28d \geq 20d^2 \geq 13$. This implies that $4d^2+6d+1 \leq (25/13)(4d^2+2d) = (50/13)d(2d+1)$. The fact that $13 \geq 2\sqrt{\pi}e^{5/4}$ and the fact that $-1-\frac{1}{4d} \geq -\frac{5}{4}$ hence ensure that

$$\frac{\sqrt{2d}(2d+1)}{4d^2(4d^2+6d+1)}\left[\frac{2}{\pi}\right]^{1/2} e^{-1-\frac{1}{4d}} \geq \frac{\sqrt{2d}}{4d^2}\left[\frac{13}{50d}\right]\left[\frac{2}{\pi}\right]^{1/2} e^{-5/4} \geq \frac{\sqrt{2d}}{4d^2}\left[\frac{2\sqrt{2}}{50d}\right] = 50^{-1}d^{-5/2}. \tag{5.57}$$

The proof of Lemma 5.8 is thus complete. □

Lemma 5.9. *Let $d \in \mathbb{N}$ and let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ and $g\colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $\varphi(x) = (2\pi)^{-d/2}\exp(-\frac{1}{2}(\sum_{j=1}^d |x_j|^2))$ and $g(x) = \sum_{j=1}^d [\max\{|x_j| - \sqrt{2d}, 0\}]^2$. Then*

$$(50)^{-1}d^{-3/2}e^{-d} \leq \int_{\mathbb{R}^d} |g(x)|^2\varphi(x)\,dx \leq 3d^2e^{-d}. \tag{5.58}$$

Proof of Lemma 5.9. Throughout this proof let $\Gamma\colon (0,\infty) \to (0,\infty)$ satisfy for all $x \in (0,\infty)$ that $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$. Note that the fact that for all $k \in \mathbb{N}$, $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_k \in \mathbb{R}$ it holds that

$$|\mathfrak{a}_1|^2 + |\mathfrak{a}_2|^2 + \ldots + |\mathfrak{a}_k|^2 \leq (|\mathfrak{a}_1| + |\mathfrak{a}_2| + \ldots + |\mathfrak{a}_k|)^2 \leq k(|\mathfrak{a}_1|^2 + |\mathfrak{a}_2|^2 + \ldots + |\mathfrak{a}_k|^2) \tag{5.59}$$

ensures that for all $(x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\sum_{j=1}^{d} \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^4 \le \left[\sum_{j=1}^{d} \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^2\right]^2 \le d\left[\sum_{j=1}^{d} \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^4\right] \tag{5.60}$$

(cf. Definition 2.1). The fact that for all $k \in \mathbb{N}$ it holds that $\int_{\mathbb{R}^k} (2\pi)^{-k/2} e^{-\frac{1}{2}\|x\|_2^2}\, dx = 1$ therefore demonstrates that

$$\begin{aligned}
& d \int_{\mathbb{R}} \big[\mathfrak{R}\big(|x| - \sqrt{2d}\big)\big]^4 (2\pi)^{-1/2} e^{-\frac{1}{2}x^2}\, dx \\
&= d \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \big[\mathfrak{R}\big(|x_1| - \sqrt{2d}\big)\big]^4 (2\pi)^{-d/2} e^{-\frac{1}{2}[\sum_{j=1}^{d} |x_j|^2]}\, dx_d \dots dx_2\, dx_1 \\
&= \sum_{j=1}^{d} \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^4 (2\pi)^{-d/2} e^{-\frac{1}{2}[\sum_{j=1}^{d} |x_j|^2]}\, dx_d \dots dx_2\, dx_1 \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left[\sum_{j=1}^{d} \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^4\right] (2\pi)^{-d/2} e^{-\frac{1}{2}[\sum_{j=1}^{d} |x_j|^2]}\, dx_d \dots dx_2\, dx_1 \\
&\le \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left[\sum_{j=1}^{d} \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^2\right]^2 (2\pi)^{-d/2} e^{-\frac{1}{2}[\sum_{j=1}^{d} |x_j|^2]}\, dx_d \dots dx_2\, dx_1 \\
&= \int_{\mathbb{R}^d} |g(x)|^2 \varphi(x)\, dx
\end{aligned} \tag{5.61}$$

and

$$\begin{aligned}
& d^2 \int_{\mathbb{R}} \big[\mathfrak{R}\big(|x| - \sqrt{2d}\big)\big]^4 (2\pi)^{-1/2} e^{-\frac{1}{2}x^2}\, dx \\
&= d^2 \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \big[\mathfrak{R}\big(|x_1| - \sqrt{2d}\big)\big]^4 (2\pi)^{-d/2} e^{-\frac{1}{2}[\sum_{j=1}^{d} |x_j|^2]}\, dx_d \dots dx_2\, dx_1 \\
&= d \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left[\sum_{j=1}^{d} \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^4\right] (2\pi)^{-d/2} e^{-\frac{1}{2}[\sum_{j=1}^{d} |x_j|^2]}\, dx_d \dots dx_2\, dx_1 \\
&\ge \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left[\sum_{j=1}^{d} \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^2\right]^2 (2\pi)^{-d/2} e^{-\frac{1}{2}[\sum_{j=1}^{d} |x_j|^2]}\, dx_d \dots dx_2\, dx_1 \\
&= \int_{\mathbb{R}^d} |g(x)|^2 \varphi(x)\, dx
\end{aligned} \tag{5.62}$$

(cf. Definition 2.21). Hence, we obtain that

$$d\int_{\mathbb{R}}\big[\mathfrak{R}\big(|x|-\sqrt{2d}\big)\big]^4(2\pi)^{-1/2}e^{-\frac{1}{2}x^2}\,dx \le \int_{\mathbb{R}^d}|g(x)|^2\varphi(x)\,dx \le d^2\int_{\mathbb{R}}\big[\mathfrak{R}\big(|x|-\sqrt{2d}\big)\big]^4(2\pi)^{-1/2}e^{-\frac{1}{2}x^2}\,dx. \tag{5.63}$$

Next observe that Lemma 5.7 (applied with $\sigma \curvearrowleft 1/2$, $s \curvearrowleft (2d)^{1/2}+(2d)^{-1/2}$ in the notation of Lemma 5.7) and Lemma 5.8 (applied with $d \curvearrowleft d$ in the notation of Lemma 5.8) ensure that

$$\begin{aligned}
&\int_{\mathbb{R}}\big[\mathfrak{R}\big(|x|-\sqrt{2d}\big)\big]^4(2\pi)^{-1/2}e^{-\frac{1}{2}x^2}\,dx\\
&=2\left[\int_0^\infty\big[\mathfrak{R}\big(|x|-\sqrt{2d}\big)\big]^4(2\pi)^{-1/2}e^{-\frac{1}{2}x^2}\,dx\right]=\left[\frac{2}{\pi}\right]^{1/2}\left[\int_{\sqrt{2d}}^\infty\big[x-\sqrt{2d}\big]^4e^{-\frac{1}{2}x^2}\,dx\right]\\
&\ge\frac{1}{4d^2}\left[\frac{2}{\pi}\right]^{1/2}\left[\int_{(2d)^{1/2}+(2d)^{-1/2}}^\infty e^{-\frac{1}{2}x^2}\,dx\right]\ge\frac{1}{4d^2}\left[\frac{2}{\pi}\right]^{1/2}\left[\frac{[(2d)^{1/2}+(2d)^{-1/2}]e^{-\frac{1}{2}(2d+2+(2d)^{-1})}}{1+(2d+2+(2d)^{-1})}\right]\\
&=e^{-d}\left[\frac{\sqrt{2d}(2d+1)}{4d^2(4d^2+6d+1)}\left[\frac{2}{\pi}\right]^{1/2}e^{-1-\frac{1}{4d}}\right]\ge 50^{-1}d^{-5/2}e^{-d}.
\end{aligned}\tag{5.64}$$

