On a constructive proof of Kolmogorov's superposition theorem

Jürgen Braun, Michael Griebel

Abstract

Kolmogorov showed in [14] that any multivariate continuous function can be represented as a superposition of one–dimensional functions, i.e.

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^{n} \psi_{q,p}(x_p) \right) .$$

The proof of this fact, however, was not constructive and it was not clear how to choose the outer and inner functions Φ_q and $\psi_{q,p}$ respectively. Sprecher gave in [27, 28] a constructive proof of Kolmogorov's superposition theorem in form of a convergent algorithm which defines the inner functions explicitly via one inner function ψ by $\psi_{p,q} := \lambda_p \psi(x_p + qa)$ with appropriate values $\lambda_p, a \in \mathbb{R}$. Basic features of this function as monotonicity and continuity were supposed to be true, but were not explicitly proved and turned out to be not valid. Köppen suggested in [16] a corrected definition of the inner function ψ and claimed, without proof, its continuity and monotonicity. In this paper we now show that these properties indeed hold for Köppen's ψ and present a correct constructive proof of Kolmogorov's superposition theorem for continuous inner functions ψ similar to Sprecher's approach.

Keywords: Kolmogorov's superposition theorem, superposition of functions, representation of functions

AMS-Classification: 26B40

1 Introduction

The description of multivariate continuous functions as a superposition of a number of continuous functions [13, 24] is closely related to Hilbert's thirteenth problem [10] from his Paris lecture in 1900. In 1957 the Russian mathematician Kolmogorov showed the remarkable fact that any continuous function f of many variables can be represented as a composition of addition and some functions of one variable [14]. The original version of this theorem can be expressed as follows:

Theorem. *Let $f : \mathbb{I}^n := [0, 1]^n \to \mathbb{R}$ be an arbitrary multivariate continuous function. Then it has the representation*

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^{n} \psi_{q,p}(x_p) \right) , \tag{1.1}$$

with continuous one–dimensional outer and inner functions Φ_q and $\psi_{q,p}$. All these functions Φ_q, $\psi_{q,p}$ are defined on the real line. The inner functions $\psi_{q,p}$ are independent of the function f.

Kolmogorov's student Arnold also made contributions [1–3] in this context that appeared at nearly the same time. Several improvements of Kolmogorov's original version were published in the following years. Lorentz showed that the outer functions Φ_q can be chosen to be the same [19, 20] while Sprecher proved that the inner functions $\psi_{q,p}$ can be replaced by $\lambda_p \psi_q$ with appropriate constants λ_p [25, 26]. A proof of Lorentz's version with one outer function that is based on the Baire category theorem was given by Hedberg [9] and Kahane [13]. A further improvement was made by Friedman [5], who showed that the inner functions can be chosen to be Lipschitz continuous. A geometric interpretation of the theorem is that the $2n+1$ inner sums $\sum_{p=1}^{n} \psi_{q,p}$ map the unit cube $\mathbb{I}^n$ homeomorphically onto a compact set $\Gamma \subset \mathbb{R}^{2n+1}$. Ostrand [23] and Tikhomirov [15] extended Kolmogorov's theorem to arbitrary n–dimensional metric compact sets. The fact that any compact set $K \subset \mathbb{R}^n$ can be homeomorphically embedded into $\mathbb{R}^{2n+1}$ was already known from the Menger–Nöbeling theorem [11].

More recently, Kolmogorov's superposition theorem found attention in neural network computation by Hecht–Nielsen's interpretation as a feed-forward network with an input layer, one hidden layer and an output layer [7, 8, 25]. However, the inner functions in all these versions of Kolmogorov's theorem are highly non-smooth. Also, the outer functions depend on the specific function f and hence are not representable in a parameterized form. Moreover, all one–dimensional functions are the limits or sums of some infinite series of functions, which cannot be computed practically. Therefore Girosi and Poggio [6] made the criticism that such an approach is not applicable in neurocomputing.

The original proof of Kolmogorov's theorem is not constructive, i.e. one can show the existence of a representation (1.1) but it cannot be used in an algorithm for numerical calculations. Kurkova [17, 18] partly eliminated these difficulties by substituting the exact representation in (1.1) with an approximation of the function f. She replaced the one–variable functions with finite linear combinations of affine transformations of a single arbitrary sigmoidal function ψ. Her direct approach also enabled an estimation of the number of hidden units (neurons) as a function of the desired accuracy and the modulus of continuity of f being approximated. In [21] a constructive algorithm is proposed that approximates a function f to any desired accuracy with one single design, which means that no additional neurons have to be added. There, also a short overview of the history of Kolmogorov's superposition theorem in neural network computing is given. Other approximative, but constructive approaches to function approximation by generalizations of Kolmogorov's superposition theorem can be found in [4, 12, 22].

Recently, Sprecher derived in [27, 28] a numerical algorithm for the implementation of both internal and external univariate functions, which promises to constructively prove Kolmogorov's superposition theorem. In these articles, the inner functions ψ_q are defined as translations of a single function ψ that is explicitly defined as an extension of a function which is defined on a dense subset of the real line. There, the r–th iteration step of Sprecher's algorithm works as follows: For a chosen appropriate basis $\gamma \in \mathbb{N}_+$, the n–dimensional unit cube $[0, 1]^n$ is divided into subcubes which are separated by small gaps whose sizes depend on γ. Also, Sprecher's definition of the inner function is based on this γ such that, for fixed q, the corresponding inner sum maps the subcubes into intervals on the real line. These intervals are then again separated by gaps. This allows the definition of a continuous outer function Φ_q^r on the intervals such that the residual f_r between f and the previous iterate is approximated on the subcubes by the superposition of the q–th outer function and an inner sum. Since the approximation error cannot be controlled on the gaps, the cubes are additionally translated by a variation of the q's. This is done such that for each point $\mathbf{x} \in [0, 1]^n$ the set of q–values for which $\mathbf{x}$ is contained in a subcube is larger than the set for which it lies in a gap. The r–th approximation is then defined as the sum over all values of q and the previous iterate. Sprecher proved convergence of this algorithm in [27, 28]. Throughout this proof, he relied on continuity and monotonicity of the resulting ψ. It can however be shown that his ψ does not possess these important properties. This was already observed by Köppen in [16] where a modified inner function ψ was suggested. Köppen claims, but does not prove the continuity of his ψ and merely comments on the termination of the recursion which defines his corrected function ψ.

In this article we close these gaps. First, since the recursion is defined on a dense subset of $\mathbb{R}$, it is necessary to show the existence of an expansion of Köppen's ψ to the real line. We give this existence proof. Moreover it is also a priori not clear that Köppen's ψ possesses continuity and monotonicity, which are necessary to proof the convergence of Sprecher's algorithm and therefore Kolmogorov's superposition theorem. We provide these properties. Altogether, we thus derive a complete constructive proof of Kolmogorov's superposition theorem along the lines of Sprecher but based on Köppen's ψ.

The remainder of this article is organized as follows: As starting point, we specify Sprecher's version of Kolmogorov's superposition theorem in section 2. Then, in section 3 we briefly repeat the definitions of the original inner function ψ and the constructive algorithm that was developed by Sprecher in [27,28]. The convergence of this algorithm would prove Kolmogorov's superposition theorem. First, we observe that Sprecher's ψ is neither continuous nor monotone increasing on the whole interval $[0,1]$. We then show that Köppen's ψ indeed exists, i.e. it is well defined and has the necessary continuity and monotonicity properties. Endowed with this knowledge, we then follow Sprecher's lead and prove the convergence of the algorithm, where the original inner function is replaced by the corrected one. This finally gives a constructive proof of Kolmogorov's superposition theorem.

2 Definitions and algorithm

2.1 A version of Kolmogorov's superposition theorem

Many different variants of Kolmogorov's superposition theorem (1.1) were developed since the first publication of this remarkable result in 1957. Some improvements can be found e.g. in [20,25]. In [5] it was shown that the inner functions $\psi_{q,p}$ can be chosen to be Lipschitz continuous with exponent one. Another variant with only one outer function and $2n+1$ inner functions was derived in [20]. A version of Kolmogorov's superposition theorem recently developed by Sprecher in [25] reads as follows:

Theorem 2.1. *Let $n \geq 2$, $m \geq 2n$ and $\gamma \geq m+2$ be given integers and let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{x}_q = (x_1 + qa, \ldots, x_n + qa)$, where $a := [\gamma(\gamma-1)]^{-1}$. Then, for any arbitrary continuous function $f : \mathbb{R}^n \to \mathbb{R}$, there exist $m+1$ continuous functions $\Phi_q : \mathbb{R} \to \mathbb{R}$, $q = 0, \ldots m$, such that*

$$f(\mathbf{x}) \;=\; \sum_{q=0}^{m} \Phi_q \circ \xi(\mathbf{x}_q)\,, \quad \textit{with} \quad \xi(\mathbf{x}_q) \;=\; \sum_{p=1}^{n} \alpha_p \, \psi(x_p + qa)\,, \tag{2.1}$$

$\alpha_1 = 1$, $\alpha_p = \sum_{r=1}^{\infty} \gamma^{-(p-1)\beta(r)}$ for $p > 1$ and $\beta(r) = (n^r - 1)/(n-1)$.

This version of Kolmogorov's superposition theorem involves m one–dimensional outer functions Φ_q and one single inner function ψ. The definition of ψ will be discussed in detail in the following.

For a fixed base $\gamma > 1$ we define for any $k \in \mathbb{N}$ the set of terminating rational numbers

$$\mathcal{D}_k \;=\; \mathcal{D}_k(\gamma) \;:=\; \left\{ d_k \in \mathbb{Q} \;:\; d_k = \sum_{r=1}^{k} i_r \gamma^{-r}, i_r \in \{0, \ldots, \gamma - 1\} \right\}. \tag{2.2}$$

Then the set

$$\mathcal{D} \ := \ \bigcup_{k\in\mathbb{N}} \mathcal{D}_k \tag{2.3}$$

is dense in $[0,1]$.

