A multivariate Riesz basis of ReLU neural networks

Cornelia Schneider* and Jan Vybíral†

March 2, 2023

Abstract

We consider the trigonometric-like system of piecewise linear functions introduced recently by Daubechies, DeVore, Foucart, Hanin, and Petrova. We provide an alternative proof that this system forms a Riesz basis of $L_2([0,1])$ based on the Gershgorin theorem. We also generalize this system to higher dimensions $d > 1$ by a construction, which avoids using (tensor) products. As a consequence, the functions from the new Riesz basis of $L_2([0,1]^d)$ can be easily represented by neural networks. Moreover, the Riesz constants of this system are independent of d, making it an attractive building block regarding future multivariate analysis of neural networks.

Key Words: Riesz basis, Rectified Linear Unit (ReLU), artificial neural networks, Euler product, Möbius function

MSC2020 Math Subject Classifications: 68T07, 42C15, 11A25.

1 Introduction

The last decades observed a tremendous success of artificial neural networks in many machine learning tasks, including computer vision [16], speech recognition [11], natural language processing [24], or games solutions [17, 21] to name just few. Despite their wide use, many of their properties are not fully understood and many aspects of their great practical performance lack a rigorous explanation. Without any doubt, a deeper insight into the theory of artificial neural networks could boost their applicability even further.

In the last decade, a growing number of authors investigated why the deep neural networks with a higher number of hidden layers approximate many interesting functions more efficiently than the shallow neural networks (with only one hidden layer) using the same number of parameters. We refer to [1, 7, 8, 10, 12, 18, 22, 25] for a number of mathematically rigorous results in this direction.

One of the astonishing properties of artificial neural networks is, that they can approximate extremely well also functions of many variables, which often allows to avoid the curse of dimensionality [3, 9, 19]. The aim of this work is to shed new light on the effectiveness of artificial neural networks for the approximation of multivariate functions by constructing a new system of functions, which forms a Riesz basis of $L_2([0,1]^d)$ for every $d \geq 1$.

To state the result, we first recall the notion of a Riesz basis (which, in turn, is a generalization of an orthonormal system and of an orthonormal basis, cf. [6]).

Definition 1.1. *Let H be a real Hilbert space. The (finite or infinite) sequence $(x_n)_n \subset H$ is called a Riesz sequence if there are two constants $A, B > 0$ such that*

$$A \sum_n \alpha_n^2 \leq \left\| \sum_n \alpha_n x_n \right\|^2 \leq B \sum_n \alpha_n^2 \tag{1}$$

*Friedrich-Alexander Universität Erlangen, Applied Mathematics III, Cauerstr. 11, 91058 Erlangen, Germany. Email: cornelia.schneider@math.fau.de

†Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Trojanova 13, 12000 Praha, Czech Republic. Email: jan.vybiral@fjfi.cvut.cz. The work of this author has been supported by the grant P202/23/04720S of the Grant Agency of the Czech Republic

for every real square summable sequence $(\alpha_n)_n$. If the closed span of $(x_n)_n$ is the whole space H, then we call it a Riesz basis.

The system, which we study in this paper, is a trigonometric-like basis, where instead of cos and sin functions we use their piecewise linear counterparts $\mathcal{C}$ and $\mathcal{S}$, which are defined as follows (cf. Figure 1).

Definition 1.2. *1. For $x \in [0,1]$, we define*

$$\mathcal{C}(x) = 4\left|x - \frac{1}{2}\right| - 1 = \begin{cases} 1 - 4x, & x \in [0, 1/2), \\ 4x - 3, & x \in [1/2, 1] \end{cases}$$

and

$$\mathcal{S}(x) = \left|2 - 4\left|x - \frac{1}{4}\right|\right| - 1 = \begin{cases} 4x, & x \in [0, 1/4), \\ 2 - 4x, & x \in [1/4, 3/4), \\ 4x - 4, & x \in [3/4, 1]. \end{cases}$$

2. For $x \in \mathbb{R}$, we extend this definition periodically, i.e. $\mathcal{C}(x) = \mathcal{C}(x - \lfloor x \rfloor)$ and $\mathcal{S}(x) = \mathcal{S}(x - \lfloor x \rfloor)$.

3. If $k \geq 1$ and $x \in \mathbb{R}$, we put $\mathcal{C}_k(x) = \mathcal{C}(kx)$ and $\mathcal{S}_k(x) = \mathcal{S}(kx)$.

Figure 1: The plot of $\mathcal{C}$, $\mathcal{S}$, $\mathcal{C}_2$ and $\mathcal{S}_2$.
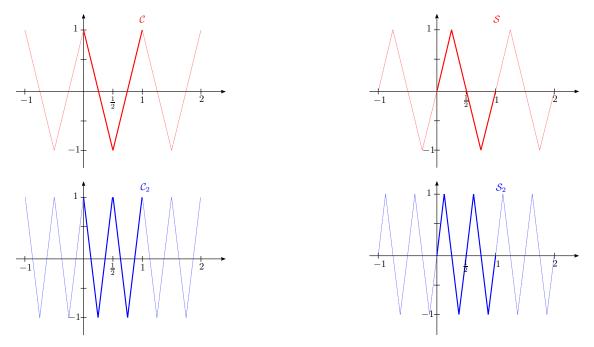
These functions were introduced and studied in [7], where it was shown that the system
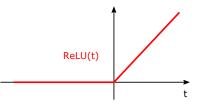
$$\overline{\mathcal{R}}_1 := \{1\} \cup \{\sqrt{3}\,\mathcal{C}_k, \sqrt{3}\,\mathcal{S}_k : k \in \mathbb{N}\} \tag{2}$$

forms a Riesz basis of $L_2([0,1])$ with the constants $A = 1/2$ and $B = 3/2$. Let us note that the factor $\sqrt{3}$ in (2) is simply a normalization factor, which ensures that all the elements of $\overline{\mathcal{R}}_1$ have unit norm in $L_2([0,1])$. The proof given in [7] is implicitly inspired by the method of analysis and synthesis operators, respectively, used in the frame theory, cf. [13]. It is one of the aims of our paper (cf. Theorem 2.2) to provide an alternative proof, which first reduces (1) to the study of spectral properties of the Gram matrix of $\overline{\mathcal{R}}_1$. The result then follows from the Gershgorin circle theorem and some elementary number theory (including Euler products and a certain Ramanujan's formula).

The main advantage of (2) in contrast to the standard trigonometric system is, that its elements can be easily identified by artificial neural networks with the REctified Linear Unit (ReLU) activation function.

Let us recall, that if $t \in \mathbb{R}$, then the ReLU function is defined as $\text{ReLU}(t) = \max(0, t)$, cf. Figure 2. On vectors, it acts component-wise

$$\text{ReLU}(x_1, \ldots, x_n) = (\text{ReLU}(x_1), \ldots, \text{ReLU}(x_n)), \quad x \in \mathbb{R}^n.$$

Figure 2: Graph of the ReLU function.

If $W \geq 2$ and $L \geq 1$ are integer parameters, then we denote by $\Upsilon^{W,L}$ the real-valued functions which can be represented by a ReLU neural network of width W and depth L, see Definition 4.1 for a precise formulation. Then [7, Theorem 6.2] shows that $\mathcal{C}_j$ and $\mathcal{S}_j$ (restricted to $[0,1]$) lie in $\Upsilon^{W,L}$ for an arbitrary $W \geq 6$ and L of the asymptotic order $\log_2(j)$.

The main aim of our work is to generalize the results of [7] to the multivariate case. There are two crucial issues which prevent us from simply taking the tensor products of the functions in $\overline{\mathcal{R}}_1$. First, the product function $(x, y) \to x \cdot y$ can only be approximated by the ReLU neural networks and, second, the ratio of the Riesz constants B and A gets exponentially large when d grows.

We propose a surprisingly simple and effective solution to these challenges. We show that (cf. Theorem 3.3) the multivariate analogue of $\overline{\mathcal{R}}_1$

$$\{1\} \cup \{\sqrt{3}\,\mathcal{C}(\alpha \cdot x) : \alpha \rhd 0\} \cup \{\sqrt{3}\,\mathcal{S}(\alpha \cdot x) : \alpha \rhd 0\}, \tag{3}$$

forms a Riesz basis of $L_2([0,1]^d)$ for every $d \geq 1$ with the same constants $A = 1/2$ and $B = 3/2$ as in the univariate case. Here, $\alpha \rhd 0$ means that the first non-zero entry of $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{Z}^d$ is positive. Finally, we show in Section 4 that the functions from (3) can be exactly reproduced by the ReLU neural networks of width W and depth L in essentially the same way as in the univariate case, i.e., there is virtually no price to pay when d grows.

2 Univariate case

It was observed in [7, Section 6], that the system of piecewise linear functions

$$\{\mathcal{C}_k, \mathcal{S}_k : k \in \mathbb{N}\}, \tag{4}$$

on one hand shares some nice properties with the trigonometric system and on the other hand can be easily reproduced by artificial neural networks with the ReLU activation function. The aim of this section is to essentially reprove Proposition 6.1 of [7], which states that this system is a Riesz basis of $L_2^0([0,1])$, the space of square integrable functions with mean zero.

Although our proof shares some technical details with [7], its main structure is different: In particular, we reduce the problem to spectral properties of the corresponding Gram matrix and then apply the Gershgorin circle theorem. Interestingly, by using some elementary number theory, we are able to completely characterize the inner products of the $\mathcal{C}_i$ and/or $\mathcal{S}_j$ functions. In contrast to [7], we complement (4) by adding the constant function, which is orthogonal to all functions from (4).

We denote by $\gcd(i, j)$ the greatest common divisor of i and j and by $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$ the standard inner product in $L_2([0,1])$.

Lemma 2.1. *Let $i, j \in \mathbb{N}$. Then*

1. *$\langle \mathcal{C}_i, \mathcal{S}_j \rangle = 0$;*

2. *$\langle \mathcal{C}_i, \mathcal{C}_j \rangle = \langle \mathcal{S}_i, \mathcal{S}_j \rangle = 0$ if $i/\gcd(i,j)$ is odd and $j/\gcd(i,j)$ is even (or vice versa), i.e., if the prime factorizations of i and j contain a different power of 2;*

3. *If $i/\gcd(i,j)$ and $j/\gcd(i,j)$ are both odd, then*

$$3 \cdot \langle \mathcal{C}_i, \mathcal{C}_j \rangle = 3 \cdot |\langle \mathcal{S}_i, \mathcal{S}_j \rangle| = \frac{\gcd(i,j)^4}{i^2 \cdot j^2}.$$

Here, the sign of $\langle \mathcal{S}_i, \mathcal{S}_j \rangle$ is negative if, and only if, $(i + j)/(2\gcd(i,j))$ is even.