Moreover, note that the integral transformation theorem and Lemma 3.1 demonstrate that

$$\begin{aligned}
&\int_{\mathbb{R}}\big[\mathfrak{R}\big(|x|-\sqrt{2d}\big)\big]^4(2\pi)^{-1/2}e^{-\frac{1}{2}x^2}\,dx\\
&=2\left[\int_0^\infty\big[\mathfrak{R}\big(|x|-\sqrt{2d}\big)\big]^4(2\pi)^{-1/2}e^{-\frac{1}{2}x^2}\,dx\right]=\left[\frac{2}{\pi}\right]^{1/2}\left[\int_{\sqrt{2d}}^\infty\big[x-\sqrt{2d}\big]^4e^{-\frac{1}{2}x^2}\,dx\right]\\
&=\left[\frac{2}{\pi}\right]^{1/2}\left[\int_0^\infty x^4\,e^{-\frac{1}{2}(x+\sqrt{2d})^2}\,dx\right]=\left[\frac{2}{\pi}\right]^{1/2}e^{-d}\left[\int_0^\infty x^4\,e^{-\frac{1}{2}(x^2+2\sqrt{2d}x)}\,dx\right]\\
&\le\left[\frac{2}{\pi}\right]^{1/2}e^{-d}\left[\int_0^\infty x^4\,e^{-\frac{1}{2}x^2}\,dx\right]=\left[\frac{4}{\sqrt{\pi}}\right]e^{-d}\left[\int_0^\infty x^{3/2}e^{-x}\,dx\right]=\left[\frac{4}{\sqrt{\pi}}\right]e^{-d}\,\Gamma\!\left(\frac{5}{2}\right)=3e^{-d}.
\end{aligned}\tag{5.65}$$

Combining this with (5.63) and (5.64) demonstrates that

$$50^{-1}d^{-3/2}e^{-d}\le\int_{\mathbb{R}^d}|g(x)|^2\varphi(x)\,dx\le 3d^2e^{-d}. \tag{5.66}$$

The proof of Lemma 5.9 is thus complete. □

5.5. ANN representations for multiplications with powers of real numbers

Lemma 5.10. *Let $n\in\mathbb{N}$, $\lambda\in\mathbb{R}$, $\Phi,\mathscr{I},\Psi\in\mathbf{N}$ satisfy $\mathscr{I}=(\lambda\circledast\mathfrak{I}_{\mathcal{O}(\Phi)})\bullet\mathbf{A}_{\lambda\mathrm{I}_{\mathcal{O}(\Phi)},0}$ and $\Psi=(\mathscr{I}^{\bullet n})\bullet\Phi$ (cf. Definitions 2.2, 2.4, 2.7, 2.8, 2.13, 2.15, and 2.17). Then*

(i) it holds that $\mathcal{I}(\Psi)=\mathcal{I}(\Phi)$,
(ii) it holds that $\mathcal{H}(\Psi)=\mathcal{H}(\Phi)+n$,
(iii) it holds that $\mathcal{P}(\Psi)\le 2\mathcal{P}(\Phi)+6n|\mathcal{O}(\Phi)|^2$,
(iv) it holds that $\|\mathcal{T}(\Psi)\|_\infty \le \max\{1,|\lambda|\}\max\{|\lambda|,\|\mathcal{T}(\Phi)\|_\infty\}$, and
(v) it holds for all $x\in\mathbb{R}^{\mathcal{I}(\Psi)}$ that $(\mathcal{R}(\Psi))(x)=\lambda^{2n}(\mathcal{R}(\Phi))(x)$

(cf. Definitions 2.21 and 2.22).

Proof of Lemma 5.10. Throughout this proof let $d,l_0,l_1,l_2\in\mathbb{N}$ satisfy $l_0=l_2=d=\mathcal{O}(\Phi)$ and $l_1=2d$, let $0_\mathfrak{n}\in\mathbb{R}^\mathfrak{n}$, $\mathfrak{n}\in\mathbb{N}$, satisfy for all $\mathfrak{n}\in\mathbb{N}$ that $0_\mathfrak{n}=0$, and let $W_k\in\mathbb{R}^{l_k\times l_{k-1}}$, $k\in\{1,2\}$, satisfy $\mathfrak{I}_d=((W_1,0_{2d}),(W_2,0_d))$ (cf. Lemma 2.14). Observe that Lemma 2.9, Proposition 2.5, and Lemma 2.18 show that

$$\mathcal{D}(\mathscr{I}^{\bullet n})=(d,2d,2d,\dots,2d,d)\in\mathbb{N}^{n+2},\qquad \mathcal{H}(\mathscr{I}^{\bullet n})=n,\qquad \mathcal{H}(\Psi)=\mathcal{H}(\Phi)+n,$$
$$\text{and}\qquad \mathcal{D}(\Psi)=(\mathbb{D}_0(\Phi),\mathbb{D}_1(\Phi),\dots,\mathbb{D}_{\mathcal{H}(\Phi)}(\Phi),\underbrace{2d,2d,\dots,2d}_{n},d)\in\mathbb{N}^{\mathcal{L}(\Phi)+n+1}. \tag{5.67}$$

Therefore, we obtain that

$$\begin{aligned}\mathcal{P}(\Psi)&=\mathcal{P}(\Phi)+\mathbb{D}_{\mathcal{L}(\Phi)}(\Phi)(\mathbb{D}_{\mathcal{H}(\Phi)}(\Phi)+1)+\underbrace{2d(2d+1)+\ldots+2d(2d+1)}_{n-1}+d(2d+1)\\&=\mathcal{P}(\Phi)+\mathbb{D}_{\mathcal{L}(\Phi)}(\Phi)(\mathbb{D}_{\mathcal{H}(\Phi)}(\Phi)+1)+(n-1)(4d^2+2d)+(2d^2+d)\\&\le 2\mathcal{P}(\Phi)+6d^2n=2\mathcal{P}(\Phi)+6n|\mathcal{O}(\Phi)|^2.\end{aligned} \tag{5.68}$$

Moreover, note that (2.3) and the fact that for all $\alpha\in\mathbb{R}$, $\phi\in\mathbf{N}$ it holds that $\alpha\circledast\phi=\mathbf{A}_{\alpha\,\mathrm{I}_{\mathcal{O}(\phi)},0}\bullet\phi$ ensure that $\mathscr{I}=((\lambda W_1,0_{2d}),(\lambda W_2,0_d))$. Therefore, we obtain that

$$\mathscr{I}^{\bullet n}=((\lambda W_1,0_{2d}),\underbrace{(\lambda^2W_1W_2,0_{2d}),\dots,(\lambda^2W_1W_2,0_{2d})}_{n-1},(\lambda W_2,0_d)). \tag{5.69}$$

Next observe that the fact that $\mathfrak{I}_d=((W_1,0_{2d}),(W_2,0_d))$, (2.8), (2.11), and (2.12) demonstrate that $\|\mathcal{T}(((W_1,0_{2d})))\|_\infty=\|\mathcal{T}(((W_2,0_d)))\|_\infty=\|\mathcal{T}(((W_1W_2,0_{2d})))\|_\infty=1$ (cf. Definitions 2.21 and 2.22). Combining this with (5.69) establishes that

$$\|\mathcal{T}(\mathscr{I}^{\bullet n})\|_\infty=\begin{cases}|\lambda| & : n=1\\ |\lambda|\max\{1,|\lambda|\} & : n>1.\end{cases} \tag{5.70}$$

Furthermore, note that the fact that $\mathfrak{I}_d=((W_1,0_{2d}),(W_2,0_d))$, (2.8), (2.11), and (2.12) show that for all $k\in\mathbb{N}$, $\mathfrak{W}\in\mathbb{R}^{d\times k}$, $\mathfrak{B}\in\mathbb{R}^d$ it holds that

$$\|\mathcal{T}(((\lambda W_1\mathfrak{W},\lambda W_1\mathfrak{B}+0_{2d})))\|_\infty=|\lambda|\|\mathcal{T}(((\mathfrak{W},\mathfrak{B})))\|_\infty. \tag{5.71}$$