In [28] Sprecher formulated an algorithm, whose convergence proves the above theorem 2.1 constructively. In this algorithm, the inner function ψ was defined point-wise on the set $\mathcal{D}$. Further investigations on this function were made in [27]. However, to make this proof work, two fundamental properties of ψ namely continuity and monotonicity are needed. Unfortunately, the inner function ψ in [27, 28] is neither continuous nor monotone. In the following, we repeat the definition of ψ here and show that it indeed does not define a continuous and monotone increasing function.

Let $\langle i_1 \rangle := 0$ and for $r \geq 2$ let

$$\langle i_r \rangle \ := \ \begin{cases} 0 & \text{when } i_r = 0, 1 \ldots, \gamma - 2\,, \\ 1 & \text{when } i_r = \gamma - 1\,. \end{cases}$$

Furthermore, we define $[i_1] := 0$ and, for $r \geq 2$,

$$[i_r] \ := \ \begin{cases} 0 & \text{when } i_r = 0, 1 \ldots, \gamma - 3\,, \\ 1 & \text{when } i_r = \gamma - 2, \gamma - 1\,, \end{cases}$$

$$\tilde{\imath}_r := i_r - (\gamma - 2)\langle i_r \rangle\,,$$

and

$$m_r \ := \ \langle i_r \rangle \left(\sum_{s=1}^{r-1} \Big([i_s] \cdot \ldots \cdot [i_{r-1}] \Big) \right)\,.$$

The function ψ is then defined on $\mathcal{D}_k$ by

$$\psi(d_k) \ := \ \sum_{r=1}^{k} \tilde{\imath}_r \, 2^{-m_r} \, \gamma^{-\beta(r-m_r)}\,. \tag{2.4}$$

Note that the definition of ψ depends on the dimension n since $\beta(\cdot)$ depends on n. For a simpler notation we dispense with an additional index. The graph of the function ψ is depicted in figure 1 for $k = 5$, $\gamma = 10$ and $n = 2$, i.e. it was calculated with the definition (2.4) on the set of rational decimal numbers $\mathcal{D}_k$. The function ψ from (2.4) has an extension to $[0,1]$, which also will be denoted by ψ if the meaning is clear from the contents.

The following calculation shows directly that this function is not continuous in contrast to the claim in [27]. With the choice $\gamma = 10$ and $n = 2$ one gets with the definition (2.4) the function values

$$\psi(0.58999) \ = \ 0.55175 \quad \text{and} \quad \psi(0.59) \ = \ 0.55\,. \tag{2.5}$$

This counter–example shows that the function ψ is not monotone increasing. We furthermore can see from the additive structure of ψ in (2.4) that

$$\psi(0.58999) \ < \ \psi(x) \quad \text{for all} \quad x \in \big(0.58999, 0.59\big)\,. \tag{2.6}$$

This shows that the function ψ is also not continuous.

Remark 2.2. *Discontinuities of ψ arise for all values $x = 0.i_1 9\,,\ i_1 = 0, \ldots, 9$.*

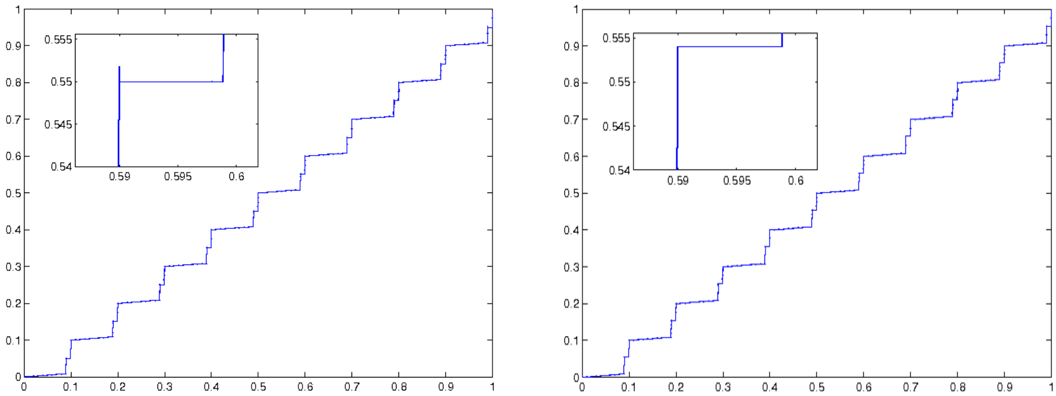

Figure 1: The graph of Sprecher's ψ from (2.4) on the interval $[0,1]$ (left) and a zoom into a smaller interval (small, left), computed for the values of the set $\mathcal{D}_5$, $\gamma = 10$ and $n = 2$. One can clearly see the non–monotonicity and discontinuity near the value $x = 0.59$. The right image shows Köppen's version from (2.7) for the same parameters and a zoom into the same region (small, right). Here, the discontinuity is no longer present.

Among other things, the convergence proof in [27,28] is based on continuity and monotonicity of ψ. As the inner function defined by Sprecher does not provide these properties the convergence proof also becomes invalid unless the definition of ψ is properly modified. To this end, Köppen suggested in [16] a corrected version of the inner function and stated its continuity. This definition of ψ is also restricted to the dense set of terminating rational numbers $\mathcal{D}$. Köppen defines recursively

$$\psi_k(d_k) \;=\; \begin{cases} d_k & \text{for } k = 1\,, \\ \psi_{k-1}(d_k - \frac{i_k}{\gamma^k}) + \frac{i_k}{\gamma^{\beta(k)}} & \text{for } k > 1 \text{ and } i_k < \gamma - 1\,, \\ \frac{1}{2}\Big(\psi_{k-1}(d_k - \frac{i_k}{\gamma^k}) + \psi_{k-1}(d_k + \frac{1}{\gamma^k}) + \frac{i_k}{\gamma^{\beta(k)}}\Big) & \text{for } k > 1 \text{ and } i_k = \gamma - 1 \end{cases} \tag{2.7}$$

and claimed that this recursion terminates. He assumed that there exists an extension from the dense set $\mathcal{D}$ to the real line as in Sprecher's construction and that this extended ψ is monotone increasing and continuous but did not give a proof for it. In the following, we provide such a proof. The function ψ_k is depicted in figure 1 for the same parameters $k = 5$, $\gamma = 10$ and $n = 2$ as before.

We first consider the existence of an extension and begin with the remark that every real number $x \in [0,1]$ has a representation

$$x \;=\; \sum_{r=1}^{\infty} \frac{i_r}{\gamma^r} = \lim_{k\to\infty} \sum_{r=1}^{k} \frac{i_r}{\gamma^r} = \lim_{k\to\infty} d_k\,.$$

For such a value x, we define the inner function

$$\psi(x) \;:=\; \lim_{k\to\infty} \psi_k(d_k) \;=\; \lim_{k\to\infty} \psi_k\left(\sum_{r=1}^{k} \frac{i_r}{\gamma^r}\right) \tag{2.8}$$

and show the existence of this limit.

For the following calculations it is advantageous to have an explicit representation of (2.7) as a sum. To this end, we need some further definitions. The values of ψ_{k-j} at the rational points d_{k-j}, $d_{k-j} + \frac{1}{\gamma^{k-j}} \in [0,1]$ are denoted as

$$\psi_{k-j} := \psi_{k-j}(d_{k-j}) \quad \text{and} \quad \psi^+_{k-j} := \psi_{k-j}(d_{k-j} + \frac{1}{\gamma^{k-j}})\,.$$

Then, the recursion (2.7) takes for $k-j>1$ the form

$$\psi_{k-j} = \begin{cases} \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \psi_{k-j-1} & \text{for} \quad i_{k-j} < \gamma - 1, \\ \frac{\gamma-2}{2\gamma^{\beta(k-j)}} + \frac{1}{2}\psi_{k-j-1} + \frac{1}{2}\psi^+_{k-j-1} & \text{for} \quad i_{k-j} = \gamma - 1 \end{cases} \tag{2.9}$$

and

$$\psi^+_{k-j} = \begin{cases} \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \frac{1}{\gamma^{\beta(k-j)}} + \psi_{k-j-1} & \text{for} \quad i_{k-j} < \gamma - 2, \\ \frac{i_{k-j}}{2\gamma^{\beta(k-j)}} + \frac{1}{2}\psi_{k-j-1} + \frac{1}{2}\psi^+_{k-j-1} & \text{for} \quad i_{k-j} = \gamma - 2, \\ \psi^+_{k-j-1} & \text{for} \quad i_{k-j} = \gamma - 1\,. \end{cases} \tag{2.10}$$

Using the values

$$s_j \;:=\; \begin{cases} 0 & \text{for} \quad i_{k-j+1} < \gamma - 2, \\ \frac{1}{2} & \text{for} \quad i_{k-j+1} = \gamma - 2, \\ 1 & \text{for} \quad i_{k-j+1} = \gamma - 1 \end{cases} \qquad \text{and} \qquad \tilde{s}_j \;:=\; \begin{cases} 0 & \text{for} \quad i_{k-j+1} < \gamma - 1, \\ \frac{1}{2} & \text{for} \quad i_{k-j+1} = \gamma - 1, \end{cases} \tag{2.11}$$

we can define the matrix and vector

$$\mathbf{M}_j \;:=\; \begin{pmatrix} (1-\tilde{s}_{j+1}) & \tilde{s}_{j+1} \\ (1-s_{j+1}) & s_{j+1} \end{pmatrix} \quad \text{and} \quad \mathbf{b}_j \;:=\; \begin{pmatrix} (1-2\tilde{s}_{j+1})\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1}\frac{\gamma-2}{\gamma^{\beta(k-j)}} \\ (1-s_{j+1})\Big[\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1-2s_{j+1})\frac{1}{\gamma^{\beta(k-j)}}\Big] \end{pmatrix}.$$