3

4. In particular, we get $\langle \mathcal{C}_i, \mathcal{C}_i \rangle = \langle \mathcal{S}_i, \mathcal{S}_i \rangle = 1/3$ for all $i \in \mathbb{N}$.

Proof. We transfer the proof to the Fourier side by exploiting the decomposition of $\mathcal{C}_i$ and $\mathcal{S}_j$ into Fourier series, cf. [7, page 166]. Let

$$c_k(x) = \sqrt{2}\cos(2\pi kx), \quad s_k(x) = \sqrt{2}\sin(2\pi kx), \quad k \in \mathbb{N}_0, \ x \in \mathbb{R}.$$

Then a standard calculation reveals that

$$\sqrt{3}\,\mathcal{C}_k = \mu \sum_{m \geq 0} \frac{1}{(2m+1)^2} c_{(2m+1)k} \quad \text{and} \quad \sqrt{3}\,\mathcal{S}_k = \mu \sum_{m \geq 0} \frac{(-1)^m}{(2m+1)^2} s_{(2m+1)k}, \tag{5}$$

where

$$\mu^2 \sum_{m \geq 0} \frac{1}{(2m+1)^4} = 1, \quad \text{i.e.} \quad \mu^2 \frac{\pi^4}{96} = 1. \tag{6}$$

Using (5), we immediately obtain that $\langle \mathcal{C}_i, \mathcal{S}_j \rangle = 0$. Furthermore,

$$3\langle \mathcal{C}_i, \mathcal{C}_j \rangle = \sum_{m,n=0}^{\infty} \frac{\mu^2}{(2m+1)^2(2n+1)^2} \langle c_{(2m+1)i}, c_{(2n+1)j} \rangle$$

$$= \sum_{m,n=0}^{\infty} \frac{\mu^2}{(2m+1)^2(2n+1)^2} \delta_{(2m+1)i,(2n+1)j}, \tag{7}$$

where $\delta_{u,v} = 1$ if $u = v$ and zero otherwise. To simplify (7), we have to find for fixed $i, j \in \mathbb{N}$ all $m, n \in \mathbb{N}_0$ such that $(2m+1)i = (2n+1)j$. First, we observe that if the prime factorizations of i and j contain a different power of two, then also $(2m+1)i$ and $(2n+1)j$ have a different power of two in their prime factorizations and therefore they differ for all $m, n \in \mathbb{N}_0$. Consequently, (7) shows that $3\langle \mathcal{C}_i, \mathcal{C}_j \rangle = 0$.

If the prime factorizations of i and j contain the same power of two, then $i/\gcd(i,j)$ and $j/\gcd(i,j)$ are both odd. We denote $g = \gcd(i,j)$ and note that i/g and j/g are coprime, i.e., that their greatest common divisor is one. We then look for all pairs $(m,n) \in \mathbb{N}_0^2$, which solve the equation

$$(2m+1) \cdot g \cdot \frac{i}{g} = (2n+1) \cdot g \cdot \frac{j}{g}.$$

All the solutions are obtained in the form

$$2m+1 = \frac{j}{g} \cdot (2l+1), \quad 2n+1 = \frac{i}{g} \cdot (2l+1), \quad l \in \mathbb{N}_0. \tag{8}$$

We insert (8) into (7) and conclude that

$$3\langle \mathcal{C}_i, \mathcal{C}_j \rangle = \sum_{l=0}^{\infty} \frac{\mu^2}{\left(\frac{i}{g}(2l+1) \right)^2 \cdot \left(\frac{j}{g}(2l+1) \right)^2}$$

$$= \mu^2 \sum_{l=0}^{\infty} \frac{1}{(2l+1)^4} \cdot \frac{1}{(i/g)^2 \cdot (j/g)^2} = \frac{1}{(i/g)^2 \cdot (j/g)^2}.$$

The calculation of $\langle \mathcal{S}_i, \mathcal{S}_j \rangle$ can be performed in a very similar way, one only needs to take care about the sign of the inner product. In particular, instead of (7) we obtain

$$3\langle \mathcal{S}_i, \mathcal{S}_j \rangle = \sum_{m,n=0}^{\infty} \frac{\mu^2(-1)^{n+m}}{(2m+1)^2(2n+1)^2} \delta_{(2m+1)i,(2n+1)j},$$

where $(-1)^{n+m} = -1$ if, and only if, $n + m = \frac{i+j}{2g}(2l+1) - 1$ is odd. Therefore, the sign is negative if, and only if, $\frac{i+j}{2g}$ even. $\qquad\square$

Next, we combine Lemma 2.1 with the Gershgorin circle theorem and provide an alternative proof of [7, Theorem 6.2], which we restate as follows.

Theorem 2.2. *The system $\mathcal{R}_1 := \{1\} \cup \{\mathcal{C}_k, \mathcal{S}_k : k \in \mathbb{N}\}$ is a Riesz basis of $L_2([0,1])$.*

Before we come to the proof, several remarks seem to be in order.

Remark 2.3.   *1. We will actually show, that the Riesz constants of the L_2-normalized system $\overline{\mathcal{R}}_1 := \{1\} \cup \{\sqrt{3}\,\mathcal{C}_k, \sqrt{3}\,\mathcal{S}_k : k \in \mathbb{N}\}$ can be chosen as $A = 1/2$ and $B = 3/2$.*

2. We divide the proof of Theorem 2.2 into several steps. In the first two steps, we show that the truncated system $\overline{\mathcal{R}}_1^N := \{1\} \cup \{\sqrt{3}\,\mathcal{C}_k, \sqrt{3}\,\mathcal{S}_k : k \leq N\}$ forms a Riesz sequence (i.e., that it satisfies (1)) with $A = 1/2$ and $B = 3/2$ being independent of $N \in \mathbb{N}$. The first step reduces this question to spectral properties of the Gram matrix of $\overline{\mathcal{R}}_1^N$ and in the second step we apply the Gershgorin theorem to bound this spectrum. The third step describes how we pass to the limit $N \to \infty$ to deduce that $\overline{\mathcal{R}}_1$ is also a Riesz sequence. Finally, the fourth step shows that $\overline{\mathcal{R}}_1$ is also a basis, i.e., that its closed linear span is $L_2([0,1])$.

3. Our proof shows that the spectrum of the Gram matrix of $\overline{\mathcal{R}}_1^N$ (for arbitrary $N \in \mathbb{N}$) is contained in $[1/2, 3/2]$. We leave it as an open problem to find out if these bounds are actually optimal. Supported by numerical evidence (see Figure 3 for details), our conjecture is that there is indeed some space for improvement.
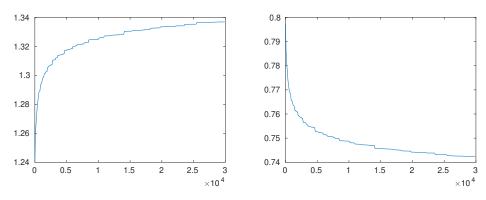
Figure 3: The largest (left) and the smallest (right) eigenvalue of $(\langle \mathcal{C}_i, \mathcal{C}_j \rangle)_{i,j=1}^N$ for $1 \leq N \leq 3 \cdot 10^4$.

Proof of Theorem 2.2: Step 1.
First we reformulate the definition of a (finite) Riesz sequence as an eigenvalue problem of its Gram matrix. This reformulation is rather straightforward and by no means new, see [15] or [26, Chapter 1.8]. Let H be a real Hilbert space and let $\{x_i\}_{i=1}^N \subset H$. Then, for every $\alpha = (\alpha_1, \ldots, \alpha_N)^T \in \mathbb{R}^N$

$$\left\| \sum_{i=1}^N \alpha_i x_i \right\|^2 = \sum_{i,j=1}^N \alpha_i \alpha_j \langle x_i, x_j \rangle = \alpha^T G \alpha,$$

where $G = (g_{i,j})_{i,j=1}^N$ with $g_{i,j} = \langle x_i, x_j \rangle$ is the Gram matrix of $\{x_i\}_{i=1}^N$. Therefore, (1) is equivalent to $A\alpha^T \alpha \leq \alpha^T G \alpha \leq B\alpha^T \alpha$ for every $\alpha \in \mathbb{R}^N$ or simply to $\sigma(G) \subset [A, B]$. To show that this is indeed true for a given Gram matrix G, we will use the Gershgorin circle theorem [14, Theorem 6.1.1], which states that

$$\sigma(G) \subset \bigcup_{i=1}^N \left[g_{i,i} - \sum_{j \neq i} |g_{i,j}|, \, g_{i,i} + \sum_{j \neq i} |g_{i,j}| \right].$$

Proof of Theorem 2.2: Step 2.

In this step, we show that $\{1\} \cup \{\sqrt{3}\,\mathcal{C}_k, \sqrt{3}\,\mathcal{S}_k : k \leq N\}$ forms a Riesz sequence for every $N \in \mathbb{N}$ with the Riesz constants independent on N. By Lemma 2.1, its Gram matrix $G_N \in \mathbb{R}^{(2N+1)\times(2N+1)}$ is a block matrix with three blocks. The first one is just a 1×1 block corresponding to the constant function, the second and the third block are $N \times N$ matrices of the inner products $\left(3\langle\mathcal{C}_i, \mathcal{C}_j\rangle\right)_{i,j=1}^{N}$ and $\left(3\langle\mathcal{S}_i, \mathcal{S}_j\rangle\right)_{i,j=1}^{N}$, respectively, i.e.,

$$
G_N = \begin{pmatrix}
1 & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \begin{matrix} \overbrace{3\langle\mathcal{C}_1, \mathcal{C}_1\rangle}^{=1} & \dots & 3\langle\mathcal{C}_1, \mathcal{C}_N\rangle \\ \vdots & \ddots & \vdots \\ 3\langle\mathcal{C}_N, \mathcal{C}_1\rangle & \dots & \underbrace{3\langle\mathcal{C}_N, \mathcal{C}_N\rangle}_{=1} \end{matrix} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \begin{matrix} \overbrace{3\langle\mathcal{S}_1, \mathcal{S}_1\rangle}^{=1} & \dots & 3\langle\mathcal{S}_1, \mathcal{S}_N\rangle \\ \vdots & \ddots & \vdots \\ 3\langle\mathcal{S}_N, \mathcal{S}_1\rangle & \dots & \underbrace{3\langle\mathcal{S}_N, \mathcal{S}_N\rangle}_{=1} \end{matrix}
\end{pmatrix}.
$$

We apply the Gershgorin theorem to G_N. Therefore, we need to estimate the row sums of G_N. For the first row we see that $g_{1,1} = 1$ and $g_{1,j} = 0$ for all $j \neq 1$.

