A Computable Kolmogorov Superposition Theorem

Vasco Brattka
Theoretische Informatik I, FernUniversität Hagen
D-58084 Hagen, Germany
e-mail: vasco.brattka@fernuni-hagen.de

Abstract

In the year 1900 in his famous lecture in Paris Hilbert formulated 23 challenging problems which inspired many ground breaking mathematical investigations in the last century. Among these problems the 13th was concerned with the solution of higher order algebraic equations. Hilbert conjectured that such equations are not solvable by functions which can be constructed as substitution of continuous functions of only two variables. Surprisingly, 57 years later Hilbert's conjecture was refuted when Kolmogorov succeeded to prove his Superposition Theorem which states that each multivariate continuous real-valued function can be represented as superposition and composition of continuous functions of only one variable. Again 30 years later this theorem got an interesting application in the theory of neural networks. We will prove a computable version of Kolmogorov's Superposition Theorem in the rigorous framework of computable analysis: each multivariate computable real-valued function can be represented as superposition and composition of computable real number functions of only one variable and such a representation can even be determined effectively. As a consequence one obtains a characterization of the computational power of feedforward neural networks with computable activation functions and weights.

Keywords: computable analysis, superpositions, neural networks.

1 Introduction

In his famous lecture in Paris 1900 Hilbert introduced 23 prominent problems as a challenge for the coming generations of mathematicians [Hil00]. The 13th problem is concerned with algebraic equations

$$a_nx^n + a_{n-1}x^{n-1} + ... + a_1x + a_0 = 0.$$

In 1900 it was well-known that equations of degree $n \leq 4$ are solvable by functions which can be represented as substitutions of purely algebraic operations but as a consequence of Galois' famous theory it was known that this is not possible for equations of degree $n \geq 5$ in general. The algebraic operations, i.e. addition, multiplication, subtraction, division and roots are operations of at most two arguments. Thus, Hilbert was led to the conjecture that algebraic equations of higher degree can in general not even be solved by functions which can be represented as substitutions of continuous functions of only two variables (cf. also [Hil27]). Especially, he conjectured that this is not possible for the equation

$$x^7 + ax^3 + bx^2 + cx + 1 = 0.$$

Surprisingly, his conjecture was refuted when Kolmogorov proved in a series of papers (in a competition with Arnol'd) his Superposition Theorem [Kol57]. Roughly speaking, the theorem states that each multivariate continuous real-valued function can be represented as a superposition and composition of continuous functions of only one variable. More precisely:

Theorem 1.1 (Kolmogorov's Superposition Theorem) *For each $n \geq 2$ there exist continuous functions $\varphi_q : [0,1] \to \mathbb{R}$, $q = 0, ..., 2n$ and constants $\lambda_p \in \mathbb{R}$, $p = 1, ..., n$ such that the following holds true: for each continuous function $f : [0,1]^n \to \mathbb{R}$ there exists a continuous function $g : [0,1] \to \mathbb{R}$ such that*

$$f(x_1, ..., x_n) = \sum_{q=0}^{2n} g\left(\sum_{p=1}^{n} \lambda_p \varphi_q(x_p)\right).$$

It is worth noticing that the functions φ_q and the constants λ_p are independent from the represented function f and do only depend on the dimension n. Actually, the given formulation is based on improved versions of the theorem which have been published by Lorentz [Lor66] and Sprecher [Spr65]. The functions φ_q can even be guaranteed to have a certain Lipschitz degree. But later on it has been noticed that the theorem does not hold true for several other classes of functions: if we substitute "continuously differentiable" or "analytic" for "continuous" in the formulation of the theorem, then the statement is not correct any longer (cf. [Lor76]).

It is natural to ask whether the theorem is still correct if we replace "continuous" by "computable". Actually, the question whether Kolmogorov's Superposition Theorem can be proved constructively, has been already discussed in the literature. Especially, Nakamura, Mines and Kreinovich have investigated the question based on a (rather informal) computability notion [NMK93]. We will revisit the question from the rigorous point of view of computable analysis [Wei00, PER89] and using the recent work of Sprecher [Spr96] (cf. also [Spr93, KS94, Spr97]). In Section 3 we formulate the precise version of the following main result of this paper:

Theorem 1.2 (Computable Kolmogorov Superposition Theorem) *For each $n \geq 2$ there exist computable functions $\varphi_q : [0,1] \to \mathbb{R}$, $q = 0, ..., 2n$ and computable constants $\lambda_p \in \mathbb{R}$, $p = 1, ..., n$ such that the following holds true: for each continuous function $f : [0,1]^n \to \mathbb{R}$ there exists a continuous function $g : [0,1] \to \mathbb{R}$ such that*

$$f(x_1, ..., x_n) = \sum_{q=0}^{2n} g\left(\sum_{p=1}^{n} \lambda_p \varphi_q(x_p)\right).$$

Given a continuous f, we can even effectively find such a continuous g and if f is computable, then this procedure leads to a computable g.