Now, the representations (2.9) and (2.10) can be brought into the more compact form

$$\begin{pmatrix} \psi_{k-j} \\ \psi^+_{k-j} \end{pmatrix} \;=\; \mathbf{M}_j \begin{pmatrix} \psi_{k-j-1} \\ \psi^+_{k-j-1} \end{pmatrix} \;+\; \mathbf{b}_j\,. \tag{2.12}$$

Next, we define the values $\theta_0 := 1$, $\theta^+_0 := 0$, $\theta_1 := 1 - \tilde{s}_1$, $\theta^+_1 := \tilde{s}_1$, and set recursively

$$\begin{pmatrix} \theta_{j+1} \\ \theta^+_{j+1} \end{pmatrix} \;:=\; \mathbf{M}_j^T \begin{pmatrix} \theta_j \\ \theta^+_j \end{pmatrix} \tag{2.13}$$

for $j = 1, \ldots, k-1$. By induction we can directly deduce from (2.13) and (2.11) the useful properties

$$\theta_j + \theta^+_j = 1 \quad \text{and} \quad \theta_j, \theta^+_j > 0\,. \tag{2.14}$$

With these definitions, the ξ–th step of the recursion can be written as the sum

$$\begin{aligned} \psi_k \;=\; & \sum_{j=0}^{\xi-1} \theta_j \Big[(1-2\tilde{s}_{j+1})\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1}\frac{\gamma-2}{\gamma^{\beta(k-j)}}\Big] \\ & + \theta^+_j \Big[(1-s_{j+1})\Big(\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1-2s_{j+1})\frac{1}{\gamma^{\beta(k-j)}}\Big)\Big] \;+\; \theta_\xi\,\psi_{k-\xi} \;+\; \theta^+_\xi\psi^+_{k-\xi}\,. \end{aligned} \tag{2.15}$$

Proof by induction. $\xi = 1$: From (2.12) for $j = 0$ we directly get

$$\psi_k \;=\; \Big[(1-2\tilde{s}_1)\frac{i_k}{\gamma^{\beta(k)}} + \tilde{s}_1\frac{\gamma-2}{\gamma^{\beta(k)}}\Big] \;+\; (1-\tilde{s}_1)\,\psi_{k-1} \;+\; \tilde{s}_1\psi^+_{k-1}\,. \qquad \surd$$

$\xi \to \xi + 1$: With (2.12) and (2.13) we have

$$\begin{aligned}
\psi_k \;=\; & \sum_{j=0}^{\xi-1} \theta_j \Big[(1-2\tilde{s}_{j+1})\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1}\frac{\gamma-2}{\gamma^{\beta(k-j)}}\Big] \\
& \quad + \theta_j^+ \Big[(1-s_{j+1})\Big(\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1-2s_{j+1})\frac{1}{\gamma^{\beta(k-j)}}\Big)\Big] \\
& + \theta_\xi \Big[(1-\tilde{s}_{\xi+1})\psi_{k-(\xi+1)} + \tilde{s}_{\xi+1}\psi^+_{k-(\xi+1)} + (1-2\tilde{s}_{\xi+1})\frac{i_{k-\xi}}{\gamma^{\beta(k-\xi)}} + \tilde{s}_{\xi+1}\frac{\gamma-2}{\gamma^{\beta(k-\xi)}}\Big] \\
& + \theta_\xi^+ \Big[(1-s_{\xi+1})\psi_{k-(\xi+1)} + s_{\xi+1}\psi^+_{k-(\xi+1)} + (1-s_{\xi+1})\Big[\frac{i_{k-\xi}}{\gamma^{\beta(k-\xi)}} + (1-2s_{\xi+1})\frac{1}{\gamma^{\beta(k-\xi)}}\Big]\Big] \\
= \; & \sum_{j=0}^{\xi} \theta_j \Big[(1-2\tilde{s}_{j+1})\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1}\frac{\gamma-2}{\gamma^{\beta(k-j)}}\Big] \\
& \quad + \theta_j^+ \Big[(1-s_{j+1})\Big(\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1-2s_{j+1})\frac{1}{\gamma^{\beta(k-j)}}\Big)\Big] \;+\; \theta_{\xi+1}\,\psi_{k-(\xi+1)} \;+\; \theta^+_{\xi+1}\psi^+_{k-(\xi+1)}\,.
\end{aligned}$$

□

Choosing $\xi = k-1$ we finally obtain a point-wise representation of the function ψ_k as the direct sum

$$\begin{aligned}
\psi_k(d_k) \;=\; & \sum_{j=0}^{k-2} \theta_j \Big[(1-2\tilde{s}_{j+1})\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1}\frac{\gamma-2}{\gamma^{\beta(k-j)}}\Big] \\
& \quad + \theta_j^+ \Big[(1-s_{j+1})\Big(\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1-2s_{j+1})\frac{1}{\gamma^{\beta(k-j)}}\Big)\Big] \;+\; \theta_{k-1}\,\frac{i_1}{\gamma} \;+\; \theta^+_{k-1}\frac{i_1+1}{\gamma}\,.
\end{aligned} \tag{2.16}$$

Now we have to show the existence of the limit (2.8). To this end, we consider the behavior of the function values ψ_k and ψ_k^+ as k tends to infinity:

Lemma 2.3. *For growing values of k one has for ψ_k defined in (2.9) and ψ_k^+ from (2.10)*

$$\psi_k^+ \;=\; \psi_k \;+\; \mathcal{O}(2^{-k})\,.$$

Proof. With (2.12), the fact that $\gamma^{\beta(j)} = \gamma^{\beta(j-1)}\gamma^{n^{j-1}}$ and $\gamma^n > 2$, we have

$$\begin{aligned}
|\psi_k^+ - \psi_k| \le\; & \frac{1}{2}|\psi^+_{k-1} - \psi_{k-1}| \;+\; \frac{\gamma-2}{\gamma^{\beta(k)}} \\
\le\; & \left(\frac{1}{2}\right)^{k-1} |\psi_1^+ - \psi_1| \;+\; \left(\frac{1}{2}\right)^{k-2} (\gamma-2)\left(\sum_{j=2}^{k} \frac{2^{j-2}}{\gamma^{\beta(j)}}\right) \\
\le\; & \left(\frac{1}{2}\right)^{k-1} |\psi_1^+ - \psi_1| \;+\; \left(\frac{1}{2}\right)^{k-2} (\gamma-2)\left(\sum_{j=0}^{\infty} \left(\frac{2}{\gamma^n}\right)^{j-2}\right) \\
=\; & \left(\frac{1}{2}\right)^{k-2} \left[\frac{1}{2\gamma} \;+\; \frac{(\gamma-2)\gamma^n}{\gamma^n-2}\right]
\end{aligned}$$

and the assertion is proved. □

If we now apply this result to arbitrary values k and k', we can show the following lemma:

Lemma 2.4. *The sequence ψ_k defined in* (2.9) *is a Cauchy sequence.*

Proof. For k, $k' \in \mathbb{N}$ and without loss of generality $k > k'$, we set $\xi := k - k'$ in (2.15). Then, we obtain by (2.14) and with lemma 2.3 the following estimate:

$$
\begin{aligned}
|\psi_k - \psi_{k'}| &\leq \; \left|\theta_{k-k'}\psi_{k'} + \theta^+_{k-k'}\psi^+_{k'} - \psi_{k'}\right| \; + \; 2(\gamma-2) \sum_{j=k'+1}^{k} \frac{1}{\gamma^{\beta(j)}} \\
&\leq \; \left|\theta_{k-k'}\psi_{k'} + \theta^+_{k-k'}\psi^+_{k'} - \psi_{k'}\right| \; + \; 2(\gamma-2) \sum_{j=k'+1}^{k} \left(\frac{1}{\gamma^n}\right)^{j-1} \\
&= \; \left|\theta_{k-k'}\psi_{k'} + \theta^+_{k-k'}\left(\psi_{k'} + \mathcal{O}(2^{-k'})\right) - \psi_{k'}\right| \; + \; \frac{2\gamma^n(\gamma-2)}{1-\gamma^n} \left(\left(\frac{1}{\gamma^n}\right)^k - \left(\frac{1}{\gamma^n}\right)^{k'} \right) \\
&\leq \; \mathcal{O}(2^{-k'}) \; + \; \frac{2\gamma^n(\gamma-2)}{1-\gamma^n} \left(\left(\frac{1}{\gamma^n}\right)^k - \left(\frac{1}{\gamma^n}\right)^{k'} \right) .
\end{aligned}
$$

The right hand side tends to 0 when $k, k' \longrightarrow \infty$. □

The real numbers $\mathbb{R}$ are complete and we therefore can infer the existence of a function value for all $x \in [0, 1]$. Thus the function ψ from (2.8) is well defined. It remains to show that this ψ is continuous and monotone increasing. This will be the topic of the following subsections.

2.2 The continuity of ψ

We now show the continuity of the inner function ψ. To this end we first recall some properties of the representations of real numbers.