Next we assume that $i \leq N$ is an odd number, i.e. that

$$
i = q_1^{\alpha_1} \dots q_n^{\alpha_n}
$$

for some primes $q_1, \dots, q_n \geq 3$ and integers $\alpha_1, \dots, \alpha_n \geq 1$. For the $(i+1)$-th row $(i = 1, \dots, N$, corresponding to $\mathcal{C}_i$) we conclude

$$
\langle\sqrt{3}\,\mathcal{C}_i, 1\rangle + \sum_{j=1}^{N} 3 \cdot \langle\mathcal{C}_i, \mathcal{C}_j\rangle + \sum_{j=1}^{N} 3 \cdot \langle\mathcal{C}_i, \mathcal{S}_j\rangle = \sum_{1 \leq j \leq N, j \text{ odd}} 3 \cdot \langle\mathcal{C}_i, \mathcal{C}_j\rangle \leq \sum_{j \in \mathbb{N}, j \text{ odd}} 3 \cdot \langle\mathcal{C}_i, \mathcal{C}_j\rangle. \qquad (9)
$$

Every odd j can be written as $j = q_1^{\beta_1} \dots q_n^{\beta_n} \cdot J$, where $\beta_1, \dots, \beta_n \geq 0$ and J is an odd integer, not divisible by any of $q_1, \dots, q_n$, i.e., with $\gcd(J, i) = 1$. Observe that with this notation

$$
\gcd(i, j) = \prod_{u=1}^{n} q_u^{\min(\alpha_u, \beta_u)}.
$$

Therefore, we can use Lemma 2.1 and rewrite (9) as

$$
\sum_{j \in \mathbb{N}, \, j \text{ odd}} 3 \cdot \langle\mathcal{C}_i, \mathcal{C}_j\rangle = \sum_{\beta_1, \dots, \beta_n=0}^{\infty} \sum_{\substack{J \in \mathbb{N}, J \text{ odd} \\ \gcd(J, i)=1}} \frac{\prod_{u=1}^{n} q_u^{4\min(\alpha_u, \beta_u)}}{\left(q_1^{\alpha_1} \cdots q_n^{\alpha_n}\right)^2 \cdot \left(q_1^{\beta_1} \dots q_n^{\beta_n} \cdot J\right)^2}
$$

$$
= \sum_{\substack{J \in \mathbb{N}, J \text{ odd} \\ \gcd(J, i)=1}} \frac{1}{J^2} \cdot \sum_{\beta_1=0}^{\infty} \frac{1}{\left[q_1^{\alpha_1+\beta_1-2\min(\alpha_1, \beta_1)}\right]^2} \cdots \sum_{\beta_n=0}^{\infty} \frac{1}{\left[q_n^{\alpha_n+\beta_n-2\min(\alpha_n, \beta_n)}\right]^2}.
$$

Next, we simplify the individual terms.

$$
\sum_{\substack{J \in \mathbb{N}, J \text{ odd} \\ \gcd(J, i)=1}} \frac{1}{J^2} = \prod_{\substack{p \geq 3: p \text{ prime} \\ p \notin \{q_1, \dots, q_n\}}} \left(1 + \frac{1}{p^2} + \frac{1}{p^4} + \dots\right) = \prod_{\substack{p \geq 3: p \text{ prime} \\ p \notin \{q_1, \dots, q_n\}}} \frac{1}{1 - 1/p^2}
$$

and

$$\sum_{\beta=0}^{\infty} \frac{1}{\left[q^{\alpha+\beta-2\min(\alpha,\beta)}\right]^2} = \sum_{\beta=0}^{\alpha} \frac{1}{q^{2(\alpha-\beta)}} + \sum_{\beta=\alpha+1}^{\infty} \frac{1}{q^{2(\beta-\alpha)}}$$

$$\leq \sum_{r=0}^{\infty} \frac{1}{q^{2r}} + \sum_{r=1}^{\infty} \frac{1}{q^{2r}} = \left(1+1/q^2\right) \cdot \frac{1}{1-1/q^2}.$$

Therefore,

$$(9) \leq \prod_{p\geq 3:p \text{ prime}} \frac{1}{1-1/p^2} \cdot \prod_{u=1}^{n} \left(1+\frac{1}{q_u^2}\right) \leq \prod_{p\geq 3:p \text{ prime}} \frac{1+1/p^2}{1-1/p^2} = \frac{3}{2}, \qquad (10)$$

where in the last step we used the following Euler product [23, Page 5] attributed already to Ramanujan

$$\prod_{p \text{ prime}} \frac{1+1/p^2}{1-1/p^2} = \frac{5}{2}.$$

Therefore, we get

$$\sum_{j\neq i} |(G_N)_{i,j}| = \sum_{j\neq i} 3\langle \mathcal{C}_i, \mathcal{C}_j\rangle = \sum_{j=1}^{N} 3\langle \mathcal{C}_i, \mathcal{C}_j\rangle - 3\langle \mathcal{C}_i, \mathcal{C}_i\rangle \leq \frac{3}{2} - 1 = \frac{1}{2}. \qquad (11)$$

If i is even it follows from Lemma 2.1, assertions 2. and 3., that the estimates above remain the same since $3 \cdot \langle \mathcal{C}_i, \mathcal{C}_j\rangle = \frac{\gcd(i,j)^4}{i^2 \cdot j^2} \neq 0$ only for j's with the same power of 2 in their prime factorization as i, which then cancels out.

If we replace in (9) $\mathcal{C}_i$ by $\mathcal{S}_i$ (corresponding to the rows $(N+1)+i$, $i=1,\ldots,N$ of the Gram matrix), we obtain instead the estimate

$$\langle \sqrt{3}\,\mathcal{S}_i, 1\rangle + \sum_{j=1}^{N} 3 \cdot \langle \mathcal{S}_i, \mathcal{C}_j\rangle + \sum_{j=1}^{N} 3 \cdot |\langle \mathcal{S}_i, \mathcal{S}_j\rangle| \leq \sum_{1\leq j\leq N, j \text{ odd}} 3 \cdot |\langle \mathcal{S}_i, \mathcal{S}_j\rangle| \leq \sum_{j\in\mathbb{N}, j \text{ odd}} 3 \cdot |\langle \mathcal{S}_i, \mathcal{S}_j\rangle|. \qquad (12)$$

The term on the right hand side of (12) can be bounded by (10) as before and we again obtain

$$\sum_{j\neq i} |(G_N)_{i,j}| = \sum_{j\neq i} 3|\langle \mathcal{S}_i, \mathcal{S}_j\rangle| = \sum_{j=1}^{N} 3|\langle \mathcal{S}_i, \mathcal{S}_j\rangle| - 3\langle \mathcal{S}_i, \mathcal{S}_i\rangle \leq \frac{3}{2} - 1 = \frac{1}{2}. \qquad (13)$$

By Gershgorin's theorem, we deduce $\sigma(G_N) \subset \{1\} \cup [\frac{1}{2}, \frac{3}{2}] = [\frac{1}{2}, \frac{3}{2}]$.

Proof of Theorem 2.2: Step 3.
The third step of the proof of Theorem 2.2, i.e., the passage to the limit $N \to \infty$, is quite standard and straightforward (cf. [15] and [4, 5] for the so-called "projection method") and is contained in the following lemma.

Lemma 2.4. *Let H be a real Hilbert space and let $(x_n)_{n=1}^{\infty} \subset H$ be an infinite sequence. If $\{x_1,\ldots,x_N\}$ is a Riesz sequence for every $N \in \mathbb{N}$ with Riesz constants A and B independent on N, then $\{x_1, x_2, \ldots\}$ is also a Riesz sequence with Riesz constants A and B.*

Proof. Let $\alpha = (\alpha_1, \alpha_2, \ldots)$ be a square-summable sequence. Since

$$\left\| \sum_{n=n_0}^{n_1} \alpha_n x_n \right\|^2 \leq B \sum_{n=n_0}^{n_1} \alpha_n^2,$$

the partial sums of $\sum_{n=1}^{\infty} \alpha_n x_n$ form a Cauchy sequence and therefore the series is convergent. Furthermore, by the triangle inequality

$$\left\| \sum_{n=1}^{N} \alpha_n x_n \right\| \to \left\| \sum_{n=1}^{\infty} \alpha_n x_n \right\|.$$

Hence, we can take the limit $N \to \infty$ in

$$A \sum_{n=1}^{N} \alpha_n^2 \le \left\| \sum_{n=1}^{N} \alpha_n x_n \right\|^2 \le B \sum_{n=1}^{N} \alpha_n^2$$

and the result follows. $\qquad\square$

Proof of Theorem 2.2: Step 4.
As the last step, we show that $\mathcal{R}_1$ is not only a Riesz sequence but also a Riesz basis, i.e., that its closed linear span is the whole space $L_2([0,1])$. We rely on the fact that the trigonometric system

$$\mathcal{T}_1 := \{1\} \cup \{\sqrt{2}\cos(2\pi k x) : k \in \mathbb{N}\} \cup \{\sqrt{2}\sin(2k\pi x) : k \in \mathbb{N}\} \tag{14}$$

forms an orthonormal basis of $L_2([0,1])$. We show that every function from (14) lies in the closed linear span of $\mathcal{R}_1$ and, therefore, the closed linear span of $\mathcal{T}_1$ is contained in the closed linear span of $\mathcal{R}_1$.