This, Lemma 2.23, (5.69), and (5.70) establish that

$$\begin{aligned}\|\mathcal{T}(\Psi)\|_\infty&=\big\|\mathcal{T}\big((\mathscr{I}^{\bullet n})\bullet\Phi\big)\big\|_\infty\le\max\big\{\|\mathcal{T}(\mathscr{I}^{\bullet n})\|_\infty,\|\mathcal{T}(\Phi)\|_\infty,|\lambda|\|\mathcal{T}(\Phi)\|_\infty\big\}\\&\le\max\big\{|\lambda|\max\{1,|\lambda|\},\|\mathcal{T}(\Phi)\|_\infty,|\lambda|\|\mathcal{T}(\Phi)\|_\infty\big\}\\&=\max\{1,|\lambda|\}\max\{|\lambda|,\|\mathcal{T}(\Phi)\|_\infty\}.\end{aligned} \tag{5.72}$$

In addition, observe that Proposition 2.5, Lemma 2.14, Lemma 2.16, and Lemma 2.18 demonstrate that for all $x\in\mathbb{R}^d$ it holds that

$$\begin{aligned}(\mathcal{R}(\mathscr{I}))(x)&=(\mathcal{R}((\lambda\circledast\mathfrak{I}_d)\bullet\mathbf{A}_{\lambda\,\mathrm{I}_d,0}))(x)=(\mathcal{R}(\lambda\circledast\mathfrak{I}_d))((\mathcal{R}(\mathbf{A}_{\lambda\,\mathrm{I}_d,0}))(x))\\&=(\mathcal{R}(\lambda\circledast\mathfrak{I}_d))(\lambda x)=\lambda[(\mathcal{R}(\mathfrak{I}_d))(\lambda x)]=\lambda[\lambda x]=\lambda^2x.\end{aligned} \tag{5.73}$$

Induction therefore shows that for all $x \in \mathbb{R}^d$ it holds that $(\mathcal{R}(\mathscr{I}^{\bullet n}))(x) = \lambda^{2n} x$. Hence, we obtain that for all $x \in \mathbb{R}^{\mathcal{I}(\Psi)}$ it holds that

$$\mathcal{R}(\Psi)(x) = (\mathcal{R}((\mathscr{I}^{\bullet n}) \bullet \Phi))(x) = (\mathcal{R}(\mathscr{I}^{\bullet n}))((\mathcal{R}(\Phi))(x)) = \lambda^{2n} (\mathcal{R}(\Phi))(x). \tag{5.74}$$

Combining this with (5.67), (5.68), and (5.72) establishes items (i), (ii), (iii), (iv), and (v). The proof of Lemma 5.10 is thus complete. □

5.6. ANN approximations for certain specific high-dimensional functions

Theorem 5.11. *Let $d \in \mathbb{N}$, $M \in \mathbb{N} \cap [2, \infty)$, $R \in [1, \infty)$, let $\varphi : \mathbb{R}^d \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $\varphi(x) = (2\pi)^{-d/2} \exp(-\frac{1}{2}(\sum_{j=1}^d |x_j|^2))$ and $g(x) = \sum_{j=1}^d [\max\{|x_j| - \sqrt{2d}, 0\}]^2$, and let $\mathfrak{g} : \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$ that $\mathfrak{g}(x) = [\int_{\mathbb{R}^d} |g(y)|^2 \varphi(y)\, dy]^{-1/2} g(x)$. Then there exists $\Phi \in \mathbf{N}$ such that*

(i) it holds that $\mathcal{R}(\Phi) \in C(\mathbb{R}^d, \mathbb{R})$,
(ii) it holds that $\mathcal{H}(\Phi) = d + M + 1$,
(iii) it holds that $\mathcal{P}(\Phi) \le 42 d^2 M + 6d$,
(iv) it holds that $\|\mathcal{T}(\Phi)\|_\infty \le 12 d^{3/2} \max\{4, R^2\}$, and
(v) it holds that $\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\, dx \le 50 d^{7/2} \big[16^{-M-1} R^4 + 105 R^{-4}\big]$

(cf. Definitions 2.2, 2.21, and 2.22).

Proof of Theorem 5.11. Throughout this proof let $\Gamma : (0, \infty) \to (0, \infty)$ satisfy for all $x \in (0, \infty)$ that $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$, let $\psi \in \mathbf{N}$ satisfy that

(I) it holds that $\mathcal{R}(\psi) \in C(\mathbb{R}, \mathbb{R})$,
(II) it holds that $\mathcal{D}(\psi) = (1, 2, \underbrace{4, \ldots, 4}_{M}, 1) \in \mathbb{N}^{M+3}$,
(III) it holds that $\|\mathcal{T}(\psi)\|_\infty \le (\sqrt{2d} + 1) \max\{4, R^2\}$,
(IV) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}(\psi))(x) = (\mathcal{R}(\psi))(-x)$,
(V) it holds for all $x \in \mathbb{R}$ with $|x| \le \sqrt{2d}$ that $\big|[\mathfrak{R}(|x| - \sqrt{2d})]^2 - (\mathcal{R}(\psi))(x)\big| = 0$,
(VI) it holds for all $x \in \mathbb{R}$ with $\sqrt{2d} \le |x| \le R + \sqrt{2d}$ that

$$\big|[\mathfrak{R}(|x| - \sqrt{2d})]^2 - (\mathcal{R}(\psi))(x)\big| \le 4^{-M-1} R^2, \tag{5.75}$$

and
(VII) it holds for all $x \in \mathbb{R}$ with $|x| \ge R + \sqrt{2d}$ that

$$\big|[\mathfrak{R}(|x| - \sqrt{2d})]^2 - (\mathcal{R}(\psi))(x)\big| \le \big[|x| - \sqrt{2d}\big]^4 R^{-2} \tag{5.76}$$

(cf. Corollary 5.5), let $\lambda \in \mathbb{R}$ satisfy $\lambda = [\int_{\mathbb{R}^d} |g(y)|^2 \varphi(y)\, dy]^{-1/(4d)}$, and let $\mathscr{I}, \Psi, \Phi \in \mathbf{N}$ satisfy $\mathscr{I} = (\lambda \circledast \mathfrak{I}_1) \bullet \mathbf{A}_{\lambda,0}$, $\Psi = \mathfrak{S}_{1,d} \bullet \mathbf{P}_d(\psi, \psi, \ldots, \psi)$, and $\Phi = (\mathscr{I}^{\bullet d}) \bullet \Psi$ (cf. Definitions 2.1, 2.2, 2.4, 2.8, 2.10, 2.13, 2.15, 2.17, 2.19, 2.21, and 2.22). Note that Lemma 5.9 (applied with $d \curvearrowleft d$, $\varphi \curvearrowleft \varphi$, $g \curvearrowleft g$ in the notation of Lemma 5.9) implies that

$$\begin{aligned} 0 < \lambda = \left[\int_{\mathbb{R}^d} |g(y)|^2 \varphi(y)\, dy \right]^{-\frac{1}{4d}} &\le \left[50^{-1} d^{-3/2} e^{-d} \right]^{-\frac{1}{4d}} = \left[50 d^{3/2} e^d \right]^{\frac{1}{4d}} \\ &\le \left[64 d^2 4^d \right]^{\frac{1}{4d}} = \left[8d\, 2^d \right]^{\frac{1}{2d}} \le \left[8^d\, 2^d \right]^{\frac{1}{2d}} = \left[16^d \right]^{\frac{1}{2d}} = \left[4^{2d} \right]^{\frac{1}{2d}} = 4. \end{aligned} \tag{5.77}$$