One should notice that we have not replaced each occurrence of "continuous" by "computable". The functions φ_q can be chosen to be computable even in the classical version of the theorem. Moreover, the effective procedure which takes f as input and produces g as output works even for an arbitrary continuous function f as input.

In 1987, exactly 30 years after Kolmogorov's Superposition Theorem has been published, Hecht-Nielsen noticed an interesting application of this theorem to the theory of neural networks [HN87, HN90]: each continuous function $f : [0,1]^n \to \mathbb{R}$ can be implemented by a feedforward neural network with continuous activation functions $t : [0,1] \to \mathbb{R}$. Similarly, using the Computable Kolmogorov Superposition Theorem we can precisely characterize the computational power of computable feedforward neural networks:

Theorem 1.3 (Computable Feedforward Neural Networks) *The class of functions $f : [0,1]^n \to \mathbb{R}$, realizable by feedforward neural networks with computable activation functions $t : [0,1] \to \mathbb{R}$ and computable weights $\lambda \in \mathbb{R}$, is exactly the class of computable functions $f : [0,1]^n \to \mathbb{R}$.*

Besides those results which capture the computational power of neural networks precisely, also approximative solutions have been studied (cf. [Kur91, Kur92]).

We close this section with a survey on the organization of this paper: in Section 2 we will introduce some basic notions from computable analysis and we will show that Sprecher's function $\varphi : [0,1] \to \mathbb{R}$, which has been used to prove Kolmogorov's Superposition Theorem [Spr96], is computable. In Section 3 we will effectivize the construction of the "outer function" g, following the proof idea of Lorentz [Lor66]. In Section 4 we discuss a superposition representation of multivariate continuous functions.

2 Computability of Sprecher's Function

In this section we will briefly recall the notion of a computable real-valued function as it is used in computable analysis and we will show that Sprecher's function φ is computable in this sense. Precise definitions can be found in Weihrauch [Wei00]; other equivalent approaches have been presented by Pour-El and Richards [PER89] and Ko [Ko91].

Roughly speaking, a partial real-valued function $f :\subseteq \mathbb{R}^n \to \mathbb{R}$ is *computable*, if there exists a Turing machine M which in the long run transforms each infinite sequence $p \in \Sigma^\omega$ representing some input $x \in \mathbb{R}^n$ into an infinite sequence $r \in \Sigma^\omega$ representing the corresponding output $f(x)$.

It is easy to see that the given definition sensitively relies on an appropriate choice of the representation. Here, we will use the so-called *Cauchy representation* $\rho :\subseteq \Sigma^\omega \to \mathbb{R}$ of the real numbers, where roughly speaking, $\rho(p) = x$ if p is a sequence of rational numbers $(q_i)_{i\in\mathbb{N}}$ (encoded over the alphabet Σ) which rapidly converges to x, i.e. $|q_k - q_i| \leq 2^{-k}$ for all $i \geq k$. A corresponding representation ρ^n for the n-dimensional Euclidean space $\mathbb{R}^n$ can be derived easily.

More generally, a *representation* of a set X is a surjective mapping $\delta :\subseteq \Sigma^\omega \to X$. Using this notion we can define the computability concept of computable analysis precisely.

Definition 2.1 (Computable functions) Let δ, δ' be representations of X, Y, respectively. A partial function $f :\subseteq X \to Y$ is called (δ, δ')–*computable*, if there exists a Turing machine M (with one-way output tape) which computes a function $f_M :\subseteq \Sigma^\omega \to \Sigma^\omega$ such that

$$\delta' F_M(p) = f\delta(p)$$

for all $p \in \mathrm{dom}(f\delta)$.

The definition can be generalized to functions of higher arity straightforwardly. The situation of the previous definition can be visualized by a commutative diagram:

$$\begin{tikzcd}
\Sigma^\omega \arrow[r, "F_M"] \arrow[d, "\delta"'] & \Sigma^\omega \arrow[d, "\delta'"] \\
X \arrow[r, "f"'] & Y
\end{tikzcd}$$

Figure 1: Computable functions

We can generalize this computability concept even to (partial) multi-valued operations $f :\subseteq X \rightrightarrows Y$, which are operations where the image $f(x)$ is a (not necessarily single-valued) subset of Y. In this case we write $\delta' F_M(p) \in f\delta(p)$, i.e. "$\in$" instead of "$=$" in the definition above, in order to define computability.