We start with the following lemma, which gives an explicit decomposition of $\cos(2\pi x)$ and $\sin(2\pi x)$ in $\mathcal{R}_1$. Its statement requires the notion of the Möbius function, which is defined for every positive integer $n \in \mathbb{N}$ as

$$\mu(n) = \begin{cases} +1, & \text{if } n \text{ is a square-free integer with an even number of prime factors,} \\ -1, & \text{if } n \text{ is a square-free integer with an odd number of prime factors,} \\ 0, & \text{if } n \text{ is not a square-free integer, i.e., if it is divisible by some squared prime.} \end{cases}$$

Lemma 2.5. *For every $l \in \mathbb{N}$, let $\overline{\mathcal{C}}_l(x) = \sqrt{3}\,\mathcal{C}_l(x)/\mu$ and $\overline{\mathcal{S}}_l(x) = \sqrt{3}\,\mathcal{S}_l(x)/\mu$, where μ is the constant from (6). Then*

$$c_1(x) := \sqrt{2}\cos(2\pi x) = \sum_{l=0}^{\infty} \frac{\mu(2l+1)}{(2l+1)^2}\,\overline{\mathcal{C}}_{2l+1}(x) \tag{15}$$

and

$$s_1(x) := \sqrt{2}\sin(2\pi x) = \sum_{l=0}^{\infty} (-1)^l \frac{\mu(2l+1)}{(2l+1)^2}\,\overline{\mathcal{S}}_{2l+1}(x) \tag{16}$$

with the convergence being in $L_2([0,1])$.

Proof. We first reformulate (5) as

$$\overline{\mathcal{C}}_1(x) = \frac{\sqrt{3}\,\mathcal{C}(x)}{\mu} = \sum_{m=1}^{\infty} \frac{\alpha_m}{m^2} \cdot c_m(x), \tag{17}$$

where $\alpha_{2m+1} = 1$ and $\alpha_{2m} = 0$. Note, that (17) converges in $L_2([0,1])$.

We show that there is a unique bounded sequence $(\beta_l)_{l=1}^{\infty}$, such that

$$c_1(x) = \sum_{l=1}^{\infty} \frac{\beta_l}{l^2} \cdot \overline{\mathcal{C}}_l(x) \tag{18}$$

8

with the convergence in $L_2([0,1])$ and that

$$\beta_l = \begin{cases} 0 & \text{for } l \in \mathbb{N} \text{ even,} \\ \mu(l) & \text{for } l \in \mathbb{N} \text{ odd.} \end{cases} \tag{19}$$

Using (17), we observe that (18) holds for bounded sequence $(\beta_l)_{l=1}^{\infty}$ if, and only if,

$$c_1(x) = \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} \frac{\beta_l \cdot \alpha_m}{l^2 \cdot m^2} c_{lm}(x). \tag{20}$$

We compare the coefficients of $c_k(x)$ on both sides of (20) and observe that (20) is equivalent to the system of equations

$$
\begin{aligned}
1 &= \beta_1 \alpha_1, \\
0 &= \sum_{(l,m):l\cdot m=n} \beta_l \alpha_m, \quad n = 2,3,\dots.
\end{aligned} \tag{21}
$$

This system could be solved by using the Möbius inversion formula [23, p. 3], but one can also proceed directly. From $1 = \beta_1 \alpha_1$ we obtain $\beta_1 = 1$ and from $\beta_1 \alpha_2 + \beta_2 \alpha_1 = 0$ we get $\beta_2 = 0$. We show by induction that $\beta_{2n} = 0$ also for all $n \geq 1$. Let this be true for all integers smaller than n. Then

$$0 = \sum_{(l,m):l\cdot m=2n} \beta_l \alpha_m = \sum_{(l,m):l\cdot m=n} \beta_{2l} \alpha_m + \sum_{(l,m):l\cdot m=n} \beta_l \alpha_{2m} = \beta_{2n} \alpha_1$$

gives $\beta_{2n} = 0$ as well. Similarly, from $0 = \beta_p \alpha_1 + \beta_1 \alpha_p$ we get $\beta_p = -1$ for every prime $p \neq 2$.

Let now $p = p_1 p_2$ with odd primes p_1, p_2. Then $\beta_p = 1$ follows from

$$0 = \beta_p\,\alpha_1 + \beta_{p_1}\alpha_{p_2} + \beta_{p_2}\alpha_{p_1} + \beta_1 \alpha_p = \beta_p \cdot 1 + (-1 \cdot 1) + (-1 \cdot 1) + 1.$$

The formula for a general $p = p_1 \cdot \dots \cdot p_k$, with distinct odd primes p_i follows by induction. Observe that p is divisible by all $p_1^{e_1} \cdot \dots \cdot p_k^{e_k}$ with $e = (e_1, \dots, e_k) \in \{0,1\}^k$. Hence

$$0 = \sum_{(l,m):l\cdot m=p} \beta_l \alpha_m = \sum_{l|p} \beta_l = \beta_p + \sum_{j=0}^{k-1} \binom{k}{j} (-1)^j = \beta_p + \sum_{j=0}^{k} \binom{k}{j} (-1)^j - (-1)^k = \beta_p - (-1)^k.$$

We conclude, that $\beta_n = \mu(n)$ for every odd square-free integer n.

Finally, we show that $\beta_n = 0$ for every positive odd integer n, which is not square-free. Therefore, we assume that $n = p_1^u \cdot q$, where p_1 is an odd prime, q is an odd integer not divisible by p_1, $u \geq 2$ and that the statement is true for all integers smaller than n. We obtain

$$0 = \sum_{(l,m):l\cdot m=n} \beta_l \alpha_m = \sum_{l|p_1^u q} \beta_l = \sum_{l|q} (\beta_l + \beta_{p_1 l} + \dots + \beta_{p_1^u l}).$$

If $l < q$ is square-free, then the first two terms in this sum have values $+1$ and -1, respectively, and the others vanish by the induction assumption. If $l < q$ is not square-free, then all the terms vanish again by assumption. Finally, if $l = q$ the same argument applies leaving us with $\beta_n = 0$. We conclude that the sequence $(\beta_l)_{l=1}^{\infty}$ given by (19) indeed satisfies the system (21) which in turn gives (18).

Finally, (16) follows from (15) using the simple relation $\mathcal{S}(x) = \mathcal{C}(x - 1/4)$. The factor $(-1)^l$ results from the relation

$$\mathcal{C}_{2l+1}\left(x - \frac{1}{4}\right) = (-1)^l \mathcal{C}\left((2l+1)x - \frac{1}{4}\right) = (-1)^l \mathcal{S}_{2l+1}(x).$$

$\square$

3 Multivariate case

The main aim of this section is to generalize Theorem 2.2 to higher dimensions $d \geq 1$ and to provide a Riesz basis of $L_2([0,1]^d)$, which is easily expressed by artificial neural networks with ReLU activation function. The most natural approach would be to consider the tensor products of the functions from $\mathcal{R}_1$, i.e., a system of functions of the form $(x,y) \rightarrow \mathcal{C}_k(x) \cdot \mathcal{C}_l(y)$ etc. Indeed, it is quite easy to show that tensor products of elements of a Riesz sequence form again a Riesz sequence [2]. This approach is quite classical in analysis and there exist many multivariate bases and systems with a tensor product structure. Unfortunately, the Riesz constants of the tensor product system are in general given as products of the Riesz constants of the univariate Riesz sequences, cf. [2, Theorem 4.1]. Applying the tensor product construction to $\mathcal{R}_1$ would therefore lead to an exponential dependence of the ratio of the Riesz constants on the dimension.

Furthermore, the tensor product approach does not fit really well to artificial neural networks. The reason is that it is surprisingly difficult to construct a neural network, which for two real inputs x and y outputs the product xy (or at least its approximation). In general, one first approximates the square function $t \rightarrow t^2$ and then applies the formula $xy = [(x+y)^2 - (x-y)^2]/4$. We refer to [10, 20, 25] for details. Therefore, we are looking for another multivariate Riesz basis of piecewise affine functions, which can be constructed without the use of (tensor) products, but where inner products with fixed vectors in $\mathbb{R}^d$ are allowed.

Before we state our results, we need some additional notation. If $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{Z}^d$, we say that $\alpha \rhd 0$ if the first non-zero entry of α is positive. The multivariate analogue of $\mathcal{R}_1$ is then defined simply as

$$\mathcal{R}_d := \{1\} \cup \{\mathcal{C}(\alpha \cdot x) : \alpha \rhd 0\} \cup \{\mathcal{S}(\alpha \cdot x) : \alpha \rhd 0\},$$

where we interpret the functions $\mathcal{C}$ and $\mathcal{S}$ as $1-$periodic functions on $\mathbb{R}$. Note that we need to restrict ourselves to indices $\alpha \rhd 0$ here, since $\mathcal{C}(\alpha \cdot x) = \mathcal{C}(-\alpha \cdot x)$ and $\mathcal{S}(\alpha \cdot x) = -\mathcal{S}(-\alpha \cdot x)$, respectively.

Furthermore, we say that two non-zero $\alpha, \beta \in \mathbb{R}^d$ are co-linear if there is $t \neq 0$ such that $\alpha = t\beta$. Obviously, if $\alpha, \beta \in \mathbb{Z}^d$ with $\alpha, \beta \rhd 0$ are co-linear, then $t > 0$. In the rest of this section, the inner product $\langle f, g \rangle = \int_{[0,1]^d} f(x)g(x)dx$ denotes the inner product in $L_2([0,1]^d)$, the space of real square integrable functions on $[0,1]^d$.

The multivariate analogue of Lemma 2.1, which characterizes the inner products of the elements of $\mathcal{R}_d$ then looks as follows.

Lemma 3.1. Let $\alpha, \beta \in \mathbb{Z}^d$ with $\alpha, \beta \rhd 0$. Then

1. $\langle \mathcal{C}(\alpha \cdot x), \mathcal{S}(\beta \cdot x) \rangle = 0$.

2. $\langle \mathcal{C}(\alpha \cdot x), \mathcal{C}(\beta \cdot x) \rangle = \langle \mathcal{S}(\alpha \cdot x), \mathcal{S}(\beta \cdot x) \rangle = 0$ if α and β are not co-linear, or if $\alpha = t\beta$, but t can not be written as a ratio of two odd positive integers.