This and Lemma 5.10 (applied with $n \curvearrowleft d$, $\lambda \curvearrowleft \lambda$, $\Phi \curvearrowleft \Psi$, $\mathscr{I} \curvearrowleft \mathscr{I}$, $\Psi \curvearrowleft \Phi$ in the notation of Lemma 5.10) ensure that for all $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ it holds that

$$(\mathcal{R}(\Phi))(x) = \lambda^{2d}(\mathcal{R}(\Psi))(x) = \left[\int_{\mathbb{R}^d} |g(y)|^2 \varphi(y)\, dy \right]^{-1/2} (\mathcal{R}(\Psi))(x). \tag{5.78}$$

Next observe that item (I), Lemma 2.20 (applied with $m \curvearrowleft 1$, $n \curvearrowleft d$ in the notation of Lemma 2.20), and Proposition 2.11 (applied with $n \curvearrowleft d$, $(\Phi_1, \Phi_2, \dots, \Phi_n) \curvearrowleft (\psi, \psi, \dots, \psi)$ in the notation of Proposition 2.11) assure that for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that $\mathcal{R}(\Psi) \in C(\mathbb{R}^d, \mathbb{R})$ and

$$(\mathcal{R}(\Psi))(x) = \sum_{j=1}^{d} (\mathcal{R}(\psi))(x_j). \tag{5.79}$$

Combining this with (5.78) establishes item (i). In the next step note that item (II), Lemma 2.20 (applied with $m \curvearrowleft 1$, $n \curvearrowleft d$ in the notation of Lemma 2.20), Proposition 2.12 (applied with $n \curvearrowleft d$, $(\Phi_1, \Phi_2, \dots, \Phi_n) \curvearrowleft (\psi, \psi, \dots, \psi)$ in the notation of Proposition 2.12), and Proposition 2.5 (applied with $\Phi_1 \curvearrowleft \mathfrak{S}_{1,d}$, $\Phi_2 \curvearrowleft \mathbf{P}_d(\psi, \psi, \dots, \psi)$ in the notation of Proposition 2.5) show that

$$\mathcal{D}(\mathbf{P}_d(\psi, \dots, \psi)) = (d, 2d, \underbrace{4d, \dots, 4d}_{M}, d) \in \mathbb{N}^{M+3} \tag{5.80}$$

and

$$\mathcal{D}(\Psi) = (d, 2d, \underbrace{4d, \dots, 4d}_{M}, 1) \in \mathbb{N}^{M+3}. \tag{5.81}$$

Therefore, we obtain that $\mathcal{H}(\Psi) = M + 1$ and

$$\begin{aligned} \mathcal{P}(\Psi) &= 2d(d+1) + 4d(2d+1) + \underbrace{4d(4d+1) + \ldots + 4d(4d+1)}_{M-1} + 1(4d+1) \\ &= 10d^2 + 10d + 1 + (M-1)(16d^2 + 4d) \leq 21d^2 M. \end{aligned} \tag{5.82}$$

Combining this with Lemma 5.10 (applied with $n \curvearrowleft d$, $\lambda \curvearrowleft \lambda$, $\Phi \curvearrowleft \Psi$, $\mathscr{I} \curvearrowleft \mathscr{I}$, $\Psi \curvearrowleft \Phi$ in the notation of Lemma 5.10) ensures that $\mathcal{H}(\Phi) = \mathcal{H}(\Psi) + d = d + M + 1$ and $\mathcal{P}(\Phi) \leq 2\mathcal{P}(\Psi) + 6d|\mathcal{O}(\Psi)|^2 \leq 42d^2 M + 6d$. This establishes items (ii) and (iii). Next observe that for all $\mathfrak{W} = (w_{i,j})_{(i,j)\in\{1,2,\dots,d\}\times\{1,2,\dots,4d\}} \in \mathbb{R}^{d\times 4d}$, $\mathfrak{B} = (b_1, b_2, \dots, b_d) \in \mathbb{R}^d$ it holds that

$$\underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}}_{\in \mathbb{R}^{1\times d}} \mathfrak{W} = \Big(\big[\textstyle\sum_{i=1}^{d} w_{i,1} \big], \big[\textstyle\sum_{i=1}^{d} w_{i,2} \big], \dots, \big[\textstyle\sum_{i=1}^{d} w_{i,4d} \big] \Big) \in \mathbb{R}^{1\times 4d} \tag{5.83}$$

and

$$\underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}}_{\in \mathbb{R}^{1\times d}} \mathfrak{B} + 0 = \big[\textstyle\sum_{i=1}^{d} b_i \big] \in \mathbb{R}. \tag{5.84}$$

The fact that $\|\mathcal{T}(\mathbf{P}_d(\psi, \psi, \dots, \psi))\|_\infty = \|\mathcal{T}(\psi)\|_\infty$ therefore implies that

$$\|\mathcal{T}(\Psi)\|_\infty = \|\mathcal{T}(\mathfrak{S}_{1,d} \bullet \mathbf{P}_d(\psi, \psi, \dots, \psi))\|_\infty \leq d\, \|\mathcal{T}(\mathbf{P}_d(\psi, \psi, \dots, \psi))\|_\infty = d\, \|\mathcal{T}(\psi)\|_\infty. \tag{5.85}$$

Combining this with item (III) assures that

$$\|\mathcal{T}(\Psi)\|_\infty \leq d\, \|\mathcal{T}(\psi)\|_\infty \leq d\big(\sqrt{2d} + 1\big) \max\{4, R^2\} \leq 3d^{3/2} \max\{4, R^2\}. \tag{5.86}$$

Lemma 5.10 (applied with $n \curvearrowleft d$, $\lambda \curvearrowleft \lambda$, $\Phi \curvearrowleft \Psi$, $\mathscr{I} \curvearrowleft \mathscr{I}$, $\Psi \curvearrowleft \Phi$ in the notation of Lemma 5.10) and (5.77) hence demonstrate that

$$\begin{aligned} \|\mathcal{T}(\Phi)\|_\infty &\le \max\{1, |\lambda|\} \max\{|\lambda|, \|\mathcal{T}(\Psi)\|_\infty\} \\ &\le 4 \max\Big\{4, 3d^{3/2} \max\{4, R^2\}\Big\} = 12 d^{3/2} \max\{4, R^2\}. \end{aligned} \tag{5.87}$$

This establishes item (iv). Moreover, note that the fact that for all $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_d \in \mathbb{R}$ it holds that $(\mathfrak{a}_1 + \mathfrak{a}_2 + \ldots + \mathfrak{a}_d)^2 \le d(|\mathfrak{a}_1|^2 + |\mathfrak{a}_2|^2 + \ldots + |\mathfrak{a}_d|^2)$ and (5.79) ensure that for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned} |(\mathcal{R}(\Psi))(x) - g(x)|^2 &= \left| \sum_{j=1}^d \Big[(\mathcal{R}(\psi))(x_j) - \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^2 \Big] \right|^2 \\ &\le d \sum_{j=1}^d \Big[(\mathcal{R}(\psi))(x_j) - \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^2 \Big]^2. \end{aligned} \tag{5.88}$$