We will call the (ρ^n, ρ)–computable functions $f :\subseteq \mathbb{R}^n \to \mathbb{R}$ *computable* for short. Computability can analogously be defined for other types of functions, as total functions $f : [a, b] \to \mathbb{R}$ and sequences $f : \mathbb{N} \to \mathbb{R}$, using representations $\rho|^{[a,b]}$ of $[a, b]$ and $\delta_\mathbb{N}$ of $\mathbb{N}$. There exist a well-known characterization of computable functions $f : [0, 1] \to \mathbb{R}$ which we will use to prove the computability of Sprecher's function φ. Roughly speaking, a function $f : [0, 1] \to \mathbb{R}$ is computable, if and only if $f|_D$ restricted to some "effectively dense" subset $D \subseteq [0, 1]$ is computable and f admits some computable modulus of continuity m, which is a function $m : \mathbb{N} \to \mathbb{N}$ such that

$$|x - y| < 2^{-m(n)} \Longrightarrow |f(x) - f(y)| < 2^{-n}$$

holds for all $x, y \in [0, 1]$ and $n \in \mathbb{N}$. As a dense subset we will use in the following the set $\mathbb{Q}_\gamma$ of rational numbers between 0 and 1 in expansion w.r.t. base γ. This set can be defined by $\mathbb{Q}_\gamma := \bigcup_{k=1}^{\infty} \mathbb{Q}_{\gamma k}$ and

$$\mathbb{Q}_{\gamma k} := \left\{ \sum_{r=1}^{k} i_r \gamma^{-r} : i_1, ..., i_k \in \{0, 1, ..., \gamma - 1\} \right\}$$

for all integers $k \geq 1$ and $\gamma \geq 2$. To make the characterization of computable functions $f : [0,1] \to \mathbb{R}$ more precise we assume that $\nu : \mathbb{N} \to \mathbb{Q}_\gamma$ is some effective standard numbering of the rational numbers in γ-expansion (cf. [Wei00]).

Proposition 2.2 (Characterization of computable functions) *Let $\gamma \geq 2$. A function $f : [0,1] \to \mathbb{R}$ is computable, if and only if the following holds:*

(1) $f \circ \nu : \mathbb{N} \to \mathbb{R}$ is a computable sequence of real numbers,

(2) f admits a computable modulus of continuity $m : \mathbb{N} \to \mathbb{N}$.

For a proof cf. [Wei00, Bra99, Ko91, PER89]. The main purpose of the remaining part of this section is to prove that there exist computable functions $\varphi_q : [0,1] \to \mathbb{R}$ which can be used for the proof of the computable version of Kolmogorov's Superposition Theorem. In the original proofs the existence of such functions φ_q has been proved without any explicit construction. Recently, Sprecher defined a concrete function $\varphi : [0,1] \to \mathbb{R}$ [Spr96, Spr97] which can be used to construct the functions φ_q. We will prove that φ is a computable function. More precisely, for each dimension $n \geq 2$ and each base $\gamma \geq 2n+2$ Sprecher has defined a separate function $\varphi : [0,1] \to \mathbb{R}$ as follows.

Definition 2.3 (Sprecher's function) Let $n \geq 2$ and $\gamma \geq 2n+2$. Let $\varphi : [0,1] \to \mathbb{R}$ be the unique continuous function which fulfills the equation

$$\varphi\left(\sum_{r=1}^{k} i_r \gamma^{-r}\right) = \sum_{r=1}^{k} \tilde{i}_r 2^{-m_r} \gamma^{-\frac{n^{r-m_r}-1}{n-1}}$$

where

$$\tilde{i}_r := i_r - (\gamma - 2)\langle i_r \rangle,$$
$$m_r := \langle i_r \rangle \left(1 + \sum_{s=1}^{r-1} \prod_{t=s}^{r-1} [i_t]\right)$$

for all $r \geq 1$, $\langle i_1 \rangle = 0$, $[i_1] = 0$ and

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

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

for $r \geq 2$.

Sprecher has proved that there exists such a unique continuous function φ, mainly because φ is strictly increasing on $\mathbb{Q}_\gamma$ and

$$|\varphi(d) - \varphi(d')| \leq \frac{1}{\gamma \cdot 2^{k-1}} \tag{1}$$

holds for all consecutive numbers $d, d' \in \mathbb{Q}_{\gamma k}$ and all $k \geq 1$. Especially, φ is strictly increasing and $\varphi([0,1]) = [0,1]$. We will use the same idea and the previous proposition to prove:

Proposition 2.4 (Computability of Sprecher's function) *For each $n \geq 2$ and $\gamma \geq 2n+2$ Sprecher's function $\varphi : [0,1] \to \mathbb{R}$ is computable.*

Proof. Let $n \geq 2$ and $\gamma \geq 2n+2$. Since a standard numbering ν allows to extract the digits $i_1, ..., i_k$ of the γ-expansion of each number $\nu(l) \in \mathbb{Q}_{\gamma k}$ from l, we can directly deduce from the definition of φ that $\varphi \circ \nu$ is a computable sequence of real numbers.