3. If $\alpha = \dfrac{2p+1}{2q+1} \cdot \beta$ with coprime integers $2p+1$ and $2q+1$ (i.e., $\gcd(2p+1, 2q+1) = 1$), then

$$3\langle \mathcal{C}(\alpha \cdot x), \mathcal{C}(\beta \cdot x) \rangle = \frac{1}{(2p+1)^2(2q+1)^2} \tag{22}$$

   and

$$3\langle \mathcal{S}(\alpha \cdot x), \mathcal{S}(\beta \cdot x) \rangle = \frac{(-1)^{p+q}}{(2p+1)^2(2q+1)^2}. \tag{23}$$

Remark 3.2. Lemma 3.1 includes Lemma 2.1 as a special case. In particular, if $d = 1$ then α and β are always co-linear. Moreover, if the prime factorizations of α and β contain different powers of 2, then we have $\alpha \neq \frac{2p+1}{2q+1}\beta$ for all integers p, q.

Proof of Lemma 3.1. First, we recall the elementary formulas

$$\cos(u_1 + \cdots + u_d) = \mathrm{Re}(e^{i(u_1 + \cdots + u_d)}) = \mathrm{Re}\left[\prod_{j=1}^{d}(\cos u_j + i \sin u_j)\right]$$

$$= \sum_{J \subset \{1,\ldots,d\}, \#J \text{ even}} (-1)^{\#J/2} \prod_{j \notin J} \cos u_j \cdot \prod_{j \in J} \sin u_j \tag{24}$$

and

$$\sin(u_1 + \cdots + u_d) = \mathrm{Im}(e^{i(u_1 + \cdots + u_d)}) = \mathrm{Im}\left[\prod_{j=1}^{d}(\cos u_j + i \sin u_j)\right]$$

$$= \sum_{J \subset \{1,\ldots,d\}, \#J \text{ odd}} (-1)^{(\#J-1)/2} \prod_{j \notin J} \cos u_j \cdot \prod_{j \in J} \sin u_j. \tag{25}$$

Next, we proceed to the proof of (22). Let us fix $\alpha, \beta \rhd 0$. We apply (5) followed by (24) and obtain

$$3\langle \mathcal{C}(\alpha \cdot x), \mathcal{C}(\beta \cdot x)\rangle = \sum_{m,n \geq 0} \frac{\mu^2}{(2m+1)^2(2n+1)^2} \int_{[0,1]^d} c_{2m+1}(\alpha \cdot x) c_{2n+1}(\beta \cdot x) dx$$

$$= \sum_{m,n \geq 0} \sum_{\substack{J,K \subset \{1,\ldots,d\} \\ \#J, \#K \text{ even}}} \frac{2\mu^2}{(2m+1)^2(2n+1)^2} \cdot (-1)^{(\#J + \#K)/2}.$$

$$\cdot \frac{1}{2^d} \cdot \int_{[0,1]^d} \prod_{j \notin J} c_{2m+1}(\alpha_j x_j) \prod_{j \in J} s_{2m+1}(\alpha_j x_j) \prod_{k \notin K} c_{2n+1}(\beta_k x_k) \prod_{k \in K} s_{2n+1}(\beta_k x_k) dx$$

$$= \sum_{m,n \geq 0} \sum_{\substack{J \subset \{1,\ldots,d\} \\ \#J \text{ even}}} \frac{2\mu^2}{(2m+1)^2(2n+1)^2}$$

$$\cdot \frac{1}{2^d} \cdot \prod_{j \notin J} \int_{[0,1]} c_{2m+1}(\alpha_j x_j) c_{2n+1}(\beta_j x_j) dx_j \cdot \prod_{j \in J} \int_{[0,1]} s_{2m+1}(\alpha_j x_j) s_{2n+1}(\beta_j x_j) dx_j. \tag{26}$$

Next, we discuss, when the last product vanishes for given J. First, this happens if $\alpha_j = 0$ or $\beta_j = 0$ for any $j \in J$. If $j \in J$ and both α_j and β_j are non-zero, then the last product vanishes also if $(2m+1)|\alpha_j| \neq (2n+1)|\beta_j|$. And finally, the product is zero also if $j \notin J$ and $(2m+1)|\alpha_j| \neq (2n+1)|\beta_j|$.

Equivalently, (26) is not equal to zero if $(2m+1)|\alpha_j| = (2n+1)|\beta_j|$ for every $j \in \{1,\ldots,d\}$ and α_j and β_j are different from zero if $j \in J$. If we denote $|\alpha| = (|\alpha_1|,\ldots,|\alpha_n|)$ (and similarly for $|\beta|$), we will therefore restrict ourselves for the rest of the proof to $m, n \geq 0$ with

$$(2m+1)|\alpha| = (2n+1)|\beta|. \tag{27}$$

If there is no pair of integers $(m,n) \in \mathbb{N}_0^2$, such that (27) holds, then $\langle \mathcal{C}(\alpha \cdot x), \mathcal{C}(\beta \cdot x)\rangle = 0$. Furthermore, we may consider only sets $J \subset \{1,\ldots,d\}$ with an even number of elements, which are subsets of $\mathrm{supp}(\alpha) = \mathrm{supp}(\beta)$.

If (27) holds, than the univariate integrals in (26) are equal to one for $j \notin J$ and $j \in \mathrm{supp}(\alpha)$. They are equal to two, if $j \notin J$ and $j \notin \mathrm{supp}(\alpha)$. And if $j \in J \subset \mathrm{supp}(\alpha)$, then the integral is $+1$ if $(2m+1)\alpha_j = (2n+1)\beta_j$ and it is equal to -1 if $(2m+1)\alpha_j = -(2n+1)\beta_j$.

We denote $D = \{j : \mathrm{sign}(\alpha_j) \cdot \mathrm{sign}(\beta_j) = -1\} \subset \mathrm{supp}(\alpha) = \mathrm{supp}(\beta)$, $\nu = \#D$ and $n = \#\mathrm{supp}(\alpha) \geq \nu$. Using this notation, we obtain

$$3\langle \mathcal{C}(\alpha \cdot x), \mathcal{C}(\beta \cdot x)\rangle = \sum_{\substack{m,n \geq 0 \\ (2m+1)|\alpha|=(2n+1)|\beta|}} \frac{2\mu^2}{(2m+1)^2(2n+1)^2 \cdot 2^n} \sum_{\substack{J \subset \mathrm{supp}(\alpha) \\ \#J \text{ even}}} (-1)^{\#(J \cap D)}.$$

If $\nu \geq 1$, we calculate

$$\sum_{\substack{J \subset \mathrm{supp}(\alpha) \\ \#J \text{ even}}} (-1)^{\#(J \cap D)} = \sum_{\substack{0 \leq a \leq n-\nu \\ 0 \leq b \leq \nu \\ a+b \text{ even}}} (-1)^b \binom{n-\nu}{a} \binom{\nu}{b}$$

$$= \sum_{b=0}^{\nu} (-1)^b \binom{\nu}{b} \sum_{\substack{0 \leq a \leq n-\nu \\ a+b \text{ even}}} \binom{n-\nu}{a} = 0,$$

where the last step follows since $\sum_{b=0}^{\nu}(-1)^b \binom{\nu}{b} = (1-1)^{\nu} = 0$. Hence, $\langle \mathcal{C}(\alpha \cdot x), \mathcal{C}(\beta \cdot x) \rangle = 0$ if there exists $1 \leq j \leq d$ with $\mathrm{sign}(\alpha_j) \cdot \mathrm{sign}(\beta_j) = -1$ and we arrive at

$$3\langle \mathcal{C}(\alpha \cdot x), \mathcal{C}(\beta \cdot x) \rangle = \sum_{\substack{m,n \geq 0 \\ (2m+1)\alpha = (2n+1)\beta}} \frac{\mu^2}{(2m+1)^2(2n+1)^2}.$$

The last sum is empty if α and β are not co-linear or if we can not write $\alpha = t\beta$, where $t > 0$ is a ratio of two odd integers. Therefore, we assume that $\alpha = \dfrac{2p+1}{2q+1}\beta$ with $p, q \in \mathbb{N}_0$ and that $2p+1$ and $2q+1$ are coprime integers. All pairs $(m,n) \in \mathbb{N}_0^2$ with $(2m+1)\alpha = (2n+1)\beta$ are then of the form

$$2m + 1 = (2q+1)(2l+1) \quad \text{and} \quad 2n + 1 = (2p+1)(2l+1), \quad l \in \mathbb{N}_0. \tag{28}$$

This finally leads to

$$3\langle \mathcal{C}(\alpha \cdot x), \mathcal{C}(\beta \cdot x) \rangle = \sum_{l=0}^{\infty} \frac{\mu^2}{(2q+1)^2(2p+1)^2(2l+1)^4},$$

which combined with (6) gives (22). As a byproduct, we also showed that $\langle \mathcal{C}(\alpha \cdot x), \mathcal{C}(\beta \cdot x) \rangle = 0$ if α and β are not co-linear with a real factor t, which can be written as a ration of two odd positive integers.

Applying the same idea to the inner product of $\mathcal{C}(\alpha \cdot x)$ and $\mathcal{S}(\beta \cdot x)$, we get a double sum over $J, K \subset \{1, \ldots, d\}$ with $\#J$ even and $\#K$ odd. Therefore, it is not possible to match the univariate integrands and their product always vanishes. Finally, (23) follows in the same way, the only essential difference being the $(-1)^{m+n}$ factor coming from (5). And an easy observation shows that under (28), the parity of $m + n$ is the same as the one of $p + q$. $\qquad \square$

We complement Lemma 3.1 by the simple observation that the constant function is orthogonal to all other elements of $\mathcal{R}_d$. The multivariate analogue of Theorem 2.2 then reads as follows.

Theorem 3.3. *Let $d \geq 1$. Then the system $\mathcal{R}_d$ forms a Riesz basis of $L_2([0,1]^d)$ with the Riesz constants independent of d. To be more specific, the Riesz constants of the normalized system*

$$\{1\} \cup \{\sqrt{3}\,\mathcal{C}(\alpha \cdot x) : \alpha \rhd 0\} \cup \{\sqrt{3}\,\mathcal{S}(\alpha \cdot x) : \alpha \rhd 0\} \tag{29}$$

can be chosen as $A = 1/2$ and $B = 3/2$ independently of d.