Let

$$
x \; := \; \sum_{r=1}^{\infty} \frac{i_r}{\gamma^r} \qquad \text{and} \qquad x_0 \; := \; \sum_{r=1}^{\infty} \frac{i_{0,r}}{\gamma^r}
$$

be the representation of the values x and x_0 in the base γ, respectively. Let $x_0 \in (0, 1)$ be given and

$$
\delta(k_0) \; := \; \min \left\{ \sum_{r=k_0+1}^{\infty} \frac{i_{0,r}}{\gamma^r}, \frac{1}{\gamma^{k_0}} - \sum_{r=k_0+1}^{\infty} \frac{i_{0,r}}{\gamma^r} \right\} .
$$

For any $x \in \left(x_0 - \delta(k_0), x_0 + \delta(k_0)\right)$ it follows that

$$
i_r \; = \; i_{0,r} \quad \text{for} \quad r = 1, \ldots, k_0 \, . \tag{2.17}
$$

Special attention has to be paid to the values $x_0 = 0$ and $x_0 = 1$. In both cases, we can choose $\delta(k_0) = \gamma^{-k_0}$. Then (2.17) holds for all $x \in \left[0, \delta(k_0)\right)$ if $x_0 = 0$ and all $x \in \left(1 - \delta(k_0), 1\right]$ if $x_0 = 1$. The three different cases are depicted in figure 2.

Altogether we thus can find for any given arbitrary $x_0 \in [0, 1]$ a δ–neighborhood

$$
U := \left(x_0 - \delta(k_0), x_0 + \delta(k_0)\right) \cap [0, 1]
$$

in which (2.17) holds. To show the continuity of the inner function ψ in x_0, we now choose this neighborhood and see from (2.16) for $x, x_0 \in U$:

$$
\begin{aligned}
|\psi(x)-\psi(x_0)| &= \lim_{k\to\infty} |\psi(d_k)-\psi(d_{0,k})| \\
&= \lim_{k\to\infty} \Bigg| \sum_{j=0}^{k-k_0-1} \theta_j \Big[(1-2\tilde{s}_{j+1})\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1}\frac{\gamma-2}{\gamma^{\beta(k-j)}}\Big] \\
&\qquad + \theta_j^+ \Big[(1-s_{j+1})\Big(\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1-2s_{j+1})\frac{1}{\gamma^{\beta(k-j)}}\Big)\Big] \\
&\quad - \sum_{j=0}^{k-k_0-1} \theta_{0,j} \Big[(1-2\tilde{s}_{0,j+1})\frac{i_{0,k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{0,j+1}\frac{\gamma-2}{\gamma^{\beta(k-j)}}\Big] \\
&\qquad + \theta_{0,j}^+ \Big[(1-s_{0,j+1})\Big(\frac{i_{0,k-j}}{\gamma^{\beta(k-j)}} + (1-2s_{0,j+1})\frac{1}{\gamma^{\beta(k-j)}}\Big)\Big]\Bigg| \\
&\le \lim_{k\to\infty} \sum_{j=0}^{k-k_0-1} \Big|\theta_j \Big[(1-2\tilde{s}_{j+1})\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1}\frac{\gamma-2}{\gamma^{\beta(k-j)}}\Big]\Big| \\
&\qquad + \Big|\theta_j^+ \Big[(1-s_{j+1})\Big(\frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1-2s_{j+1})\frac{1}{\gamma^{\beta(k-j)}}\Big)\Big]\Big| \\
&\quad + \lim_{k\to\infty} \sum_{j=0}^{k-k_0-1} \Big|\theta_{0,j} \Big[(1-2\tilde{s}_{0,j+1})\frac{i_{0,k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{0,j+1}\frac{\gamma-2}{\gamma^{\beta(k-j)}}\Big]\Big| \\
&\qquad + \Big|\theta_{0,j}^+ \Big[(1-s_{0,j+1})\Big(\frac{i_{0,k-j}}{\gamma^{\beta(k-j)}} + (1-2s_{0,j+1})\frac{1}{\gamma^{\beta(k-j)}}\Big)\Big]\Big| \\
&\le \lim_{k\to\infty} \frac{4\gamma^n(\gamma-2)}{1-\gamma^n} \left| \left(\frac{1}{\gamma^n}\right)^k - \left(\frac{1}{\gamma^n}\right)^{k_0} \right| = \frac{4\gamma^n(\gamma-2)}{1-\gamma^n} \left(\frac{1}{\gamma^n}\right)^{k_0}.
\end{aligned}
\tag{2.18}
$$

Note that the estimation of the last two sums was derived in a similar way to that of the proof of lemma 2.4.

In conclusion we can find for any given $\varepsilon > 0$ a $k_0 \in \mathbb{N}$ and thus a $\delta(k_0) > 0$ such that $|\psi(x)-\psi(x_0)| < \varepsilon$ whenever $x, x_0 \in U = \big(x_0 - \delta(k_0), x_0 + \delta(k_0)\big) \cap [0,1]$. This is just the definition of continuity of ψ in $x_0 \in (0,1)$. Since the interval U is only open to the right if $x_0 = 0$ and open to the left if $x_0 = 1$, the inequality (2.18) also shows for these two cases continuity from the right and from the left, respectively. We hence have proved the following theorem:

Theorem 2.5. *The inner function ψ from* (2.8) *is continuous on* $[0,1]$.

2.3 The monotonicity of ψ

A further crucial property of the function ψ is its monotonicity. We show this first on the dense subset $\mathcal{D} \subset \mathbb{R}$ of terminating rational numbers. Note that the values ψ_k and ψ_k^+ from (2.9) and (2.10) are evaluations of ψ on the dense subset of rational numbers in $[0,1]$.

Lemma 2.6. *For every $k \in \mathbb{N}$, there holds*

$$\psi_k^+ \;\ge\; \psi_k + \frac{1}{\gamma^{\beta(k)}}\,.$$

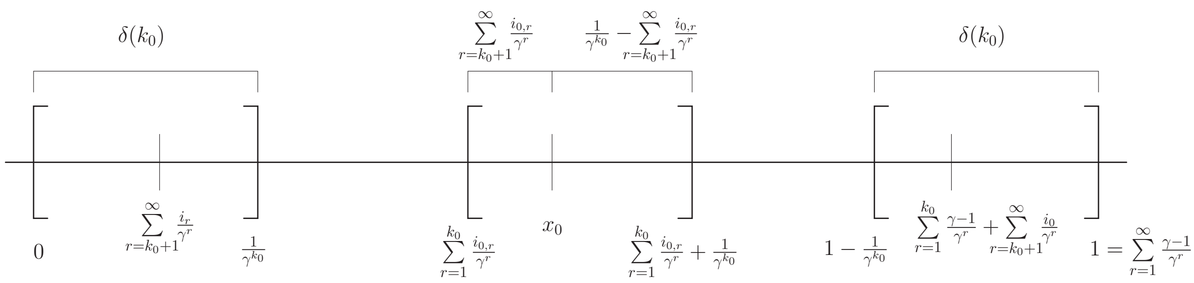

Figure 2: The figure shows the interval $[0,1]$. For any two values x_1 and x_2 that both lie in one of the depicted small intervals it holds that $i_{1,r} = i_{2,r}$ for $r = 1, \ldots, k_0$. The three intervals represent the possible cases that occur in the proof of theorem 2.5.

Proof by induction. $k = 1$:

$$\psi_1^+ - \psi_1 \;=\; \psi_1(d_1 + \frac{1}{\gamma}) - \psi_1(d_1) \;=\; d_1 + \frac{1}{\gamma} - d_1 \;=\; \frac{1}{\gamma} \;=\; \frac{1}{\gamma^{\beta(1)}} \qquad \surd$$

$k \to k+1$:

$$\begin{aligned}
\psi_{k+1}^+ - \psi_{k+1} \;&=\; (s_0 - \tilde{s}_0)(\psi_k^+ - \psi_k) + \frac{1}{\gamma^{\beta(k+1)}}\big((2\tilde{s}_0 - s_0)i_{k+1} + (1-s_0)(1-2s_0) - \tilde{s}_0(\gamma-2)\big) \\
&=\; \begin{cases} \frac{1}{\gamma^{\beta(k+1)}} & \text{for} \quad i_{k+1} < \gamma - 2 \quad (s_0 = \tilde{s}_0 = 0)\,, \\ \frac{1}{2}(\psi_k^+ - \psi_k) - \frac{1}{2}\frac{\gamma-2}{\gamma^{\beta(k+1)}} & \text{for} \quad i_{k+1} = \gamma - 2 \quad (s_0 = \frac{1}{2},\ \tilde{s}_0 = 0)\,, \\ \frac{1}{2}(\psi_k^+ - \psi_k) - \frac{1}{2}\frac{\gamma-2}{\gamma^{\beta(k+1)}} & \text{for} \quad i_{k+1} = \gamma - 1 \quad (s_0 = 1,\ \tilde{s}_0 = \frac{1}{2})\,. \end{cases}
\end{aligned}$$

For the first case $i_{k+1} < \gamma - 2$, the assertion is trivial. For the other two cases, we have

$$\frac{1}{2}(\psi_k^+ - \psi_k) - \frac{1}{2}\frac{\gamma-2}{\gamma^{\beta(k+1)}} \;\geq\; \frac{1}{2}(\psi_k^+ - \psi_k) - \frac{1}{2}\frac{\gamma-2}{\gamma^{\beta(k+1)}} \;\geq\; \frac{1}{2}\left(\frac{1}{\gamma^{\beta(k)}} - \frac{\gamma-2}{\gamma^{\beta(k+1)}}\right) \;\geq\; \frac{1}{\gamma^{\beta(k+1)}}\,. \qquad \surd$$

Here, the validity of the last estimate can be obtained from

$$\begin{aligned}
&\frac{1}{2}\left(\frac{1}{\gamma^{\beta(k)}} - \frac{\gamma-2}{\gamma^{\beta(k+1)}}\right) \;\geq\; \frac{1}{\gamma^{\beta(k+1)}} \quad \Leftrightarrow \quad \frac{1}{2}\left(\frac{\gamma^{n^{k-1}}}{\gamma^{\beta(k+1)}} - \frac{\gamma-2}{\gamma^{\beta(k+1)}}\right) \;\geq\; \frac{1}{\gamma^{\beta(k+1)}} \\
\Leftrightarrow \quad & \gamma^{n^{k-1}} - \gamma + 2 \;\geq\; 2 \quad \Leftrightarrow \quad \gamma^{n^{k-1}} \;\geq\; \gamma \quad \Leftrightarrow \quad n^{k-1} \;\geq\; 1\,. \qquad \surd
\end{aligned}$$

$\square$

We have thus shown that ψ is strictly monotone increasing on a dense subset of $[0,1]$. Since the function is continuous, this holds for the whole interval $[0,1]$. This proves the following theorem:

Theorem 2.7. *The function ψ from (2.8) is monotone increasing on $[0,1]$.*

In summary, we have demonstrated that the inner function ψ defined by Sprecher (c.f. [27,28]) is neither continuous nor monotone increasing, whereas the definition (2.8) of ψ by Köppen from [16] possesses these properties.