Now, let $c \in \mathbb{N}$ be a number such that $c \geq \log_2 \gamma$. Then $m : \mathbb{N} \to \mathbb{N}$ with $m(k) = c \cdot k$ is a computable function. Let $x, y \in [0,1]$ and $k \in \mathbb{N}$ with $|x-y| < 2^{-m(k)}$ where w.l.o.g. $x \leq y$. Then $|x-y| < \gamma^{-k}$ and there exist consecutive numbers $d < d'$ in $\mathbb{Q}_{\gamma k}$ such that $d \leq x \leq y \leq d'$, where at least the first or the last inequality is strict. Since φ is strictly increasing by definition, we obtain by Equation 1

$$|\varphi(x) - \varphi(y)| < |\varphi(d) - \varphi(d')| \leq \frac{1}{\gamma \cdot 2^{k-1}} \leq \frac{1}{2^k}$$

Thus, the function m is a computable modulus of continuity of φ and by Proposition 2.2 φ is a computable function. □

The proof especially shows that Sprecher's function φ is Lipschitz continuous of order $1/\log_2 \gamma$, i.e.

$$|\varphi(x) - \varphi(y)| \leq b \cdot |x-y|^{1/\log_2 \gamma}$$

for some suitable constant $b \in \mathbb{N}$. But on the other hand, it is easy to see that φ is not differentiable. Figure 2 shows the graph of Sprecher's function for $n = 2$ and $\gamma = 10$ together with a part of the graph magnified by factor 10.

The important property of Sprecher's function is that certain intervals I which correspond to the "plateaus" in the graph of φ are mapped to very small intervals $\varphi(I)$. To be more precise, define for each fixed $n \geq 2$ and $\gamma \geq 2n+1$

$$I_i^k := \left[i \cdot \gamma^{-k}, i \cdot \gamma^{-k} + \delta_k\right]$$

for all $k \geq 1$, and $i = 1, ..., \gamma^k - 1$, where

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

Sprecher has proved [Spr96] that disjoint intervals I_i^k of each fixed precision k are mapped to small disjoint intervals $\varphi(I_i^k)$. This "separation property" will implicitly be used to prove the Fundamental Lemma 2.7 below.

It is well-known that one can define computable functions by case distinction, provided that both functions coincide at the border. More precisely, if $f : [0,1] \to \mathbb{R}$ and $g : [1,2] \to \mathbb{R}$ are computable functions with $f(1) = g(1)$, then $h : [0,2] \to \mathbb{R}$ defined by

$$h(x) := \begin{cases} f(x) & \text{if } 0 \leq x \leq 1 \\ g(x) & \text{if } 1 < x \leq 2 \end{cases}$$

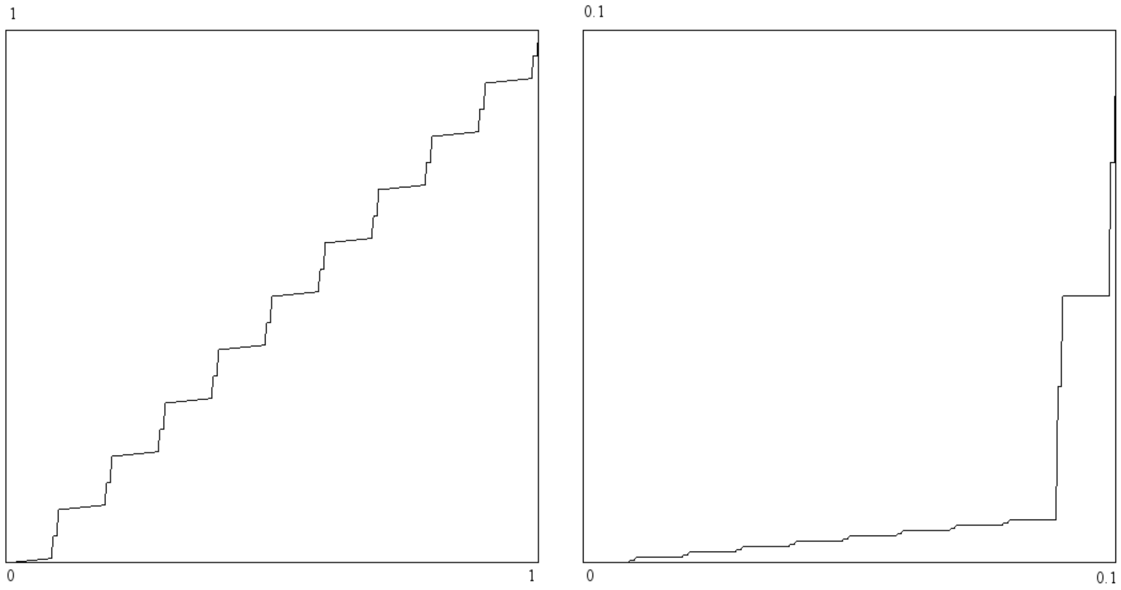

Figure 2: The graph of Sprecher's function φ for $n = 2$ and $\gamma = 10$.

is a computable function too [Wei00]. Since $\varphi(1) = 1$ we can extend Sprecher's function to the interval $[0, 2]$ by $\varphi(x + 1) := \varphi(x) + 1$ for all $x \in (0, 1]$. Thus, from now on we can assume without loss of generality that Sprecher's function is a computable function $\varphi : [0, 2] \to \mathbb{R}$, if necessary. Especially, we will use this fact in the following definition where we define functions $\xi_q : [0, 1]^n \to \mathbb{R}$ with the help of Sprecher's function φ. Essentially, this is the point where the superposition takes place, i.e. the reduction of the number of variables.