Proof. Step 1. Using Lemma 3.1 together with the Gershgorin circle theorem, it is surprisingly simple to prove Theorem 3.3 with slightly worse constants A and B, cf. Remark 3.4. To improve the Riesz constants to $A = 1/2$ and $B = 3/2$, we proceed more carefully. Let us denote by $\mathbb{P}$ the set of primes and by $\mathbb{P}' = \mathbb{P} \setminus \{2\}$ the set of odd primes. Then every odd $p \in \mathbb{N}$ can be written as $p = p_1^{k_1} \cdot \ldots \cdot p_n^{k_n}$ with $p_1, \ldots, p_n \in \mathbb{P}'$ and $k_1, \ldots, k_n \in \mathbb{N}$. If $p = 1$, then we choose $n = 0$ and interpret the empty product as one.

We use the following observation. To a fixed $\alpha \rhd 0$ and a pair of odd coprimed integers $2p+1$ and $2q+1$, there exists at most one $\beta \rhd 0$ such that $\alpha = (2p+1)/(2q+1) \cdot \beta$. Then we obtain

$$
\begin{aligned}
3 \sum_{\beta \rhd 0} \langle \mathcal{C}(\alpha \cdot x), \mathcal{C}(\beta \cdot x) \rangle &\leq \sum_{2p+1, 2q+1 \text{ coprimes}} \frac{1}{(2p+1)^2 (2q+1)^2} = \sum_{p,q \geq 1 \text{ odd coprimes}} \frac{1}{p^2 q^2} \\
&= \sum_{\{p_1,\ldots,p_n\} \subset \mathbb{P}'} \sum_{k_1,\ldots,k_n=1}^{\infty} \frac{1}{p_1^{2k_1} \cdots p_n^{2k_n}} \sum_{\substack{q \geq 1, q \text{ odd} \\ \gcd(q, p_1 \ldots p_n)=1}} \frac{1}{q^2} \\
&= \sum_{\{p_1,\ldots,p_n\} \subset \mathbb{P}'} \prod_{u=1}^{n} \left(\sum_{k_u=1}^{\infty} \frac{1}{p_u^{2k_u}} \right) \prod_{p \in \mathbb{P}' \setminus \{p_1,\ldots,p_n\}} \left(\sum_{k=0}^{\infty} \frac{1}{p^{2k}} \right) \\
&= \sum_{\{p_1,\ldots,p_n\} \subset \mathbb{P}'} \prod_{u=1}^{n} \frac{1}{p_u^2 (1 - 1/p_u^2)} \prod_{p \in \mathbb{P}' \setminus \{p_1,\ldots,p_n\}} \frac{1}{1 - 1/p^2} \\
&= \sum_{\{p_1,\ldots,p_n\} \subset \mathbb{P}'} \frac{1}{p_1^2 \cdots p_n^2} \prod_{p \in \mathbb{P}'} \frac{1}{1 - 1/p^2} \\
&= \prod_{p \in \mathbb{P}'} \frac{1}{1 - 1/p^2} \prod_{p \in \mathbb{P}'} (1 + 1/p^2) = \frac{3}{2},
\end{aligned}
$$

where the last step follows from (10). Using the Gershgorin circle theorem in the same way as in the proof of Theorem 2.2 then gives the bounds $1/2 \leq A \leq B \leq 3/2$, independent of d.

Step 2. We show that $\mathcal{R}_d$ is also a Riesz basis. The system

$$
\{1\} \cup \{c_k(x) : k \in \mathbb{N}\} \cup \{s_k(x) : k \in \mathbb{N}\} \tag{30}
$$

with $c_k(x) = \sqrt{2} \cos(2\pi k x)$ and $s_k(x) = \sqrt{2} \sin(2\pi k x)$ is an orthonormal basis of $L_2([0,1])$. Therefore, all possible tensor products of the functions from (30) form an orthonormal basis of $L_2([0,1]^d)$. For this system we use the following notation

$$
\left\{ \prod_{l \in L} c_{k_l}(x_l) \prod_{m \in M} s_{k_m}(x_m) : L, M \subset \{1, \ldots, d\} \text{ disjoint}, k_l, k_m \in \mathbb{N} \right\}. \tag{31}
$$

Now we show that every function from (31) can be found in the closed linear span of $\mathcal{R}_d$. This will imply the completeness of $\mathcal{R}_d$. First, we again recall two simple formulas

$$
\begin{aligned}
\prod_{u=1}^{n} \cos(\varphi_u) &= \prod_{u=1}^{n} \frac{e^{i\varphi_u} + e^{-i\varphi_u}}{2} = \frac{1}{2^n} \sum_{e \in \{-1,+1\}^n} \exp(i[e_1 \varphi_1 + \cdots + e_n \varphi_n]) \\
&= \frac{1}{2^n} \sum_{e \in \{-1,+1\}^n} \cos(e_1 \varphi_1 + \cdots + e_n \varphi_n)
\end{aligned}
$$

and, similarly,

$$
\prod_{u=1}^{n} \sin(\varphi_u) = \frac{(-1)^{\lfloor \frac{n}{2} \rfloor}}{2^n} \cdot \begin{cases} \displaystyle\sum_{e \in \{-1,+1\}^n} \cos(e_1 \varphi_1 + \cdots + e_n \varphi_n) \cdot \prod_{j=1}^{n} e_j & \text{if } n \text{ is even}, \\ \displaystyle\sum_{e \in \{-1,+1\}^n} \sin(e_1 \varphi_1 + \cdots + e_n \varphi_n) \cdot \prod_{j=1}^{n} e_j & \text{if } n \text{ is odd}. \end{cases}
$$

If $\#M$ is even, we use the elementary property $\cos(\alpha)\cos(\beta) = (\cos(\alpha+\beta)+\cos(\alpha-\beta))/2$ and obtain

$$\prod_{l\in L}c_{k_l}(x_l)\prod_{m\in M}s_{k_m}(x_m) = 2^{(\#L+\#M)/2}\prod_{l\in L}\cos(2\pi k_l x_l)\prod_{m\in M}\sin(2\pi k_m x_m)$$

$$= \frac{(-1)^{\lfloor \#M/2\rfloor}}{2^{(\#L+\#M)/2}}\sum_{e\in\{-1,+1\}^{\#L}}\cos\left(\sum_{l\in L}e_l\cdot 2\pi k_l x_l\right)\cdot\sum_{e'\in\{-1,+1\}^{\#M}}\cos\left(\sum_{m\in M}e'_m\cdot 2\pi k_m x_m\right)\cdot\prod_{m\in M}e'_m$$

$$= \frac{(-1)^{\lfloor \#M/2\rfloor}}{2^{(\#L+\#M)/2}}\sum_{e\in\{-1,+1\}^{\#L+\#M}}\cos\left(\sum_{l\in L}e_l\cdot 2\pi k_l x_l\right)\cdot\cos\left(\sum_{m\in M}e_m\cdot 2\pi k_m x_m\right)\cdot\prod_{m\in M}e_m$$

$$= \frac{(-1)^{\lfloor \#M/2\rfloor}}{2\cdot 2^{(\#L+\#M)/2}}\sum_{e\in\{-1,+1\}^{\#L+\#M}}\cos\left(\sum_{l\in L}e_l\cdot 2\pi k_l x_l+\sum_{m\in M}e_m\cdot 2\pi k_m x_m\right)\cdot\prod_{m\in M}e_m$$

$$+ \frac{(-1)^{\lfloor \#M/2\rfloor}}{2\cdot 2^{(\#L+\#M)/2}}\sum_{e\in\{-1,+1\}^{\#L+\#M}}\cos\left(\sum_{l\in L}e_l\cdot 2\pi k_l x_l-\sum_{m\in M}e_m\cdot 2\pi k_m x_m\right)\cdot\prod_{m\in M}e_m$$

$$= \frac{(-1)^{\lfloor \#M/2\rfloor}}{2^{(\#L+\#M)/2}}\sum_{e\in\{-1,+1\}^{\#L+\#M}}\cos\left(\sum_{u\in L\cup M}e_u\cdot 2\pi k_u x_u\right)\cdot\prod_{m\in M}e_m$$

$$= \frac{(-1)^{\lfloor \#M/2\rfloor}}{2^{(\#L+\#M)/2}}\sum_{e\in\{-1,+1\}^{\#L+\#M}}\sum_{l=0}^{\infty}\frac{\mu(2l+1)}{(2l+1)^2}\frac{\sqrt{3}}{\sqrt{2}\mu}\mathcal{C}_{2l+1}\left(\sum_{u\in L\cup M}e_u k_u x_u\right)\cdot\prod_{m\in M}e_m, \tag{32}$$

where in the last step we used the Fourier decomposition from (15). If $\#M$ is odd, we use instead the formula $\cos(\alpha)\sin(\beta) = (\sin(\alpha+\beta)+\sin(\beta-\alpha))/2$, which yields

$$\prod_{l\in L}c_{k_l}(x_l)\prod_{m\in M}s_{k_m}(x_m)$$

$$= \frac{(-1)^{\lfloor \#M/2\rfloor}}{2^{(\#L+\#M)/2}}\sum_{e\in\{-1,+1\}^{\#L+\#M}}\sum_{l=0}^{\infty}(-1)^l\frac{\mu(2l+1)}{(2l+1)^2}\frac{\sqrt{3}}{\sqrt{2}\mu}\mathcal{S}_{2l+1}\left(\sum_{u\in L\cup M}e_u k_u x_u\right)\cdot\prod_{m\in M}e_m.$$

We interprete these formulas as a decomposition of a basis function from (31) into $\mathcal{R}_d$, which converges in $L_2([0,1]^d)$. Reasoning similarly as in the proof of Theorem 2.2, this finishes the argument. $\square$

Remark 3.4. *We observe that Lemma 3.1 implies for fixed $\alpha \gtrdot 0$*

$$3\sum_{\beta\neq\alpha}|\langle\mathcal{C}(\alpha\cdot x),\mathcal{C}(\beta\cdot x)\rangle| \leq \sum_{p,q\geq 0}\frac{1}{(2p+1)^2(2q+1)^2}-1 = \frac{\pi^4}{8^2}-1 = 0.522....$$

Using Gershgorin's theorem similarly as in the proof of Theorem 2.2, we could have obtained quite easily that (29) is a Riesz sequence with constants $A = 2 - \pi^4/64$ and $B = \pi^4/64$.