3 The algorithm of Sprecher

We will now demonstrate that Sprecher's constructive algorithm from [28] with Köppen's definition of the inner function ψ from [16] is indeed convergent. We start with a review of Sprecher's algorithm, where $\alpha_1 = 1$, $\alpha_p = \sum_{r=1}^{\infty} \gamma^{-(p-1)\beta(r)}$ for $p = 2 \ldots n$, $\beta(r) = (n^r - 1)/(n-1)$ and $a = [\gamma(\gamma-1)]^{-1}$ are defined as in section 2. Additionally, some new definitions are needed.

Definition 3.1. *Let $\sigma : \mathbb{R} \to \mathbb{R}$ be an arbitrary continuous function with $\sigma(x) \equiv 0$ when $x \leq 0$, and $\sigma(x) \equiv 1$ when $x \geq 1$. For $q \in \{0, \ldots, m\}$ and $k \in \mathbb{N}$ given, define*

$$d_{k,p}^q \ := \ d_{k,p} + q \sum_{r=2}^{k} \gamma^{-r}$$

and set $\mathbf{d}_k^q = (d_{k,1}^q, \ldots, d_{k,n}^q)$. Then for each number $\xi(\mathbf{d}_k^q) := \sum_{p=1}^n \alpha_p \, \psi(d_{k,p}^q)$ we set

$$b_k \ := \ \left(\sum_{r=k+1}^{\infty} \gamma^{-\beta(r)} \right) \left(\sum_{p=1}^{n} \alpha_p \right) \qquad \text{and}$$

$$\omega(\mathbf{d}_{k_r}^q; y_q) \ := \ \sigma\Big(\gamma^{\beta(k+1)} \big(y_q - \xi(\mathbf{d}_k^q)\big) + 1\Big) - \sigma\Big(\gamma^{\beta(k+1)} \big(y_q - \xi(\mathbf{d}_k^q) - (\gamma - 2)\, b_k\big)\Big) \,. \tag{3.1}$$

We are now in the position to present the algorithm of Sprecher which implements the representation of an arbitrary multivariate function f as superposition of single variable functions. Let $\| \cdot \|$ denote the usual maximum norm of functions and let $f : \mathbb{I}^n \to \mathbb{R}$ be a given continuous function with known uniform maximum norm $\| f \|$. Furthermore, let η and ε be fixed real numbers such that $0 < \frac{m-n+1}{m+1}\varepsilon + \frac{2n}{m+1} \leq \eta < 1$ which implies $\varepsilon < 1 - \frac{n}{m-n+1}$.

Algorithm 3.2. *Starting with $f_0 \equiv f$, for $r = 1, 2, 3, \ldots$, iterate the following steps:*

I. *Given the function $f_{r-1}(\mathbf{x})$, determine an integer k_r such that for any two points $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$ with $\| \mathbf{x} - \mathbf{x}' \| \leq \gamma^{-k_r}$ it holds that $|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{x}')| \leq \varepsilon \| f_{r-1} \|$. This determines rational coordinate points $\mathbf{d}_{k_r}^q = (d_{k_r,1}^q, \ldots, d_{k_r,n}^q)$.*

II. *For $q = 0, 1, \ldots, m$:*

II–1 Compute the values $\psi(d_{k_r,p}^q)$.

II–2 Compute the linear combinations $\xi(\mathbf{d}_{k_r}^q) = \sum_{p=1}^n \alpha_p \, \psi(d_{k_r,p}^q)$.

II–3 Compute the functions $\omega(\mathbf{d}_{k_r}^q; y_q)$.

III. *III–1 Compute for $q = 0, \ldots, m$ the functions*

$$\Phi_q^r(y_q) \ = \ \frac{1}{m+1} \sum f_{r-1}(\mathbf{d}_{k_r}) \, \omega(\mathbf{d}_{k_r}^q; y_q) \,, \tag{3.2}$$

where the sum is taken over all values $\mathbf{d}_{k_r}^q \in \mathcal{D}_{k_r}^n$.

III–2 Substitute for $q = 1, \ldots, m$ the transfer functions $\xi(\mathbf{x}_q)$ and compute the functions

$$\Phi_q^r \circ \xi(\mathbf{x}_q) \ := \ \frac{1}{m+1} \sum f_{r-1}(\mathbf{d}_{k_r}) \, \omega(\mathbf{d}_{k_r}^q; \xi(\mathbf{x}_q)) \,.$$

Again, the sum is built over all values $\mathbf{d}_{k_r}^q \in \mathcal{D}_{k_r}^n$.

III–3 Compute the function

$$f_r(\mathbf{x}) \ := \ f(\mathbf{x}) - \sum_{q=0}^{m} \sum_{j=1}^{r} \Phi_q^j \circ \xi(\mathbf{x}_q) \,. \tag{3.3}$$

This completes the r–th iteration loop and gives the r–th approximation to f. Now replace r by $r+1$ and go to step I.

The convergence of the series $\{f_r\}$ for $r \to \infty$ to the limit $\lim_{r\to\infty} f_r =: g \equiv 0$ is equivalent to the validity of theorem 2.1. The following convergence proof essentially follows [27,28]. It differs however in the arguments that refer to the inner function ψ which is now given by (2.8), i.e. we always refer to Köppen's definition (2.8), if we use the inner function ψ.

The main argument for convergence is the validity of the following theorem:

Theorem 3.3. *For the approximations f_r, $r = 0, 1, 2, \ldots$ defined in step* III–3 *of Algorithm 3.2 there holds the estimate*

$$\| f_r \| \; = \; \left\| f_{r-1}(\mathbf{x}) - \sum_{q=0}^{m} \Phi_q^r \circ \xi(\mathbf{x}_q) \right\| \le \eta \| f_{r-1} \| \, .$$

To prove this theorem, some preliminary work is necessary. To this end, note that a key to the numerical implementation of Algorithm 3.2 is the minimum distance of images of rational grid points $\mathbf{d}_k$ under the mapping ξ. We omit the superscript of $\mathbf{d}_k^q$ here for convenience, since $\mathbf{d}_k^q \in \mathcal{D}_k^n$ and the result holds for all $\mathbf{d}_k \in \mathcal{D}_k^n$. This distance can be bounded from below. The estimate is given in the following lemma.

Lemma 3.4. *For each integer $k \in \mathbb{N}$, set*

$$\mu_k \; := \; \sum_{p=1}^{n} \alpha_p \big[\psi(d_{k,p}) - \psi(d'_{k,p}) \big] \, , \tag{3.4}$$

where $d_{k,p}, d'_{k,p} \in \mathcal{D}_k$. Then

$$\min |\mu_k| \; \ge \; \gamma^{-n\beta(k)} \tag{3.5}$$

where the minimum is taken over all pairs $\mathbf{d}_{k,p}, \mathbf{d}'_{k,p} \in \mathcal{D}_k^n$ for which

$$\sum_{p=1}^{n} |d_{k,p} - d'_{k,p}| \; \neq \; 0 \, . \tag{3.6}$$

Proof. Since for each k the set $\mathcal{D}_k$ is finite, a unique minimum exists. For each $k \in \mathbb{N}$, let $d_{k,p}, d'_{k,p} \in \mathcal{D}_k$ and $A_{k,p} := \psi(d_{k,p}) - \psi(d'_{k,p})$ for $p = 1, \ldots, n$. Since ψ is monotone increasing, we know that $A_{k,p} \neq 0$ for all admissible values of p. Now from lemma 2.6 it follows directly that

$$\min_{\mathcal{D}_k} |A_{k,p}| \; = \; \gamma^{-\beta(k)} \, , \tag{3.7}$$

where for each fixed d, k the minimum is taken over the decimals for which $|d_{k,p} - d'_{k,p}| \neq 0$. The upper bound

$$\min |\mu_k| \; \le \; \alpha_n \, \gamma^{-\beta(k)} \tag{3.8}$$

can be gained from the definition of the μ_k and the fact that $1 = \alpha_1 > \alpha_2 > \ldots > \alpha_n$ as follows: Since $|\mu_k| \le \sum_{p=1}^{n} \alpha_p |A_{k,p}|$ we can see from (3.7) and (3.8) that a minimum of $|\mu_k|$ can only occur if $A_{k,T} \neq 0$ for some $T \in \{2, \ldots, n\}$.