Definition 2.5 (Kolmogorov superpositions) Let $n \geq 2$ and $\gamma \geq 2n + 2$. Define

(1) $\lambda_p \in \mathbb{R}$ for $p = 1, 2, ..., n$ and $\lambda \in \mathbb{R}$ by

$$\lambda_1 := \frac{1}{2}, \quad \lambda_{p+1} := \frac{1}{2} \sum_{r=1}^{\infty} \gamma^{-p\frac{n^r - 1}{n-1}}, \quad \lambda := \sum_{p=1}^{n} \lambda_p,$$

(2) $\varphi_q : [0, 1] \to \mathbb{R}$ for $q = 0, 1, ..., 2n$ by

$$\varphi_q(x) := \frac{1}{2n+1} \left(\frac{1}{2}\varphi\left(x + \frac{q}{\gamma(\gamma - 1)}\right) + \frac{1}{\lambda} q \right),$$

(3) $\xi_q : [0, 1]^n \to \mathbb{R}$ for $q = 0, 1, ..., 2n$ by

$$\xi_q(x_1, ..., x_n) := \sum_{p=1}^{n} \lambda_p \varphi_q(x_p).$$

Let us consider an example: for instance for $n = 2$, $\gamma = 10$ and $q = 0$ we obtain

$$\xi_0(x, y) = \frac{1}{20}\varphi(x) + \frac{\lambda_2}{10}\varphi(y)$$

with constant

$$\lambda_2 = \frac{1}{2}\sum_{r=1}^{\infty} 10^{-(2^r-1)} = 0.050500050000000500000000000000005...$$

Since we will use the functions φ_q and λ for the computable version of Kolmogorov's Superposition Theorem 1.2 as stated in the introduction, we have to prove that these functions and constants are computable.

Proposition 2.6 (Computability of the Kolmogorov superposition) *Let $n \geq 2$ and $\gamma \geq 2n + 2$. Then the following holds:*

(1) $\lambda_p \in \mathbb{R}$ is computable for all $p = 1, 2, ..., n$ and $\lambda \in \mathbb{R}$ is computable,

(2) $\varphi_q : [0, 1] \to \mathbb{R}$ is computable for all $q = 0, 1, ..., 2n$,

(3) $\xi_q : [0, 1]^n \to \mathbb{R}$ is computable for all $q = 0, 1, ..., 2n$.

Proof. Let $n \geq 2$ and $\gamma \geq 2n + 2$.

(1) Obviously, λ_1 is computable. We note that for $p \geq 1$ and $\gamma > 2$ and thus

$$\left|\sum_{r=1}^{k} \gamma^{-p\frac{n^r-1}{n-1}} - \sum_{r=1}^{i} \gamma^{-p\frac{n^r-1}{n-1}}\right| < \sum_{r=k+1}^{i} 2^{-\frac{n^r-1}{n-1}} \leq \sum_{r=k+1}^{i} 2^{-r} < 2^{-k}$$

for all $i > k$. Thus, $(\sum_{r=1}^{k} \gamma^{-p\frac{n^r-1}{n-1}})_{k\in\mathbb{N}}$ is a rapidly converging and computable sequence of real numbers and the limit $2\lambda_{p+1}$ is a computable real number. Hence, λ is computable too.

(2) Since $\frac{q}{\gamma(\gamma-1)} < 1$ for $q = 0, 1, ..., 2n+1$ the functions φ_q are well-defined, if we assume that Sprecher's function φ is defined on $[0, 2]$. Moreover, it is straightforward to prove that φ_q is computable for $q = 0, 1, ..., 2n + 1$ using some standard closure properties of computable functions [Wei00] and the fact that λ is computable.

(3) Follows directly from (1) and (2).

$\square$

The next step is to transfer the nice separation property of Sprecher's function to the Kolmogorov superpositions ξ_q. Essentially, this has already been done by Sprecher [Spr96]. We use his results to prove the Fundamental Lemma 2.7 below. Roughly speaking, the idea is that ξ_q maps certain cubes S to very small intervals $\xi_q(S)$, and such that certain disjoint cubes are mapped to disjoint intervals. We will define these cubes with the help of the intervals I_i^k defined above. Since later on we will need coverings of $[0,1]^n$, it is not sufficient to consider purely the intervals I_i^k, but we will need $2n+1$ families I_{qi}^k of such intervals where $I_{qi}^k := I_i^k - aq$ is a slight translation of I_i^k defined precisely as follows: for each fixed $n \geq 2$ and $\gamma \geq 2n+1$ let

$$I_{qi}^k := \left[i \cdot \gamma^{-k} - aq, i \cdot \gamma^{-k} + \delta_k - aq\right] \cap [0,1]$$

for all $k \geq 1$, $q = 0, 1, ..., 2n$ and $i = 1, ..., \gamma^k$, where δ_k is defined as above and

$$a := \frac{1}{\gamma(\gamma-1)}.$$

Figure 3 visualizes the translated intervals I_{qi}^1 in case of $n=2$ and $\gamma = 10$ (row q displays the intervals $I_{q1}^1, ..., I_{q10}^1$).