4 Neural networks

In this section we finally address the question in which classes of neural networks we can find the elements of the new Riesz basis $\mathcal{R}_d$. Therefore, we first fix some notation and recall what was shown in [7, Sect. 6] (in case $d = 1$).

A function $f : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ is called affine, if it can be written as $f(x) = Mx + b$, where $M \in \mathbb{R}^{n_2\times n_1}$ is a matrix and $b \in \mathbb{R}^{n_2}$. The following definition formalizes the notion of ReLU neural networks with width W and depth L, cf. Figure 4 and 5.

14

Definition 4.1. *Let d, W, L be positive integers. Then a feed-forward* ReLU *network $\mathcal{N}$ with width W and depth L is a collection of $L + 1$ affine mappings $A^{(0)}, \dots, A^{(L)}$, where $A^{(0)} : \mathbb{R}^d \to \mathbb{R}^W$, $A^{(j)} : \mathbb{R}^W \to \mathbb{R}^W$ for $j = 1, \dots, L-1$ and $A^{(L)} : \mathbb{R}^W \to \mathbb{R}$. Each such a network $\mathcal{N}$ generates a function of d variables*

$$A^{(L)} \circ \mathrm{ReLU} \circ A^{(L-1)} \circ \cdots \circ \mathrm{ReLU} \circ A^{(0)}.$$

Moreover, we denote by $\Upsilon^{W,L}$ the set of all functions, which are generated in this way by some feed-forward ReLU *network with width W and depth L.*
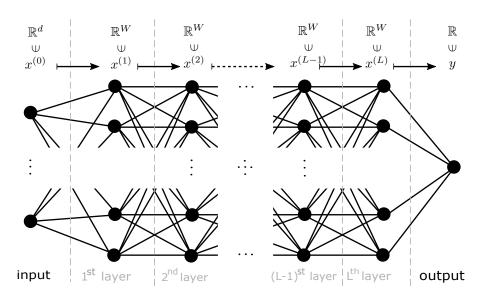

Figure 4: Feed-forward ReLU network with length L, width W
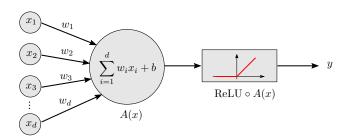

Figure 5: Close-up: ReLU activation function acting in neuron in 1st hidden layer

Every $S \in \Upsilon^{W,L}$ is a continuous piecewise affine function on $\mathbb{R}^d$. If the affine mappings associated to S are denoted by $A^{(l)}$ with $l = 0, \dots, L$, then the value $S(x^{(0)})$ is computed for each input $x := x^{(0)} \in \mathbb{R}^d$ after the calculation of a series of intermediate vectors $x^{(l)} := \mathrm{ReLU}(A^{(l-1)}x^{(l-1)}) \in \mathbb{R}^W$, $l = 1, \dots, L$, called vectors of activation at layer l. Finally the output $S(x)$ is produced as $S(x) := x^{(L+1)} = A^{(L)}x^{(L)}$.

We collect some properties of the sets $\Upsilon^{W,L}$, which are needed in the sequel.

Proposition 4.2. *Let $W \geq 2$.*

(i) *Let $\mathcal{Y}_1 \in \Upsilon^{W,L_1}, \dots, \mathcal{Y}_k \in \Upsilon^{W,L_k}$. Then the composition of the $\mathcal{Y}_i$ satisfies*

$$\mathcal{Y}_k \circ \cdots \circ \mathcal{Y}_1 \in \Upsilon^{W,L}, \qquad L = L_1 + \cdots + L_k.$$

(ii) *Let $L \geq 1$. Then $\Upsilon^{W,L} \subset \Upsilon^{W,L+1}$.*

Proof. The proof of (i) can be found in [7, Prop. 4.2] for $d = 1$. The proof for general $d \geq 1$ follows virtually without any change.

For the proof of (ii), we consider the identity function $\mathrm{id}(x) = x$ for $x \in \mathbb{R}$. We rewrite it as

$$\mathrm{id}(x) = \begin{bmatrix} 1 & -1 \end{bmatrix} \mathrm{ReLU} \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} x \right\} \tag{33}$$

to conclude that $\mathrm{id} \in \Upsilon^{2,1}$. An easy modification of (33) also shows that $\mathrm{id} \in \Upsilon^{W,1}$ for every $W \geq 2$. The result then follows by (i). $\qquad\square$

An important example and building block for our constructions to follow is the *hat function* $H : [0,1] \to \mathbb{R}$

$$H(x) := \begin{cases} 2x, & 0 \leq x \leq \frac{1}{2}, \\ 2(1-x), & \frac{1}{2} < x \leq 1, \end{cases}$$

which was already used in connection with feed-forward neural networks with ReLU activation function by [22], cf. also [7, 20, 25]. From the representation

$$H(x) = \begin{bmatrix} 2 & -4 \end{bmatrix} \mathrm{ReLU} \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} x + \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix} \right\}, \quad x \in [0,1]$$

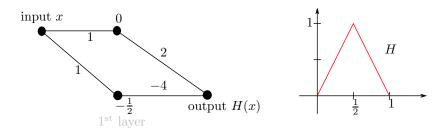we see, that H belongs to $\Upsilon^{2,1}$, see Figure 6.

Figure 6: Computational graph showing $H \in \Upsilon^{2,1}$ and usual graph associated with H

Furthermore, since $H \in \Upsilon^{2,1}$, we deduce from Proposition 4.2(i) that the k-fold composition $H^{\circ k} := H \circ H \circ \cdots \circ H$ belongs to $\Upsilon^{2,k}$, cf. Figure 7.
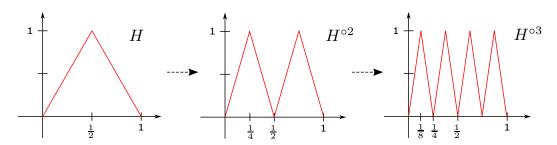

Figure 7: The graphs of H, $H^{\circ 2}$, and $H^{\circ 3}$

In particular, $H^{\circ m}$ is a sawtooth function taking alternatively the values 0 and 1 at its breakpoints $l2^{-m}$, $l = 0, 1, \ldots, 2^m$, cf. [10, Lemma III.1] or [22, Lemma 2.4]. Moreover, since the restriction of the function $(2^m x - \lfloor 2^m x \rfloor)$ on each interval $[l2^{-m}, (l+1)2^{-m})$ is a linear function passing through $l2^{-m}$ with slope 2^m and $\mathcal{C}(0) = \mathcal{C}(1)$, we have the coincidence

$$\mathcal{C}_{2^m}(x) = \mathcal{C}(2^m x - \lfloor 2^m x \rfloor) = \mathcal{C}(H^{\circ m}(x)), \quad x \in [0,1],$$

16

cf. [7, Page 147]. Therefore, since H and $\mathcal{C} = 1 - 2H$ belong to $\Upsilon^{2,1}$, we obtain from Proposition 4.2(i) that $\mathcal{C}_{2^m} \in \Upsilon^{2,m+1}$ for $m = \mathbb{N}_0$. Concerning $\mathcal{S}$, we deduce $\mathcal{S} \in \Upsilon^{2,2}$ from the identity $\mathcal{S}(x) = \mathcal{C}_2(\frac{x}{2} + \frac{3}{8})$, $x \in [0,1]$, which ultimately yields $\mathcal{S}_{2^m} \in \Upsilon^{2,m+2}$. Moreover, for arbitrary $j \in \mathbb{N}$, we choose the smallest $m \in \mathbb{N}$ such that $j \leq 2^m$ and in view of $\mathcal{C}_j(x) = \mathcal{C}_{2^m}(j2^{-m}x)$, $j \leq 2^m$, see that the following holds.

Lemma 4.3. *Let* $j \in \mathbb{N}$. *Then, restricted to* $[0,1]$,

$$\mathcal{C}_j \in \Upsilon^{2, \lceil \log_2 j \rceil + 1} \quad and \quad \mathcal{S}_j \in \Upsilon^{2, \lceil \log_2 j \rceil + 2}$$

and all the entries of weight matrices and the bias vectors are bounded by 8.

This was already observed in the proof of [7, Theorem 6.2]. We now provide a multivariate version of Lemma 4.3.

Lemma 4.4. *Let* $d > 1$ *and* $\alpha \in \mathbb{Z}^d \setminus \{0\}$. *Then, restricted to* $x \in [0,1]^d$,

$$\mathcal{C}(\alpha \cdot x) \in \Upsilon^{2, \lceil \log_2 \|\alpha\|_1 \rceil + 2} \quad and \quad \mathcal{S}(\alpha \cdot x) \in \Upsilon^{2, \lceil \log_2 \|\alpha\|_1 \rceil + 3},$$

where $\|\alpha\|_1 = |\alpha_1| + \ldots + |\alpha_d|$. *Also in this case, the weights and biases are bounded by 8.*

Proof. We first extend the functions $\mathcal{C}_{2^m}$, $\mathcal{S}_{2^m}$ from $[0,1]$ to the interval $[-1,1]$ by putting

$$\tilde{\mathcal{C}}_{2^m}(x) := \mathcal{C}_{2^{m+1}}\left(\frac{x+1}{2}\right), \quad x \in [-1,1],$$

and deduce that $\tilde{\mathcal{C}}_{2^m} \in \Upsilon^{2,m+2}$. Let now $x \in [0,1]^d$, then from $\alpha \cdot x = \alpha_1 x_1 + \ldots + \alpha_d x_d$ we get

$$\alpha \cdot x \in \big[-\|\alpha\|_1, \|\alpha\|_1 \big].$$

Moreover, using the fact that $\mathcal{C}(\alpha \cdot x) = \tilde{\mathcal{C}}_{2^m}(2^{-m}\alpha \cdot x)$, we choose $m := \lceil \log_2 \|\alpha\|_1 \rceil$ and obtain

$$2^{-m}\alpha \cdot x \in \left[-\frac{\|\alpha\|_1}{2^m}, \frac{\|\alpha\|_1}{2^m} \right] \subset [-1,1],$$

which implies $\mathcal{C}(\alpha \cdot x) \in \Upsilon^{2, \lceil \log_2 \|\alpha\|_1 \rceil + 2}$. The result for $\mathcal{S}(\alpha \cdot x)$ follows by similar considerations. $\square$

Remark 4.5. *Let us point out that we only have an implicit dependence of the length* L *of our approximating neural network on the dimension* d *(which is displayed by the fact that it logarithmically depends on the* ℓ_1-norm of α*).*

Finally, using Lemma 4.4 we obtain a multivariate analogue of [7, Theorem 6.2], where it was shown that one can reproduce linear combinations of $\mathcal{C}_k$ and $\mathcal{S}_k$ via ReLU networks with a good control of the depth L.