Let us now denote the k–th remainder of α_p by

$$\varepsilon_{k,p} \; := \; \sum_{r=k+1}^{\infty} \gamma^{-(p-1)\beta(r)} \tag{3.9}$$

such that

$$\alpha_p - \varepsilon_{k,p} \;=\; \sum_{r=1}^{k} \gamma^{-(p-1)\beta(r)} \tag{3.10}$$

and consider the expression

$$A_{k,1} + \sum_{p=2}^{T} (\alpha_p - \varepsilon_{k,p}) A_{k,p}\,. \tag{3.11}$$

We claim the following:

$$\text{If } A_{k,T} \neq 0 \quad \text{then} \quad A_{k,1} + \sum_{p=2}^{T} (\alpha_p - \varepsilon_{k,p}) A_{k,p} \neq 0\,,$$

i.e. the term $(\alpha_T - \varepsilon_{k,T})A_{k,T}$ cannot be annihilated by the preceding terms in the sum. To show this, an application of (3.10) leads to

$$\alpha_T - \varepsilon_{k,T} \;=\; \gamma^{-(T-1)} + \gamma^{-(T-1)\beta(2)} + \ldots + \gamma^{-(T-1)\beta(k)}\,.$$

Also note that, for the choice $k = 1$ and $i_{1,T} = \gamma - 1$ as well as $i'_{1,T} = 0$ in (2.16), the largest possible term in the expansion of $|A_{k,T}|$ in powers of γ^{-1} is

$$\frac{\gamma - 1}{\gamma}\,.$$

Therefore, $(\alpha_T - \varepsilon_{k,T})|A_{k,T}|$ contains at least one term τ such that

$$0 \;<\; \tau \;\leq\; \gamma^{-(T-1)\beta(k)} \frac{\gamma - 1}{\gamma}\,.$$

But according to (3.7) and (3.10) the smallest possible term of $(\alpha_p - \varepsilon_{k,p})|A_{k,p}|$ for $p < T$ is

$$\gamma^{-(T-2)\beta(k)} \gamma^{-\beta(k)} \;=\; \gamma^{-(T-1)\beta(k)}$$

so that the assertion holds and (3.11) indeed does not vanish.

If $|i_{k,T} - i'_{k,T}| = 1$, we have without loss of generality in the representation (2.16) the values

i'_k	i_k	$\tilde{s}'_1$	s'_1	θ'_0	θ'^{+}_0	$\tilde{s}_1$	s_1	θ_0	θ^{+}_0
$\gamma - 2$	$\gamma - 1$	0	$\frac{1}{2}$	1	0	$\frac{1}{2}$	1	1	0
$\gamma - 3$	$\gamma - 2$	0	0	1	0	0	$\frac{1}{2}$	1	0
$\gamma - 4$	$\gamma - 3$	0	0	1	0	0	0	1	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

and we can directly infer that the expansion of (3.11) in powers of γ^{-1} contains the term

$$\gamma^{-(T-1)\beta(k)} \gamma^{-\beta(k)} = \gamma^{-T\beta(k)}\,. \tag{3.12}$$

We now show that this is the smallest term in the sum (3.11). To this end, we use the representation (2.16) for $A_{k,p}$ and factor out $\gamma^{-\beta(k-j)}$ for each j. Since θ_j and θ_j^{+} become smaller than 2^{-j}, we can bound each term in the sum (3.11) from below by $\gamma^{-\beta(k-j)}2^{-j}$. The further estimation $\gamma^{-\beta(k-j)}2^{-j} > \gamma^{-\beta(k)}$ shows that (3.12) is indeed the smallest term in the sum and hence cannot be annihilated by other terms in (3.11). Therefore,

$$\left| A_{k,1} + \sum_{p=2}^{T} (\alpha_p - \varepsilon_{k,p}) A_{k,p} \right| \;\geq\; \gamma^{-T\beta(k)} .$$

But this implies that also

$$\left| A_{k,1} + \sum_{p=2}^{T} \alpha_p A_{k,p} \right| \;\geq\; \gamma^{-T\beta(k)}$$

since all possible terms in the expansion of $\sum_{p=2}^{T} \varepsilon_{k,p} A_{k,p}$ in powers of γ^{-1} are too small to annihilate $\gamma^{-T\beta(k)}$. Thus, choosing $T = n$ the lemma is proven. □

The linear combinations $\xi(\mathbf{d}_k^q)$ of the inner functions serve for each $q = 0, \ldots, m$ as a mapping from the hypercube $\mathbb{I}^n$ to $\mathbb{R}$. Therefore, further knowledge on the structure of this mapping is necessary. To this end, we need the following lemma:

Lemma 3.5. *For each integer $k \in \mathbb{N}$, let*

$$\delta_k \;:=\; \frac{\gamma - 2}{(\gamma - 1)\,\gamma^k} . \tag{3.13}$$

Then for all $d_k \in \mathcal{D}_k$ and $\varepsilon_{k,2}$ as given in (3.9) *we have*

$$\psi(d_k + \delta_k) \;=\; \psi(d_k) + (\gamma - 2)\,\varepsilon_{k,2} .$$

Proof. The proof relies mainly on the continuity of ψ and some direct calculations. If we express δ_k as an infinite sum we have

$$d_k + \delta_k \;=\; \lim_{k_0 \to \infty} \left\{ d_k + \sum_{r=1}^{k_0} \frac{\gamma - 2}{\gamma^{k+r}} \right\} \;=:\; \lim_{k_0 \to \infty} d_{k_0} .$$

Since ψ is continuous we get

$$\psi\left(\lim_{k_0 \to \infty} d_{k_0} \right) \;=\; \lim_{k_0 \to \infty} \psi(d_{k_0})$$

and since $i_{k+r} = \gamma - 2$ for $r = 1, \ldots k_0$, it follows directly that $\tilde{s}_r = 0$ for $j = 0, \ldots, k_0 - k$. Therefore $\theta_j^+ = 0$ and $\theta_j = 1$ for $j = 0, \ldots, k_0 - k$. With the representation (2.15) and the choice $\xi = k_0 - k$, the assertion follows. □

As a direct consequence of this lemma, we have the following corollary, in which the one-dimensional case is treated.

Corollary 3.6. *For each integer $k \in \mathbb{N}$ and $d_k \in \mathcal{D}_k$, the pairwise disjoint intervals*

$$E_k(d_k) \;:=\; [d_k, d_k + \delta_k] \tag{3.14}$$

are mapped by ψ into the pairwise disjoint image intervals

$$H_k(d_k) \;:=\; [\psi(d_k), \psi(d_k) + (\gamma - 2)\,\varepsilon_{k,2}] . \tag{3.15}$$

Proof. From their definition it follows directly that the intervals $E_k(d_k)$ are pairwise disjoint. The corollary then follows from lemma 3.4 and lemma 3.5. $\square$

We now generalize this result to the multidimensional case.

Lemma 3.7. *For each fixed integer $k \in \mathbb{N}$ and $\mathbf{d}_k \in \mathcal{D}_k^n$, the pairwise disjoint cubes*

$$S_k(\mathbf{d}_k) \;:=\; \prod_{p=1}^{n} E_k(d_{k,p}) \tag{3.16}$$

in $\mathbb{I}^n$ are mapped by $\sum_{p=1}^{n} \alpha_p\, \psi(d_{k,p})$ into the pairwise disjoint intervals

$$T_k(\mathbf{d}_k) \;:=\; \left[\; \sum_{p=1}^{n} \alpha_p\, \psi(d_{k,p})\;,\; \sum_{p=1}^{n} \alpha_p\, \psi(d_{k,p}) + \left(\sum_{p=1}^{n} \alpha_p\right)(\gamma-2)\,\varepsilon_{k,2} \right]. \tag{3.17}$$

Proof. This lemma is a consequence of the previous results and can be found in detail in [27]. $\square$

We now consider Algorithm 3.2 again. We need one more ingredient:

Lemma 3.8. *For each value of q and r, there holds the following estimate:*

$$\|\, \Phi_q^r(y_q)\,\| \;\le\; \frac{1}{m+1} \|\, f_{r-1}\,\|\,.$$

Proof. The support of each function $\omega(\mathbf{d}_k^q; y_q)$ is the open interval

$$U_k^q(\mathbf{d}_k^q) \;:=\; \Big(\; \xi(\mathbf{d}_k^q) - \gamma^{-\beta(k+1)}\;,\; \xi(\mathbf{d}_k^q) + (\gamma-2)b_k + \gamma^{-\beta(k+1)} \;\Big)\,.$$

Then, by lemma 3.7 the following holds: If $\xi(\mathbf{d}_k^q) \neq \xi(\mathbf{d}_k^{'q})$ then $U_k^q(\mathbf{d}_k^q) \cap U_k^q(\mathbf{d}_k^{'q}) = \emptyset$. With this property and the fact that $0 \le \omega(\mathbf{d}_k^q; y_q) \le 1$, we derive from (3.2):

$$\left\| \frac{1}{m+1} \sum f_{r-1}(\mathbf{d}_{k_r}^q)\, \omega(\mathbf{d}_{k_r}^q; y_q) \right\| \;=\; \frac{1}{m+1} \max |f_{r-1}(\mathbf{d}_{k_r})|\,.$$

Here, the sum is taken over all values $\mathbf{d}_{k_r}^q \in \mathcal{D}_{k_r}^n$ and the maximum over all $\mathbf{d}_{k_r} \in \mathcal{D}_{k_r}^n$. The lemma then follows from the definition of the maximum norm, see also [28], lemma 1. $\square$

We are now ready to prove theorem 3.3, compare also [28].