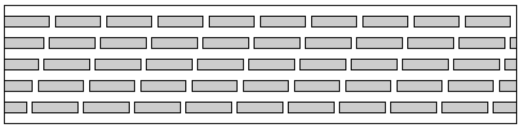

Figure 3: The intervals I_{qi}^1 for $n=2$ and $\gamma = 10$.

Now the following lemma states the separation property of the Kolmogorov superpositions which can be deduced from the separation properties of Sprecher's function φ on corresponding intervals.

Lemma 2.7 (Fundamental Lemma) *For each fixed $n \geq 2$, $\gamma \geq 2n+2$ and $k \geq 1$, the intervals $\xi_q(S_{qi_1...i_n}^k)$ with*

$$S_{qi_1...i_n}^k := I_{qi_1}^k \times I_{qi_2}^k \times ... \times I_{qi_n}^k,\ i_1, ..., i_n = 1, ..., \gamma^k,\ q = 0, ..., 2n$$

are pairwise disjoint. Moreover, for each $k \geq 1$ each point $x \in [0,1]^n$ is covered for at least $n+1$ values of $q = 0, ..., 2n$ by some cube $S_{qi_1...i_n}^k, i_1, ..., i_n = 1, ..., \gamma^k$.

Proof. Let $q \in \{0, ..., 2n\}$. In [Spr96] Sprecher has proved that the intervals $\Xi(S_{qi_1...i_n}^k)$ with $i_1, ..., i_n = 1, ..., \gamma^k$ are pairwise disjoint, where $\Xi : [0,1]^n \to \mathbb{R}$ is the function defined by

$$\Xi(x_1, ..., x_n) := \sum_{p=1}^{n} 2\lambda_p \varphi(x_p + aq).$$

Since $2\lambda = \sum_{p=1}^{n} 2\lambda_p < 2$ and $\varphi[0,2] = [0,2]$, we obtain range$(\Xi) \subseteq [0,4)$ and

$$\begin{aligned}
\xi_q(x_1,...,x_n) &= \sum_{p=1}^{n} \lambda_p \frac{1}{2n+1}\left(\frac{1}{2}\varphi\left(x_p + aq\right) + \frac{1}{\lambda}q\right) \\
&= \frac{1}{2n+1}\left(\frac{1}{4}\sum_{p=1}^{n} 2\lambda_p \varphi(x_p + aq) + q\frac{1}{\lambda}\sum_{p=1}^{n}\lambda_p\right) \\
&= \frac{1}{2n+1}\left(\frac{1}{4}\Xi(x_1,...,x_n) + q\right).
\end{aligned}$$

Thus range$(\xi_q) \subseteq \left[\frac{q}{2n+1}, \frac{q+1}{2n+1}\right)$ and $\xi_q(S^k_{qi_1...i_n})$ is just an affine transformation of range(Ξ). Hence the intervals $\xi_q(S^k_{qi_1...i_n})$ with $i_1,...,i_n = 1,...,\gamma^k, q = 0,...,2n$ are disjoint.

Since the distance of two consecutive intervals I^k_{qi} and I^k_{qi+1} is $1/((\gamma-1)\gamma^k) \leq a$ and the length of I^k_{qi} is $\delta_k > 2na$, it follows that each point $x \in [0,1]$ is covered for at least $2n$ of the $2n+1$ values $q = 0,...,2n$ by some interval I^k_{qi}, $i = 1,...,\gamma^k$. This implies that each point $x \in [0,1]^n$ is covered for at least $n+1$ of the $2n+1$ values $q = 0,...,2n$ by some cube $S^k_{qi_1...i_n}$, $i_1,...,i_n = 1,...,\gamma^k$. $\square$

Figure 4 shows the squares S^k_{qij} with delay $q = 0$, precision $k = 1$, dimension $n = 2$ and base $\gamma = 10$.

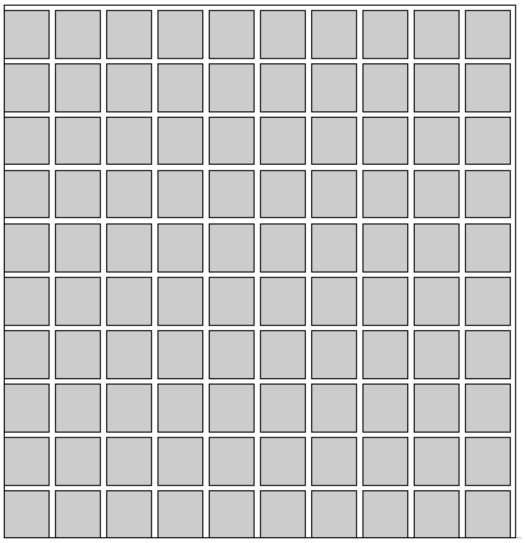

Figure 4: The squares S^1_{0ij} for $n = 2$ and $\gamma = 10$.