Combining this with the fact that for all $k \in \mathbb{N}$ it holds that $\int_{\mathbb{R}^k} (2\pi)^{-k/2} e^{-\frac{1}{2}\|x\|_2^2}\, dx = 1$, Lemma 5.9 (applied with $d \curvearrowleft d$, $\varphi \curvearrowleft \varphi$, $g \curvearrowleft g$ in the notation of Lemma 5.9), and (5.78) implies that

$$\begin{aligned} &\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\, dx \\ &= \left[\int_{\mathbb{R}^d} |g(y)|^2 \varphi(y)\, dy \right]^{-1} \int_{\mathbb{R}^d} |(\mathcal{R}(\Psi))(x) - g(x)|^2 \varphi(x)\, dx \\ &\le 50 d^{3/2} e^d \int_{\mathbb{R}^d} |(\mathcal{R}(\Psi))(x) - g(x)|^2 \varphi(x)\, dx \\ &\le 50 d^{5/2} e^d \int_{\mathbb{R}^d} \left[\sum_{j=1}^d \Big[(\mathcal{R}(\psi))(x_j) - \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^2 \Big]^2 \right] \varphi(x_1, x_2, \ldots, x_d)\, d(x_1, x_2, \ldots, x_d) \\ &= 50 d^{5/2} e^d \left[\sum_{j=1}^d \int_{\mathbb{R}^d} \Big[(\mathcal{R}(\psi))(x_j) - \big[\mathfrak{R}\big(|x_j| - \sqrt{2d}\big)\big]^2 \Big]^2 \varphi(x_1, x_2, \ldots, x_d)\, d(x_1, x_2, \ldots, x_d) \right] \\ &= 50 d^{7/2} e^d \int_{\mathbb{R}^d} \Big[(\mathcal{R}(\psi))(x_1) - \big[\mathfrak{R}\big(|x_1| - \sqrt{2d}\big)\big]^2 \Big]^2 \varphi(x_1, x_2, \ldots, x_d)\, d(x_1, x_2, \ldots, x_d) \\ &= 25 \sqrt{\frac{2}{\pi}}\, d^{7/2} e^d \int_{\mathbb{R}} \Big[(\mathcal{R}(\psi))(x) - \big[\mathfrak{R}\big(|x| - \sqrt{2d}\big)\big]^2 \Big]^2 e^{-\frac{1}{2}x^2}\, dx. \end{aligned} \tag{5.89}$$

The integral transformation theorem and items (IV), (V), (VI), and (VII) therefore demonstrate that

$$\begin{aligned} &\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\, dx \\ &\le 25 \sqrt{\frac{2}{\pi}}\, d^{7/2} e^d \int_{\mathbb{R}} \Big[(\mathcal{R}(\psi))(x) - \big[\mathfrak{R}\big(|x| - \sqrt{2d}\big)\big]^2 \Big]^2 e^{-\frac{1}{2}x^2}\, dx \end{aligned}$$

$$
\begin{aligned}
&= 50\sqrt{\frac{2}{\pi}}\, d^{7/2} e^{d} \int_{\sqrt{2d}}^{\infty} \Big[(\mathcal{R}(\psi))(x) - \big[\mathfrak{R}\big(|x| - \sqrt{2d}\big)\big]^2\Big]^2 e^{-\frac{1}{2}x^2}\, dx \\
&= 50\sqrt{\frac{2}{\pi}}\, d^{7/2} e^{d} \int_{\sqrt{2d}}^{R+\sqrt{2d}} \Big[(\mathcal{R}(\psi))(x) - \big[\mathfrak{R}\big(|x| - \sqrt{2d}\big)\big]^2\Big]^2 e^{-\frac{1}{2}x^2}\, dx \\
&\quad + 50\sqrt{\frac{2}{\pi}}\, d^{7/2} e^{d} \int_{R+\sqrt{2d}}^{\infty} \Big[(\mathcal{R}(\psi))(x) - \big[\mathfrak{R}\big(|x| - \sqrt{2d}\big)\big]^2\Big]^2 e^{-\frac{1}{2}x^2}\, dx \\
&\le 50\sqrt{\frac{2}{\pi}}\, d^{7/2} e^{d} \left[4^{-2M-2} R^4 \int_{\sqrt{2d}}^{R+\sqrt{2d}} e^{-\frac{1}{2}x^2}\, dx + R^{-4} \int_{R+\sqrt{2d}}^{\infty} \big[x - \sqrt{2d}\big]^8 e^{-\frac{1}{2}x^2}\, dx \right] \\
&= 50\sqrt{\frac{2}{\pi}}\, d^{7/2} e^{d} \left[16^{-M-1} R^4 \int_{0}^{R} e^{-\frac{1}{2}(x^2 + 2x\sqrt{2d} + 2d)}\, dx + R^{-4} \int_{R}^{\infty} x^8 e^{-\frac{1}{2}(x^2 + 2x\sqrt{2d} + 2d)}\, dx \right] \\
&= 50\sqrt{\frac{2}{\pi}}\, d^{7/2} \left[16^{-M-1} R^4 \int_{0}^{R} e^{-\frac{1}{2}(x^2 + 2x\sqrt{2d})}\, dx + R^{-4} \int_{R}^{\infty} x^8 e^{-\frac{1}{2}(x^2 + 2x\sqrt{2d})}\, dx \right] \\
&\le 50\sqrt{\frac{2}{\pi}}\, d^{7/2} \left[16^{-M-1} R^4 \int_{0}^{\infty} e^{-\frac{1}{2}x^2}\, dx + R^{-4} \int_{0}^{\infty} x^8 e^{-\frac{1}{2}x^2}\, dx \right].
\end{aligned}
\tag{5.90}
$$

Next observe that the integral transformation theorem and Lemma 3.1 ensure that

$$
\begin{aligned}
\int_{0}^{\infty} x^8 e^{-\frac{1}{2}x^2}\, dx &= 8\sqrt{2} \int_{0}^{\infty} x^{7/2} e^{-x}\, dx = 8\sqrt{2}\, \Gamma\!\left(\frac{9}{2}\right) \\
&= 8\sqrt{2}\left[\frac{7}{2}\right]\left[\frac{5}{2}\right]\left[\frac{3}{2}\right]\left[\frac{1}{2}\right]\Gamma\!\left(\frac{1}{2}\right) = \frac{105\sqrt{\pi}}{\sqrt{2}}.
\end{aligned}
\tag{5.91}
$$

The fact that $\int_0^\infty e^{-\frac{1}{2}x^2} dx = \sqrt{\frac{\pi}{2}} \int_{\mathbb{R}} (2\pi)^{-1/2} e^{-\frac{1}{2}x^2} dx = \sqrt{\frac{\pi}{2}}$ and (5.90) therefore assure that

$$
\begin{aligned}
&\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\, dx \\
&\le 50\sqrt{\frac{2}{\pi}}\, d^{7/2} \left[16^{-M-1} R^4 \int_{0}^{\infty} e^{-\frac{1}{2}x^2}\, dx + R^{-4} \int_{0}^{\infty} x^8 e^{-\frac{1}{2}x^2}\, dx \right] \\
&= 50 d^{7/2} \left[16^{-M-1} R^4 + 105 R^{-4} \right].
\end{aligned}
\tag{5.92}
$$

This establishes item (v). The proof of Theorem 5.11 is thus complete. □

Corollary 5.12. *Let $\varepsilon \in (0, 1]$, $\mathfrak{C} \in [1000\varepsilon^{-1}, \infty)$, $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $d \in \mathbb{N}$, let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ and $g\colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $\varphi(x) = (2\pi)^{-d/2} \exp(-\frac{1}{2}(\sum_{j=1}^{d} |x_j|^2))$ and $g(x) = \sum_{j=1}^{d} [\max\{|x_j| - \sqrt{2d}, 0\}]^2$, and let $\mathfrak{g}\colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$ that $\mathfrak{g}(x) = [\int_{\mathbb{R}^d} |g(y)|^2 \varphi(y)\, dy]^{-1/2} g(x)$. Then* *there exists* $\Phi \in \mathbf{N}$ *such that* $\mathcal{I}(\Phi) = d$, $\mathcal{O}(\Phi) = 1$, $d \leq \mathcal{H}(\Phi) \leq \mathfrak{c}d$, $\|\mathcal{T}(\Phi)\|_\infty \leq \mathfrak{c}d^{\mathfrak{c}}$, $\mathcal{P}(\Phi) \leq \mathfrak{c}d^3$, *and* $[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\,dx]^{1/2} \leq \varepsilon$ *(cf. Definitions 2.2, 2.21, and 2.22).*