Proof of theorem 3.3. For simplicity, we include the value $d_k = 1$ in the definition of the rational numbers $\mathcal{D}_k$. Consider now for each integer q and $a = [\gamma(\gamma-1)]^{-1}$ as in theorem 2.1 the family of closed intervals

$$E_k^q(d_k^q) \;:=\; \left[\; d_k^q - q\,a\;,\; d_k^q - q\,a + \delta_k \;\right]. \tag{3.18}$$

With $\delta_k = (\gamma-2)(\gamma-1)^{-1}\gamma^{-k}$ we can see that

$$E_k^q(d_k^q) \;=\; \left[\; d_k - \frac{q}{\gamma-1}\gamma^{-k}\;,\; d_k - \frac{q}{\gamma-1}\gamma^{-k} + \frac{\gamma-2}{\gamma-1}\gamma^{-k} \;\right]$$

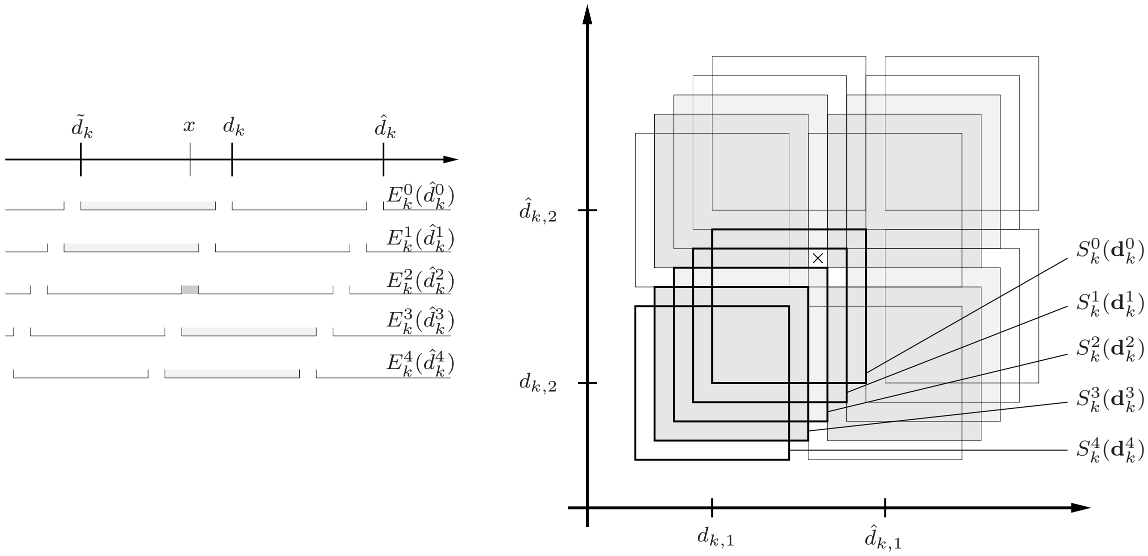

Figure 3: Let k be a fixed integer, $m = 4$, $\gamma = 10$ and $\tilde{d}_{k,i} := d_{k,i} - \gamma^{-k}$, $\hat{d}_{k,i} := d_{k,i} + \gamma^{-k}$, $i \in \{1, 2\}$. The left figure depicts the intervals $E_k^q(d_k^q)$ for $q = 1, \ldots, m$. The subscript i indicating the coordinate direction is omitted for this one–dimensional case. The point x is contained in the intervals $E_k^0(\tilde{d}_k^0)$, $E_k^1(\tilde{d}_k^1)$, $E_k^3(d_k^3)$, $E_k^4(d_k^4)$ (shaded) and in the gap $G_k^2(\tilde{d}_k^2)$ (dark shaded). The figure on the right shows the cubes $S_k^q(\mathbf{d}_k^q)$ for $n = 2$, $q = 1, \ldots, m$ and different values $\mathbf{d}_k \in \mathcal{D}_k^n$. For $q \in \{2, 3\}$, the marked point is not contained in any of the cubes from the set $\left\{S_k^q(\mathbf{d}_k^q) \,:\, \mathbf{d}_k \in \mathcal{D}_k^n\right\}$.

and that these intervals are separated by gaps $G_k^q(d_k^q) := \left(d_k^q - q\,a + \delta_k\,,\, d_k^q - q\,a + \gamma^{-k}\right)$ of width $(\gamma-1)^{-1}\gamma^{-k}$, compare figure 3. With the intervals E_k^q we obtain for each k and $q = 0, \ldots, m$ the closed (Cartesian product) cubes

$$S_k^q(\mathbf{d}_k^q) \;:=\; E_k^q(d_{k,1}^q) \times \ldots \times E_k^q(d_{k,n}^q)\,,$$

whose images under $\xi(\mathbf{x}_q) = \sum_{p=1}^n \alpha_p\, \psi(x_p + qa)$ are the disjoint closed intervals

$$T_k^q(\mathbf{d}_k^q) \;=\; \left[\, \xi(\mathbf{d}_k^q)\;,\; \xi(\mathbf{d}_k^q) + (\gamma - 2)\, b_k \,\right]\,,$$

as derived in lemma 3.7. For the two–dimensional case, the cubes $S_k^q(\mathbf{d}_k^q)$ are depicted in figure 3.

Now let k be fixed. The mapping $\xi(\mathbf{x}_q)$ associates to each cube $S_k^q(\mathbf{d}_k^q)$ from the coordinate space a unique image $T_k^q(\mathbf{d}_k^q)$ on the real line. For fixed q the images of any two cubes from the set $\left\{S_k^q(\mathbf{d}_k^q) \,:\, \mathbf{d}_k \in \mathcal{D}_k^n\right\}$ have empty intersections. This allows a local approximation of the target function $f(\mathbf{x})$ on these images $T_k^q(\mathbf{d}_k^q)$ for $\mathbf{x} \in S_k^q(\mathbf{d}_k^q)$. However, as the outer functions Φ_q^r have to be continuous, these images have to be separated by gaps in which $f(\mathbf{x})$ cannot be approximated. Thus an error is introduced that cannot be made arbitrarily small. This deficiency is eliminated by the affine translations of the cubes $S_k^q(\mathbf{d}_k^q)$ through the variation of the q's. To explain this in more detail, let $x \in [0, 1]$ be an arbitrary point. With (3.18) we see that the gaps $G_k^q(d_k^q)$ which separate the intervals do not intersect for variable q. Therefore, there exists only one value q_* such that $x \in G_k^{q_*}(d_k^{q_*})$. This implies that for the remaining m values of q there holds $x \in E_k^q(d_k^q)$ for some d_k. See figure 3 (left) for an illustration of this fact. If we now consider an arbitrary point $\mathbf{x} \in [0, 1]^n$, we see that there exist at least $m - n + 1$ different values q_j, $j = 1, \ldots, m - n + 1$ for which $\mathbf{x} \in S_k^{q_j}(\mathbf{d}_k^{q_j})$ for some $\mathbf{d}_k$, see figure 3 (right). Note that the points $\mathbf{d}_k$ can differ for different values q_j. From (3.18) we see that $\mathbf{d}_k \in S_k^{q_j}(\mathbf{d}_k^{q_j})$.

Now we consider step I of Algorithm 3.2. To this end, remember that η and ε are fixed numbers such that $0 < \frac{m-n+1}{m+1}\varepsilon + \frac{2n}{m+1} \leq \eta < 1$. Let k_r be the integer given in step I with the associated assumption that $|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{x}')| \leq \varepsilon \| f_{r-1} \|$ when $|x_p - x'_p| \leq \gamma^{-k_r}$ for $p = 1, \ldots, n$. Let $\mathbf{x} \in [0,1]^n$ be an arbitrary point and let q_j, $j = 1, \ldots, m-n+1$, denote the values of q such that $\mathbf{x} \in S^{q_j}_{k_r}(\mathbf{d}^{q_j}_{k_r})$. For the point $\mathbf{d}_{k_r} \in S^{q_j}_{k_r}(\mathbf{d}^{q_j}_{k_r})$ we have

$$|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{d}_{k_r})| \;\leq\; \varepsilon \| f_{r-1} \| \tag{3.19}$$

and for $\mathbf{x}$ it holds that $\xi(\mathbf{x}_{q_j}) \in T^{q_j}_{k_r}(\mathbf{d}^{q_j}_{k_r})$. The support $U^{q_j}_{k_r}(\mathbf{d}^{q_j}_{k_r})$ of the function $\omega(d^{q_j}_{k_r}; y_{q_j})$ contains the interval $T^{q_j}_{k_r}(\mathbf{d}^{q_j}_{k_r})$. Furthermore, from definition (3.1) we see that ω is constant on that interval. With (3.3) we then get

$$\begin{aligned} \Phi^r_{q_j} \circ \xi(\mathbf{x}_{q_j}) \;&=\; \frac{1}{m+1} \sum_{\mathbf{d}^{q_j}_{k_r}} f_{r-1}(\mathbf{d}_{k_r})\, \omega(\mathbf{d}^{q_j}_{k_r}; \xi(\mathbf{x}_{q_j})) \\ &=\; \frac{1}{m+1} f_{r-1}(\mathbf{d}_{k_r})\,. \end{aligned} \tag{3.20}$$

Together with (3.19) this shows

$$\left| \frac{1}{m+1} f_{r-1}(\mathbf{x}) - \Phi^r_{q_j} \circ \xi(\mathbf{x}_{q_j}) \right| \;\leq\; \frac{\varepsilon}{m+1} \| f_{r-1} \| \tag{3.21}$$

for all q_j, $j = 1, \ldots, m-n+1$. Note that this estimate does not hold for the remaining values of q for which $\mathbf{x}$ is not contained in the cube $S^q_{k_r}(\mathbf{d}^{q_j}_{k_r})$. Let us now denote these values by $\bar{q}_j$, $j = 1, \ldots, n$. We can apply lemma 3.8 and with the special choice of the values ε and η we obtain the estimate

$$\begin{aligned} |f_r(\mathbf{x})| \;&=\; \left| f_{r-1}(\mathbf{x}) - \sum_{q=0}^{m} \Phi^r_q \circ \xi(\mathbf{x}_q) \right| \\ &=\; \left| \sum_{q=0}^{m} \frac{1}{m+1} f_{r-1}(\mathbf{x}) - \sum_{j=1}^{m-n+1} \Phi^r_{q_j} \circ \xi(\mathbf{x}_{q_j}) - \sum_{j=1}^{n} \Phi^r_{\bar{q}_j} \circ \xi(\mathbf{x}_{\bar{q}_j}) \right| \\ &\leq\; \left| \frac{n}{m+1} f_{r-1}(\mathbf{x}) + \sum_{j=1}^{m-n+1} \frac{1}{m+1} f_{r-1}(\mathbf{x}) - \Phi^r_{q_j} \circ \xi(\mathbf{x}_{q_j}) \right| + \frac{n}{m+1} \| f_{r-1} \| \\ &\leq\; \left[\frac{m-n+1}{m+1} \varepsilon + \frac{2n}{m+1} \right] \| f_{r-1} \| \;\leq\; \eta \, \| f_{r-1} \|\,. \end{aligned} \tag{3.22}$$

This completes the proof of theorem 3.3. $\square$

We now state a fact that follows directly from the previous results.