In the next section we will use the Fundamental Lemma to prove the computational version of Kolmogorov's Superposition Theorem.

3 Computable Kolmogorov Superposition Theorem

In this section we will use the functions defined in the previous section to prove a computable version of Kolmogorov's Superposition Theorem. For each $n \geq 2$, $\gamma \geq 2n+2$ and each function $g : [0,1] \to \mathbb{R}$ we define a function $h_g : [0,1]^n \to \mathbb{R}$ by $h_g := \sum_{q=0}^{2n} g \circ \xi_q$, i.e.

$$h_g(x_1, ..., x_n) := \sum_{q=0}^{2n} g\xi_q(x_1, ..., x_n) = \sum_{q=0}^{2n} g\left(\sum_{p=1}^{n} \lambda_p \varphi_q(x_p)\right),$$

where $\xi_q, \varphi_q, \lambda_p$ are the functions and constants, respectively, as defined in the previous section. Then the statement of Kolmogorov's Superposition Theorem could be reformulated as follows: for each continuous f there exists some continuous g such that $f = h_g$. As a preparation of the proof we start with Lorentz' Lemma, which has been originally proved by Lorentz [Lor66]. For completeness, we adapt the proof to our setting. If X is a topological space then we denote by $\mathcal{C}(X)$ the set of continuous functions $f : X \to \mathbb{R}$ and by

$$||f|| := \sup_{x \in [0,1]^n} |f(x)|$$

we denote the supremum norm for functions $f \in \mathcal{C}([0,1]^n)$.

Lemma 3.1 (Lorentz' Lemma) *Let $n \geq 2$, $\gamma \geq 2n+2$ and $\theta := (2n+1)/(2n+2)$. For each continuous function $f : [0,1]^n \to \mathbb{R}$ there exists some continuous function $g : [0,1] \to \mathbb{R}$ such that*

$$||f - h_g|| < \theta ||f|| \text{ and } ||g|| < \frac{1}{n}||f||.$$

Proof. Let $\varepsilon > 0$ be such that $\frac{n}{n+1} + \varepsilon < \theta$. Since f is uniformly continuous on $[0,1]^n$, we can choose some $k \geq 1$ such that the oscillation of f on each cube $S^k_{qi_1...i_n}$ is less than $\varepsilon||f||$, i.e.

$$\max_{\substack{i_1,...,i_n=1,...,\gamma^k \\ q=0,...,2n}} \operatorname{diam} f(S^k_{qi_1...i_n}) < \varepsilon||f||.$$

Let $c^k_{qi_1...i_n}$ be the center of the cube $S^k_{qi_1...i_n}$. Now we can define $g : [0,1] \to \mathbb{R}$ by

$$g(x) := \tfrac{1}{n+1} f(c^k_{qi_1...i_n}) \text{ for all } x \in \xi_q(S^k_{qi_1...i_n})$$

and $q = 0, ..., 2n$ and $i_1, ..., i_n = 1, ..., \gamma^k$ and by linearization for all other $x \in [0,1]$. By the Fundamental Lemma 2.7 g is well-defined. Moreover, for each $x \in [0,1]^n$ there are at least $n+1$ values of $q = 0, ..., 2n$ such that $x \in S^k_{qi_1...i_n}$ for some $i_1, ..., i_n = 1, ..., \gamma^k$. For these $n+1$ values of q we obtain $g\xi_q(x) = \frac{1}{n+1} f(c^k_{qi_1...i_n})$ and $|f(c^k_{qi_1...i_n}) - f(x)| < \varepsilon||f||$. For the remaining n values of q we obtain $|g\xi_q(x)| \leq \frac{1}{n+1}||f||$. Altogether,

$$||f - h_g|| = \left|\left| f - \sum_{q=0}^{2n} g\xi_q \right|\right| < \frac{n+1}{n+1}\varepsilon||f|| + \frac{n}{n+1}||f|| < \theta||f||$$

□

To conclude the effective version of Kolmogorov's Superposition Theorem from Lorentz' Lemma, we have to use some further notions from computable analysis. We want to show that the operator, which maps f to a corresponding function g such that $f = h_g$, is computable. To express computability of such operators $\mathcal{C}[0,1]^n \to \mathcal{C}[0,1]$ we need a representation of $\mathcal{C}^n := \mathcal{C}([0,1]^n)$ (for short we write $\mathcal{C} := \mathcal{C}^1$). We will use the so-called *Cauchy representation* $\delta^n :\subseteq \Sigma^{\mathbb{N}} \to \mathcal{C}^n$ (cf. [Wei00]). Roughly speaking, $\delta^n(p) = f$, if and only if p is an (appropriately encoded) sequence of rational polynomials $(p_i)_{i\in\mathbb{N}}$, $p_i \in \mathbb{Q}[x_1, ..., x_n]$ which converges rapidly to f, i.e. $||p_i - p_k|| \leq 2^{-k}$ for all $i > k$. Computability on the space $\mathcal{C}^n$ will from now on be understood w.r.t. δ^n. An important property of the representation δ^n is that it allows evaluation and type conversion: the function $\mathcal{C}^n \times [0,1]^n \to \mathbb{R}, (f, x) \mapsto f(x)$ is computable (i.e. (δ^n, ρ^n, ρ)–computable) and for any representation δ of X a function $F : X \times [0,1]^n \to \mathbb{R}$ is (δ, ρ^n, ρ)–computable, if and only if the function $G : X \to \mathcal{C}^n$ with $G(x)(y) := F(x, y)$ is (δ, δ^n)–computable (cf. [Wei00]). For the proof of the computable Kolmogorov Superposition Theorem we will additionally use a notion of effectivity for subsets.