Proof of Corollary 5.12. Throughout this proof let $M \in \mathbb{N} \cap [2,\infty)$, $R \in [1,\infty)$ satisfy $M = \max((-\infty, R] \cap \mathbb{N})$ and $R = 9d\varepsilon^{-1/2}$. Note that Theorem 5.11 (applied with $d \curvearrowleft d$, $M \curvearrowleft M$, $R \curvearrowleft R$, $\varphi \curvearrowleft \varphi$, $g \curvearrowleft g$, $\mathfrak{g} \curvearrowleft \mathfrak{g}$ in the notation of Theorem 5.11) ensures that there exists $\Phi \in \mathbf{N}$ which satisfies that

(I) it holds that $\mathcal{R}(\Phi) \in C(\mathbb{R}^d, \mathbb{R})$,
(II) it holds that $\mathcal{H}(\Phi) = d + M + 1$,
(III) it holds that $\mathcal{P}(\Phi) \leq 42d^2M + 6d$,
(IV) it holds that $\|\mathcal{T}(\Phi)\|_\infty \leq 12d^{3/2}\max\{4, R^2\}$, and
(V) it holds that $\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\,dx \leq 50d^{7/2}\big[16^{-M-1}R^4 + 105R^{-4}\big]$

(cf. Definitions 2.2, 2.21, and 2.22). Therefore, we obtain that $\mathcal{I}(\Phi) = d$, $\mathcal{O}(\Phi) = 1$, $d \leq \mathcal{H}(\Phi) = d + M + 1 \leq d + R + 1 = d + 9d\varepsilon^{-1/2} + 1 \leq 11d\varepsilon^{-1/2} \leq \mathfrak{C}d \leq \mathfrak{c}d$, $\|\mathcal{T}(\Phi)\|_\infty \leq 12d^{3/2}\max\{4, R^2\} = 972d^{7/2}\varepsilon^{-1} \leq \mathfrak{C}d^{\mathfrak{C}} \leq \mathfrak{c}d^{\mathfrak{c}}$, and $\mathcal{P}(\Phi) \leq 42d^2M + 6d \leq 42d^2R + 6d = 378d^3\varepsilon^{-1/2} + 6d \leq 384d^3\varepsilon^{-1/2} \leq \mathfrak{C}d^3 \leq \mathfrak{c}d^3$. Moreover, observe that the fact that for all $x \in [4,\infty)$ it holds that $x^2 \leq 2^x$, the assumption that $M = \max((-\infty, R] \cap \mathbb{N})$, and the assumption that $R = 9d\varepsilon^{-1/2}$ show that $16^{-M-1}R^4 + 105R^{-4} \leq 106R^{-4} = 106\varepsilon^2(9d)^{-4} \leq (50d^{7/2})^{-1}\varepsilon^2$. Combining this with item (V) implies that $[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}(x)|^2 \varphi(x)\,dx]^{1/2} \leq [50d^{7/2}[16^{-M-1}R^4 + 105R^{-4}]]^{1/2} \leq \varepsilon$. The proof of Corollary 5.12 is thus complete. □

6. Lower and upper bounds for the number of ANN parameters in the approximation of high-dimensional functions

In this section we combine the lower bounds for the number of parameters of certain ANNs of Corollary 4.9 with the upper bounds of Corollary 5.12 to establish Theorem 6.1, the main ANN approximation result this paper. Theorem 1.1 is then an immediate consequence of Theorem 6.1, respectively Corollary 6.2.

6.1. ANN approximations with specifying the target functions

Theorem 6.1. *Let* $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, *and* $f_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, *satisfy for all* $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ *that* $\varphi_d(x) = (2\pi)^{-d/2}\exp(-\frac{1}{2}(\sum_{j=1}^d |x_j|^2))$ *and* $f_d(x) = \sum_{j=1}^d [\max\{|x_j| - \sqrt{2d}, 0\}]^2$, *let* $\mathfrak{f}_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, *satisfy for all* $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ *that* $\mathfrak{f}_d(x) = [\int_{\mathbb{R}^d} |f_d(y)|^2 \varphi_d(y)\,dy]^{-1/2} f_d(x)$, *and let* $\delta \in (0,1]$, $\varepsilon \in (0, 1/2]$. *Then there exists* $\mathfrak{C} \in (0,\infty)$ *such that*

(i) it holds for all $\mathfrak{c} \in [\mathfrak{C},\infty)$, $d \in \mathbb{N}$ *that*

$$\min\left\{ p \in \mathbb{N} \colon \begin{bmatrix} \exists\, \Phi \in \mathbf{N} \colon\ p = \mathcal{P}(\Phi),\ \mathcal{I}(\Phi) = d,\ \mathcal{O}(\Phi) = 1, \\ d \leq \mathcal{H}(\Phi) \leq \mathfrak{c}d,\ \|\mathcal{T}(\Phi)\|_\infty \leq \mathfrak{c}d^{\mathfrak{c}}, \\ [\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{f}_d(x)|^2 \varphi_d(x)\,dx]^{1/2} \leq \varepsilon \end{bmatrix} \right\} \leq \mathfrak{c}d^3 \tag{6.1}$$

and

(ii) it holds for all $\mathfrak{c} \in [\mathfrak{C},\infty)$, $d \in \mathbb{N}$ *that*

$$\min\left\{ p \in \mathbb{N} \colon \begin{bmatrix} \exists\, \Phi \in \mathbf{N} \colon\ p = \mathcal{P}(\Phi),\ \mathcal{I}(\Phi) = d,\ \mathcal{O}(\Phi) = 1, \\ \mathcal{H}(\Phi) \leq \mathfrak{c}d^{1-\delta},\ \|\mathcal{T}(\Phi)\|_\infty \leq \mathfrak{c}d^{\mathfrak{c}}, \\ [\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{f}_d(x)|^2 \varphi_d(x)\,dx]^{1/2} \leq \varepsilon \end{bmatrix} \right\} \geq (1 + \mathfrak{c}^{-3})^{(d^\delta)} \tag{6.2}$$

(cf. Definitions 2.2, 2.21, and 2.22).

Proof of Theorem 6.1. Throughout this proof let $\mathfrak{C} \in [100(\delta \ln(1.03))^{-2}, \infty) \cap [1000\varepsilon^{-1}, \infty)$, $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $d \in \mathbb{N}$ satisfy $2\mathfrak{C}^{5/\delta} \leq (1.03)^{\sqrt{\mathfrak{C}}}$. Note that Corollary 5.12 (applied with $\varepsilon \curvearrowleft \varepsilon$, $\mathfrak{C} \curvearrowleft \mathfrak{C}$, $\mathfrak{c} \curvearrowleft \mathfrak{c}$, $d \curvearrowleft d$, $\varphi \curvearrowleft \varphi_d$, $g \curvearrowleft f_d$, $\mathfrak{g} \curvearrowleft \mathfrak{f}_d$ in the notation of Corollary 5.12) assures that there exists $\Phi \in \mathbf{N}$ such that $\mathcal{I}(\Phi) = d$, $\mathcal{O}(\Phi) = 1$, $d \leq \mathcal{H}(\Phi) \leq \mathfrak{c}d$, $\|\mathcal{T}(\Phi)\|_\infty \leq \mathfrak{c}d^{\mathfrak{c}}$, $\mathcal{P}(\Phi) \leq \mathfrak{c}d^3$, and $[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{g}_d(x)|^2 \varphi_d(x)\,dx]^{1/2} \leq \varepsilon$ (cf. Definitions 2.2, 2.21, and 2.22). This establishes item (i). Moreover, observe that Corollary 4.9 (applied with $\varphi_d \curvearrowleft \varphi_d$, $g_d \curvearrowleft f_d$, $\mathfrak{g}_d \curvearrowleft \mathfrak{f}_d$, $\delta \curvearrowleft \delta$, $\mathfrak{C} \curvearrowleft \mathfrak{C}$ in the notation of Corollary 4.9) ensures that for all $\Phi \in \mathbf{N}$ with $\mathcal{I}(\Phi) = d$, $\mathcal{O}(\Phi) = 1$, $\mathcal{H}(\Phi) \leq \mathfrak{c}d^{1-\delta}$, $\|\mathcal{T}(\Phi)\|_\infty \leq \mathfrak{c}d^{\mathfrak{c}}$, and $[\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{f}_d(x)|^2 \varphi_d(x)\,dx]^{1/2} \leq \varepsilon$ it holds that $\mathcal{P}(\Phi) \geq (1 + \mathfrak{c}^{-3})^{d^\delta}$. This establishes item (ii). The proof of Theorem 6.1 is thus complete. □