Corollary 3.9. *For $j = 1, 2, 3, \ldots$ there hold the following estimates:*

$$\| \Phi^r_q(y_q) \| \;\leq\; \frac{1}{m+1} \eta^{r-1} \| f \| \tag{3.23}$$

and

$$\| f_r \| \;=\; \left\| f(\mathbf{x}) - \sum_{q=0}^{m} \sum_{j=1}^{r} \Phi^j_q \circ \xi(\mathbf{x}_q) \right\| \;\leq\; \eta^r \, \| f \|\,. \tag{3.24}$$

Proof. Remember that $f_0 \equiv f$. The first estimate follows from lemma 3.8 and a recursive application of theorem 3.3. The second estimate can be derived from the definition (3.3) of f_r and again a recursive application of theorem 3.3. $\square$

We finally are in the position to prove theorem 2.1.

Proof of theorem 2.1. From corollary 3.9 and the fact that $\eta < 1$ it follows that, for all $q = 1, \ldots, m$, we have

$$\left\| \sum_{j=1}^{r} \Phi_q^j(y_q) \right\| \leq \sum_{j=1}^{r} \| \Phi_q^j(y_q) \| \leq \frac{1}{m+1} \| f \| \sum_{j=0}^{r-1} \eta^j < \frac{1}{m+1} \| f \| \sum_{j=0}^{\infty} \eta^j < \infty \,. \tag{3.25}$$

The functions $\Phi_q^j(y_q)$ are continuous and therefore each series $\sum_{j=1}^{r} \Phi_q^j(y_q)$ converges absolutely to a continuous function $\Phi_q(y_q)$ as $r \to \infty$. Since $\eta < 1$ we see from the second estimate in corollary 3.9 that $f_r \to 0$ for $r \to \infty$. This proves Sprecher's version of Kolmogorov's superposition theorem with Köppen's inner function ψ. □

4 Conclusion and outlook

In this paper we filled mathematical gaps in the articles of Köppen [16] and Sprecher [27,28] on Kolmogorov's superposition theorem. We first showed that Sprecher's original inner function ψ is not continuous and monotone increasing. Thus the convergence proof of the algorithm from [28] that implements (2.1) constructively is incomplete. We therefore considered a corrected version of ψ as suggested in [16]. We showed that this function is well defined, continuous and monotone increasing. Then, we carried the approach for a constructive proof of Kolmogorov's superposition theorem from [27, 28] over to the new continuous and monotone ψ and showed convergence. Altogether we gave a mathematically correct, constructive proof of Kolmogorov's superposition theorem.

The present result is, to our knowledge, the first correct constructive proof of (2.1) and thus of (1.1). It however still involves (with $r \to \infty$) an in general infinite number of iterations. Thus, any finite numerical application of algorithm 3.2 can only give an *approximation* of a n–dimensional function up to an arbitrary accuracy $\tilde{\epsilon} > 0$ (compare corollary 3.9). While the number of iterations in algorithm 3.2 to achieve this desired accuracy is independent of the function f and its smoothness, the number k_r which is determined in step I can become very large for oscillating functions. This reflects the dependency of the costs of algorithm 3.2 on the smoothness of the function f: In step II the functions $\omega(\mathbf{d}_{k_r}^q, y_q)$ are computed for all rational values $\mathbf{d}_{k_r}^q$ which can be interpreted as a construction of basis functions on a regular grid in the unit cube $[0, 1]^n$. Since the number of grid-points in a regular grid increases exponentially with the dimensionality n, the overall costs of the algorithm increase at least with the same rate for $n \to \infty$. This makes algorithm 3.2 highly inefficient in higher dimensions. To overcome this problem and thus to benefit numerically from the constructive nature of the proof further approximations to the outer functions in (2.1) have to be made. This will be discussed in a forthcoming paper.

References

[1] V. Arnold. On functions of three variables. *Dokl. Akad. Nauk SSSR*, 114:679 – 681, 1957. English translation: American Math. Soc. Transl (2), 28, 51 - 54, 1963.

[2] V. Arnold. On the representation of functions of several variables by superpositions of functions of fewer variables. *Mat. Prosveshchenie*, 3:41 – 61, 1958.

[3] V. Arnold. On the representation of continuous functions of three variables by superpositions of continuous functions of two variables. *Mat. Sb.*, 48:3 – 74, 1959. English translation: American Math. Soc. Transl. (2), 28, 61 - 147, 1963.

[4] R. J. P. de Figueiredo. Implications and applications of Kolmogorov's superposition theorem. *IEEE Transactions on Automatic Control*, AC-25(6), 1980.

[5] B. Fridman. An improvement on the smoothness of the functions in Kolmogorov's theorem on superpositions. *Dokl. Akad. Nauk SSSR*, 177:1019–1022, 1967. English translation: Soviet Math. Dokl. (8), 1550-1553, 1967.

[6] F. Girosi and T. Poggio. Representation properties of networks: Kolmogorov's theorem is irrelevant. *Neural Comp.*, 1:465 – 469, 1989.

[7] R. Hecht-Nielsen. Counter propagation networks. *Proceedings of the International Conference on Neural Networks II*, pages 19 – 32, 1987.

[8] R. Hecht-Nielsen. Kolmogorov's mapping neural network existence theorem. *Proceedings of the International Conference on Neural Networks III*, pages 11 – 14, 1987.

[9] T. Hedberg. The Kolmogorov superposition theorem, Appendix II to H. S. Shapiro, Topics in Approximation Theory. *Lecture Notes in Math.*, 187:267 – 275, 1971.

[10] D. Hilbert. Mathematical problems. *Bull. Amer. Math. Soc.*, 8:461 – 462, 1902.

[11] W. Hurewicz and H. Wallman. *Dimension Theory.* Princeton University Press, Princeton, NJ, 1948.

[12] B. Igelnik and N. Parikh. Kolmogorov's spline network. *IEEE transactions on Neural Networks*, 14:725 – 733, 2003.

[13] S. Khavinson. Best approximation by linear superpositions. *Translations of Mathematical Monographs*, 159, 1997. AMS.

[14] A. N. Kolmogorov. On the representation of continuous functions of many variables by superpositions of continuous functions of one variable and addition. *Doklay Akademii Nauk USSR*, 14(5):953 – 956, 1957.

[15] A. N. Kolmogorov and V. M. Tikhomirov. ϵ–entropy and ϵ–capacity of sets in function spaces. *Uspekhi Mat. Nauk*, 13(2):3 – 86, 1959. English translation: American Math. Soc. Transl. 17, (2), 277 - 364, 1961.

[16] M. Köppen. On the training of a Kolmogorov network. *ICANN 2002, Lecture Notes In Computer Science*, 2415:474–479, 2002.

[17] V. Kurkova. Kolmogorov's theorem is relevant. *Neural Computation*, 3:617 – 622, 1991.

[18] V. Kurkova. Kolmogorov's theorem and multilayer neural networks. *Neural Networks*, 5:501 – 506, 1992.

[19] G. Lorentz. Approximation of functions. *Holt, Rinehart & Winston*, 1966.

[20] G. Lorentz, M. Golitschek, and Y. Makovoz. *Constructive Approximation.* 1996.

[21] M. Nakamura, R. Mines, and V. Kreinovich. Guaranteed intervals for Kolmogorov's theorem (and their possible relation to neural networks). *Interval Comput.*, 3:183–199, 1993.

[22] M. Nees. Approximative versions of Kolmogorov's superposition theorem, proved constructively. *Journal of Computational and Applied Mathematics*, 54:239 – 250, 1994.

[23] P. A. Ostrand. Dimension of metric spaces and Hilbert's problem 13. *Bull. Amer. Math. Soc.*, 71:619 – 622, 1965.

[24] T. Rassias and J. Simsa. Finite sum decompositions in mathematical analysis. *Pure and applied mathematics*, 1995.

[25] D. A. Sprecher. On the structure of continuous functions of several variables. *Transactions Amer. Math. Soc*, 115(3):340 – 355, 1965.

[26] D. A. Sprecher. An improvement in the superposition theorem of Kolmogorov. *Journal of Mathematical Analysis and Applications*, 38:208 – 213, 1972.

[27] D. A. Sprecher. A numerical implementation of Kolmogorov's superpositions. *Neural Networks*, 9(5):765 – 772, 1996.

[28] D. A. Sprecher. A numerical implementation of Kolmogorov's superpositions II. *Neural Networks*, 10(3):447 – 457, 1997.