Theorem 4.6. *Let* $d \geq 1$ *and let* $k, l \geq 0$ *be integers with* $k + l \geq 1$. *Let* $\{\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_l\} \subset \mathbb{Z}^d \setminus \{0\}$. *Then the function*

$$f(x) = \sum_{i=1}^{k} a_i \mathcal{C}(\alpha_i \cdot x) + \sum_{j=1}^{l} b_j \mathcal{S}(\beta_j \cdot x), \quad x \in [0,1]^d$$

belongs to $\Upsilon^{W,L}$ *with*

$$W = 2(k+l) \quad and \quad L = \max_{\substack{i=1,\ldots,k; \\ j=1,\ldots,l}} \big\{ \lceil \log_2(\|\alpha_i\|_1) \rceil + 2, \lceil \log_2(\|\beta_j\|_1) \rceil + 3 \big\}$$

and the weights and biases in this network are bounded by $\max_{\substack{i=1,\ldots,k; \\ j=1,\ldots,l}} \{8|a_i|, 8|b_j|, 8\}$.

17

Proof. By Proposition 4.2 (ii) and Lemma 4.4, $\mathcal{C}(\alpha_i \cdot x) \in \Upsilon^{2,L}$ and $\mathcal{S}(\beta_j \cdot x) \in \Upsilon^{2,L}$ for every $i = 1, \ldots, k$ and $j = 1, \ldots, l$ if we restrict x to $[0,1]^d$. By Definition 4.1 we have the corresponding representations for $x \in [0,1]^d$ and all admissible i's and j's

$$\mathcal{C}(\alpha_i \cdot x) = A_i^{(L)} \circ \mathrm{ReLU} \circ A_i^{(L-1)} \circ \cdots \circ \mathrm{ReLU} \circ A_i^{(0)}(x) \tag{34}$$

and

$$\mathcal{S}(\beta_j \cdot x) = B_j^{(L)} \circ \mathrm{ReLU} \circ B_j^{(L-1)} \circ \cdots \circ \mathrm{ReLU} \circ B_j^{(0)}(x). \tag{35}$$

If $z = (z_1, \ldots, z_n)$ is a vector in $\mathbb{R}^n$ and $1 \leq u < v \leq n$ are integers, then we denote by $z_{u,v} = (z_u, z_v)$ the restriction of z to the set $\{u, v\}$. Furthermore, we denote $x^{(0)} = (x_1, \ldots, x_d)$ and stack the networks (34) and (35) on top of each other. In this way, we obtain a series of intermediate vectors $x^{(1)}, \ldots, x^{(L)} \in \mathbb{R}^W$

$$x^{(1)} = \mathrm{ReLU} \left[A_1^{(0)} x^{(0)} \quad \ldots \quad A_k^{(0)} x^{(0)} \quad B_1^{(0)} x^{(0)} \quad \ldots \quad B_l^{(0)} x^{(0)} \right]^T,$$

$$x^{(2)} = \mathrm{ReLU} \left[A_1^{(1)} x_{1,2}^{(1)} \quad \ldots \quad A_k^{(1)} x_{2k-1,2k}^{(1)} \quad B_1^{(1)} x_{2k+1,2k+2}^{(1)} \quad \ldots \quad B_l^{(1)} x_{2k+2l-1,2k+2l}^{(1)} \right]^T,$$

$$\vdots$$

$$x^{(L)} = \mathrm{ReLU} \left[A_1^{(L-1)} x_{1,2}^{(L-1)} \quad \ldots \quad A_k^{(L-1)} x_{2k-1,2k}^{(L-1)} \quad B_1^{(L-1)} x_{2k+1,2k+2}^{(L-1)} \quad \ldots \quad B_l^{(L-1)} x_{2k+2l-1,2k+2l}^{(L-1)} \right]^T.$$

Finally, the result follows by observing that

$$y = f(x) = \sum_{i=1}^{k} a_i A_i^{(L)} x_{2i-1,2i}^{(L)} + \sum_{j=1}^{l} b_j B_j^{(L)} x_{2j-1+2k,2j+2k}^{(L)},$$
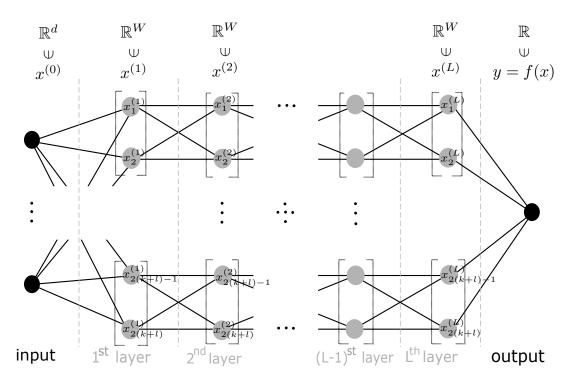
see Figure 8.

Figure 8: Feed-forward ReLU network producing $y = f(x)$

Since by Lemma 4.4, all the weights and biases used in the calculation of $x^{(1)}, \ldots, x^{(L)}$ are bounded by 8, the weights in the last step are then bounded by $\max_{i=1,\ldots,k} 8|a_i|$ and $\max_{j=1,\ldots,l} 8|b_j|$, respectively. $\qquad \square$

References

[1] H. Bölcskei, P. Grohs, G. Kutyniok, and P. Petersen, *Optimal approximation with sparsely connected deep neural networks*, SIAM J. Math. Data Sci. 1 (2019), no. 1, 8–45.

[2] A. Bourouihiya, *The tensor product of frames*, Sampl. Theory Signal Image Process. 7 (2008), no. 1, 65–76.

[3] P. Beneventano, P. Cheridito, R. Graeber, A. Jentzen, and B. Kuckuck, *Deep neural network approximation theory for high-dimensional functions*, available at `arXiv:2112.14523`.

[4] O. Christensen, *Frames and projection method*, Appl. Comput. Harmon. Anal. 1 (1993), 50–53.

[5] O. Christensen, *Frames containing a Riesz basis and approximation of the frame coefficients using finite dimensional methods*, J. Math. Anal. Appl. 199 (1996), 256–270.

[6] O. Christensen, An introduction to frames and Riesz bases, Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2003.

[7] I. Daubechies, R. DeVore, S. Foucart, B. Hanin, and G. Petrova, *Nonlinear Approximation and (Deep) ReLU Networks*, Constr. Appr. 55 (2022), 127–172.

[8] R. DeVore, B. Hanin, and G. Petrova, *Neural network approximation*, Acta Numer. 30 (2021), 327–444.

[9] D. Dung and V. K. Nguyen, *Deep ReLU neural networks in high-dimensional approximation*, Neural Netw. 142 (2021), 619–635.

[10] D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei, *Deep neural network approximation theory*, IEEE Trans. Inform. Theory 67 (2021), no. 5, 2581–2623.

[11] A. Graves, A.-R. Mohamed, and G. E. Hinton, *Speech recognition with deep recurrent neural networks*, In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2013), 6645–6649.

[12] R. Gribonval, G. Kutyniok, M. Nielsen, and F. Voigtlaender, *Approximation spaces of deep neural networks*, Constr. Approx. 55 (2022), no. 1, 259–367

[13] K. Gröchenig, Foundations of time-frequency analysis, Applied and Numerical Harmonic Analysis, Birkhäuser Boston, Inc., Boston, MA, 2001.

[14] R. A. Horn and C. R. Johnson, Matrix analysis, Cambridge University Press, Cambridge, 1985.

[15] H. O. Kim and J. K. Lim, *New characterizations of Riesz bases*, Appl. Comput. Harmon. Anal. 4 (1997), no. 2, 222–229.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Imagenet classification with deep convolutional neural networks*, In: Advances in Neural Information Processing Systems (NIPS) (2012), 1–9.

[17] M. Moravčík, M. Schmid, N. Burch, V. Lisý, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, and M. Bowling, *Deepstack: Expert-level artificial intelligence in heads-up no-limit poker*, Science 356, no. 6337 (2017), 508–513.

[18] P. Petersen and F. Voigtlaender, *Optimal approximation of piecewise smooth functions using deep ReLU neural networks*, Neural Netw. 108 (2018), 296–330.

[19] T. Poggio, H. Mhaskar, L. Rosasco, B. Miranda, Q. Liao, *Why and when can deep-but dot shallow-networks avoid the curse of dimensionality: a review*, Int. J. Autom. Comput. 14 (2017), no. 5, 503–519.

[20] J. Schmidt-Hieber, *Nonparametric regression using deep neural networks with ReLU activation function*, Ann. Statist. 48 (2020), no. 4, 1875–1897.

[21] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser et al., *Mastering the game of Go with deep neural networks and tree search*, Nature 529, no. 7587 (2016), 484–489.

[22] M. Telgarsky, *Representation benefits of deep feedforward networks*, available at `arXiv 1509.08101`.

[23] E. C. Titchmarsh, The theory of the Riemann zeta-function, The Clarendon Press, Oxford University Press, New York, 1986.

[24] Y. Wu, M. Schuster, Z. Chen, Q. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al., *Google's neural machine translation system: Bridging the gap between human and machine translation*, available at `arXiv:1609.08144`.

[25] D. Yarotsky, *Error bounds for approximations with deep ReLU networks*, Neural Netw. 94 (2017), 103–114.

[26] R. M. Young, An introduction to nonharmonic Fourier series, Pure and Applied Mathematics 93, Academic Press, Inc., New York-London, 1980.