6.2. ANN approximations without specifying the target functions

Corollary 6.2. *Let $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $\varphi_d(x) = (2\pi)^{-d/2} \exp(-\frac{1}{2}\|x\|_2^2)$ (cf. Definition 2.21). Then there exist continuously differentiable $\mathfrak{f}_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, such that for all $\delta \in (0, 1]$, $\varepsilon \in (0, 1/2]$ there exists $\mathfrak{C} \in (0, \infty)$ such that*

(i) it holds for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $d \in \mathbb{N}$ that

$$\min\left\{ p \in \mathbb{N} \colon \begin{bmatrix} \exists\, \Phi \in \mathbf{N} \colon\ p = \mathcal{P}(\Phi),\ \mathcal{I}(\Phi) = d,\ \mathcal{O}(\Phi) = 1, \\ d \leq \mathcal{H}(\Phi) \leq \mathfrak{c}d,\ \|\mathcal{T}(\Phi)\|_\infty \leq \mathfrak{c}d^{\mathfrak{c}}, \\ [\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{f}_d(x)|^2 \varphi_d(x)\,dx]^{1/2} \leq \varepsilon \end{bmatrix} \right\} \leq \mathfrak{c}d^3 \tag{6.3}$$

and

(ii) it holds for all $\mathfrak{c} \in [\mathfrak{C}, \infty)$, $d \in \mathbb{N}$ that

$$\min\left\{ p \in \mathbb{N} \colon \begin{bmatrix} \exists\, \Phi \in \mathbf{N} \colon\ p = \mathcal{P}(\Phi),\ \mathcal{I}(\Phi) = d,\ \mathcal{O}(\Phi) = 1, \\ \mathcal{H}(\Phi) \leq \mathfrak{c}d^{1-\delta},\ \|\mathcal{T}(\Phi)\|_\infty \leq \mathfrak{c}d^{\mathfrak{c}}, \\ [\int_{\mathbb{R}^d} |(\mathcal{R}(\Phi))(x) - \mathfrak{f}_d(x)|^2 \varphi_d(x)\,dx]^{1/2} \leq \varepsilon \end{bmatrix} \right\} \geq (1 + \mathfrak{c}^{-3})^{(d^\delta)} \tag{6.4}$$

(cf. Definitions 2.2 and 2.22).

Proof of Corollary 6.2. Throughout this proof let $f_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $f_d(x) = \sum_{j=1}^d [\max\{|x_j| - \sqrt{2d}, 0\}]^2$, let $\mathfrak{f}_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $\mathfrak{f}_d(x) = f_d(x)[\int_{\mathbb{R}^d} |f_d(y)|^2 \varphi_d(y)\,dy]^{-1/2}$, and let $\delta \in (0, 1]$, $\varepsilon \in (0, 1/2]$. Note that Theorem 6.1 (applied with $(\varphi_d)_{d \in \mathbb{N}} \curvearrowleft (\varphi_d)_{d \in \mathbb{N}}$, $(f_d)_{d \in \mathbb{N}} \curvearrowleft (f_d)_{d \in \mathbb{N}}$, $(\mathfrak{f}_d)_{d \in \mathbb{N}} \curvearrowleft (\mathfrak{f}_d)_{d \in \mathbb{N}}$, $\delta \curvearrowleft \delta$, $\varepsilon \curvearrowleft \varepsilon$ in the notation of Theorem 6.1) establishes items (i) and (ii). The proof of Corollary 6.2 is thus complete. □

Acknowledgements

Benno Kuckuck and Philippe von Wurstemberger are gratefully acknowledged for their helpful assistance regarding Lemma 2.23. Joshua Lee Padgett is gratefully acknowledged for his helpful assistance regarding Lemma 5.1 and Lemma 5.2. The third author gratefully acknowledges the Cluster of Excellence EXC 2044-390685587, Mathematics Münster: Dynamics-Geometry-Structure funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) and the startup fund project of Shenzhen Research Institute of Big Data under grant No. T00120220001. The fourth author acknowledges funding by the Austrian Science Fund (FWF) through the projects P 30148 and I 3403.

References

[1] G.E. Andrews, R. Askey, R. Roy, Special Functions, Encyclopedia of Mathematics and Its Applications, vol. 71, Cambridge University Press, Cambridge, 1999.

[2] A. Barron, Neural net approximation, in: Proceedings of the 7th Yale Workshop on Adaptive and Learning Systems, 1992, pp. 69–72.
[3] A. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Trans. Inf. Theory 39 (3) (1993) 930–945.
[4] A. Barron, Approximation and estimation bounds for artificial neural networks, Mach. Learn. 14 (1) (Jan 1994) 115–133.
[5] C. Beck, M. Hutzenthaler, A. Jentzen, B. Kuckuck, An overview on deep learning-based approximation methods for partial differential equations, Discrete Contin. Dyn. Syst., Ser. B 28 (6) (2023) 3697–3746.
[6] C. Beck, A. Jentzen, B. Kuckuck, Full error analysis for the training of deep neural networks, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 25 (2) (2022) 2150020.
[7] R. Bellman, Dynamic Programming, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 2010, Reprint of the 1957 edition, With a new introduction by Stuart Dreyfus.
[8] P. Beneventano, P. Cheridito, A. Jentzen, P. von Wurstemberger, High-dimensional approximation spaces of artificial neural networks and applications to partial differential equations, arXiv:2012.04326, 2020, 32 pages.
[9] J. Berner, P. Grohs, A. Jentzen, Analysis of the generalization error: empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations, SIAM J. Math. Data Sci. 3 (2) (2020) 631–657.
[10] M. Bianchini, F. Scarselli, On the complexity of neural network classifiers: a comparison between shallow and deep architectures, IEEE Trans. Neural Netw. Learn. Syst. 25 (8) (2014) 1553–1565.
[11] H. Bölcskei, P. Grohs, G. Kutyniok, P. Petersen, Optimal approximation with sparsely connected deep neural networks, SIAM J. Math. Data Sci. 1 (1) (2019) 8–45.
[12] P. Cheridito, A. Jentzen, F. Rossmannek, Efficient approximation of high-dimensional functions with neural networks, IEEE Trans. Neural Netw. Learn. Syst. 33 (7) (2022) 3079–3093.
[13] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Syst. 4 (1989) 303–314.
[14] A. Daniely, Depth separation for neural networks, in: S. Kale, O. Shamir (Eds.), Proceedings of the 2017 Conference on Learning Theory, PMLR, Amsterdam, Netherlands, 07–10 Jul 2017, in: Proceedings of Machine Learning Research, vol. 65, 2017, pp. 690–696.
[15] M.J. Donahue, C. Darken, L. Gurvits, E. Sontag, Rates of convex approximation in non-Hilbert spaces, Constr. Approx. 13 (2) (1997) 187–220.
[16] W. E, J. Han, A. Jentzen, Algorithms for solving high dimensional PDEs: from nonlinear Monte Carlo to machine learning, Nonlinearity 35 (1) (2022) 278–310.
[17] D. Elbrächter, P. Grohs, A. Jentzen, C. Schwab, DNN expression rate analysis of high-dimensional PDEs: application to option pricing, Constr. Approx. 55 (1) (2022) 3–71.
[18] D. Elbrächter, D. Perekrestenko, P. Grohs, H. Bölcskei, Deep Neural network approximation theory, arXiv:1901.02220, 2020, 74 pages.
[19] R. Eldan, O. Shamir, The power of depth for feedforward neural networks, in: V. Feldman, A. Rakhlin, O. Shamir (Eds.), 29th Annual Conference on Learning Theory, Columbia University New York, New York, USA, 23–26 Jun 2016, in: Proceedings of Machine Learning Research, PMLR, vol. 49, 2016, pp. 907–940.
[20] K.-I. Funahashi, On the approximate realization of continuous mappings by neural networks, Neural Netw. 2 (3) (1989) 183–192.
[21] F. Girosi, G. Anzellotti, Rates of convergence for radial basis functions and neural networks, in: R.J. Mammone (Ed.), Artificial Neural Networks for Speech and Vision, Chapman & Hall, 1993, pp. 97–113.
[22] L. Gonon, P. Grohs, A. Jentzen, D. Kofler, D. Šiška, Uniform error estimates for artificial neural network approximations for heat equations, IMA J. Numer. Anal. 42 (3) (2022) 1991–2054.
[23] L. Gonon, C.Schwab, Deep ReLU network expression rates for option prices in high-dimensional, exponential Lévy models, Tech. Rep. 2020-52, Seminar for Applied Mathematics, ETH, Zürich, Switzerland, 2020.
[24] P. Grohs, L. Herrmann, Deep neural network approximation for high-dimensional elliptic PDEs with boundary conditions, arXiv:2007.05384, 2020, 22 pages.
[25] P. Grohs, F. Hornung, A. Jentzen, P. von Wurstemberger, A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations, Mem. Amer. Math. Soc. (2018), in press, arXiv:1809.02362, 124 pages.
[26] P. Grohs, F. Hornung, A. Jentzen, P. Zimmermann, Space-time error estimates for deep neural network approximations for differential equations, Adv. Comput. Math. 49 (1) (2023) 4.
[27] P. Grohs, A. Jentzen, D. Salimova, Deep neural network approximations for solutions of PDEs based on Monte Carlo algorithms, Part. Differ. Equ. Appl. 3 (4) (2022) 45.
[28] N.J. Guliyev, V.E. Ismailov, Approximation capability of two hidden layer feedforward neural networks with fixed weights, Neurocomputing 316 (2018) 262–269.
[29] N.J. Guliyev, V.E. Ismailov, On the approximation by single hidden layer feedforward neural networks with fixed weights, Neural Netw. 98 (2018) 296–304.
[30] L. Gurvits, P. Koiran, Approximation and learning of convex superpositions, J. Comput. Syst. Sci. 55 (1) (1997) 161–170.
[31] K. Hornik, Approximation capabilities of multilayer feedforward networks, Neural Netw. 2 (4) (1991) 251–257.
[32] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Netw. 5 (2) (1989) 359–366.
[33] F. Hornung, A. Jentzen, D. Salimova, Space-time deep neural network approximations for high-dimensional partial differential equations, arXiv:2006.02199, 2020, 52 pages.
[34] M. Hutzenthaler, A. Jentzen, T. Kruse, T.A. Nguyen, A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations, Ser. Partial Differ. Equ. Appl. 1 (2020) 1–34.

[35] A. Jentzen, D. Salimova, T. Welti, A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients, Commun. Math. Sci. 19 (5) (2021) 1167–1205.
[36] L.K. Jones, A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training, Ann. Stat. 20 (1) (1992) 608–613.
[37] P.C. Kainen, V. Kůrková, M. Sanguineti, Complexity of Gaussian-radial-basis networks approximating smooth functions, J. Complex. 25 (1) (2009) 63–74.
[38] P.C. Kainen, V. Kůrková, M. Sanguineti, Dependence of computational models on input dimension: tractability of approximation and optimization tasks, IEEE Trans. Inf. Theory 58 (2) (Feb 2012) 1203–1214.
[39] A. Klenke, Probability Theory, Universitext, Springer-Verlag London Ltd., London, 2008, A comprehensive course, Translated from the 2006 German original.
[40] J.M. Klusowski, A.R. Barron, Approximation by combinations of ReLU and squared ReLU ridge functions with ℓ^1 and ℓ^0 controls, IEEE Trans. Inf. Theory 64 (12) (Dec 2018) 7649–7656.
[41] V. Kůrková, Minimization of error functionals over perceptron networks, Neural Comput. 20 (1) (Jan 2008) 252–270.
[42] V. Kůrková, P.C. Kainen, V. Kreinovich, Estimates of the number of hidden units and variation with respect to half-spaces, Neural Netw. 10 (6) (1997) 1061–1068.
[43] V. Kůrková, M. Sanguineti, Comparison of worst case errors in linear and neural network approximation, IEEE Trans. Inf. Theory 48 (1) (Jan 2002) 264–275.
[44] V. Kůrková, M. Sanguineti, Geometric upper bounds on rates of variable-basis approximation, IEEE Trans. Inf. Theory 54 (12) (Dec 2008) 5681–5688.
[45] G. Kutyniok, P. Petersen, M. Raslan, R. Schneider, A theoretical analysis of deep neural networks and parametric PDEs, arXiv:1904.00377, 2019, 39 pages.
[46] M. Leshno, V.Y. Lin, A. Pinkus, S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural Netw. 6 (6) (1993) 861–867.
[47] B. Li, S. Tang, H. Yu, Better approximations of high dimensional smooth functions by deep neural networks with rectified power units, arXiv:1903.05858, 2019, 28 pages.
[48] V. Maiorov, A. Pinkus, Lower bounds for approximation by MLP neural networks, Neurocomputing 25 (1) (1999) 81–91.
[49] H.N. Mhaskar, Q. Liao, T.A. Poggio, Learning functions: when is deep better than shallow, CoRR, arXiv:1603.00988v4 [abs], 2016.
[50] H.N. Mhaskar, T.A. Poggio, Deep vs. shallow networks: an approximation theory perspective, CoRR, arXiv:1608.03287 [abs], 2016.
[51] E. Novak, H. Woźniakowski, Tractability of Multivariate Problems. Vol. 1: Linear Information, EMS Tracts in Mathematics, vol. 6, European Mathematical Society (EMS), Zürich, 2008.
[52] E. Novak, H. Woźniakowski, Tractability of Multivariate Problems. Volume II: Standard Information for Functionals, EMS Tracts in Mathematics, vol. 12, European Mathematical Society (EMS), Zürich, 2010.
[53] P. Petersen, F. Voigtlaender, Optimal approximation of piecewise smooth functions using deep ReLU neural networks, Neural Netw. 108 (2018) 296–330.
[54] A. Pinkus, Approximation theory of the MLP model in neural networks, Acta Numer. 8 (1999) 143–195.
[55] C. Reisinger, Y. Zhang, Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems, arXiv:1903.06652, 2019, 39 pages.
[56] H. Robbins, A remark on Stirling's formula, Am. Math. Mon. 62 (1955) 26–29.
[57] I. Safran, O. Shamir, Depth-width tradeoffs in approximating natural functions with neural networks, in: D. Precup, Y.W. Teh (Eds.), Proceedings of the 34th International Conference on Machine Learning, International Convention Centre, Sydney, Australia, 06–11 Aug 2017, in: Proceedings of Machine Learning Research, PMLR, vol. 70, 2017, pp. 2979–2987.
[58] E.T. Whittaker, G.N. Watson, An introduction to the general theory of infinite processes and of analytic functions; with an account of the principal transcendental functions, in: A Course of Modern Analysis, in: Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1996, Reprint of the fourth (1927) edition.
[59] D. Yarotsky, Error bounds for approximations with deep ReLU networks, Neural Netw. 94 (2017) 103–114.