Definition 3.2 (Recursively enumerable open subsets) Let δ, δ' be representations of X, Y, respectively. An open subset $U \subseteq X \times Y$ is called (δ, δ')-*r.e.*, if there is a (δ, δ', ρ)–computable function $F : X \times Y \to \mathbb{R}$ such that $(X \times Y) \setminus U = f^{-1}\{0\}$.

For instance the set $\{(x, y) \in \mathbb{R}^2 : x < y\}$ is an r.e. open subset (usually we will not mention the representations if they are clear from the context). It is straightforward to generalize this definition to subsets of higher arity. For spaces Y like $\mathbb{N}, \mathbb{R}, \mathcal{C}^n$ we have the following uniformization property: if $U \subseteq X \times Y$ is an r.e. open subset such that for any $x \in X$ there exists some $y \in Y$ such that $(x, y) \in U$, then there is a computable multi-valued operation $S : X \rightrightarrows Y$ such that $\mathrm{graph}(S) \subseteq U$, i.e. $(x, y) \in Y$ for any $x \in X$ and $y \in S(x)$ (cf. [Bra99]).

Finally, we mention a closure property of computable multi-valued operations which we will use. To express this closure property we use a canonical sequence representation: whenever $\delta :\subseteq \Sigma^\omega \to X$ is a representation of X, we assume that $\delta^\omega :\subseteq \Sigma^\omega \to X^{\mathbb{N}}$ is the associated sequence representation (cf. [Wei00]).

Lemma 3.3 (Sequential iteration for multi-valued operations) *Let δ be a representation of X. If $F : X \rightrightarrows X$ is (δ, δ)–computable, then $G : X \rightrightarrows X^{\mathbb{N}}$, defined by*

$$G(x) := \{(x_n)_{n\in\mathbb{N}} : x_0 = x \textit{ and } (\forall n)\ x_{n+1} \in F(x_n)\}$$

is (δ, δ^ω)–computable.

This can be proved straightforwardly (cf. [Bra99]). Now we are prepared to prove the computable Kolmogorov Superposition Theorem.

Theorem 3.4 (Computable Kolmogorov Superposition Theorem) *Let $n \geq 2$ and $\gamma \geq 2n+2$. There exists a computable multi-valued operation $K : \mathcal{C}^n \rightrightarrows \mathcal{C}$ such that $f = \sum_{q=0}^{2n} g \circ \xi_q$, i.e.*

$$f(x_1, ..., x_n) = \sum_{q=0}^{2n} g\left(\sum_{p=1}^{n} \lambda_p \varphi_q(x_p)\right)$$

for all $f \in \mathcal{C}^n$ and $g \in K(f)$.

Proof. Using the concepts of evaluation and type conversion and Proposition 2.6 we can conclude that the function $h : \mathcal{C} \to \mathcal{C}^n, g \mapsto h_g$ is computable. Thus, Lorentz' Lemma 3.1 and the computability of the supremum norm $\mathcal{C}^n \to \mathbb{R}, f \mapsto ||f||$ does directly imply that the set

$$M := \left\{(f, g) \in \mathcal{C}^n \times \mathcal{C} : ||f - h_g|| < \theta||f|| \text{ and } ||g|| < \frac{1}{n}||f||\right\}$$

with $\theta := (2n+1)/(2n+2)$ is an r.e. open set, i.e. there exists a computable function $F : \mathcal{C}^n \times \mathcal{C} \to \mathbb{R}$, such that $F^{-1}\{0\} = (\mathcal{C}^n \times \mathcal{C}) \setminus M$. This again implies by uniformization and Lorentz' Lemma 3.1 that there exists a computable multi-valued operation $L : \mathcal{C}^n \rightrightarrows \mathcal{C}$ such that $(f, g) \in M$ for all $f \in \mathcal{C}^n$ and $g \in L(f)$.

Now let $f \in \mathcal{C}^n$. We define a sequence $(g_i)_{i \in \mathbb{N}}$ of functions in $\mathcal{C}$ inductively as follows: we choose some $g_0 \in L(f)$ and if g_i is already defined, we choose some

$$g_{i+1} \in L(f - h_{g_0} - h_{g_1} - ... - h_{g_i}).$$

By Lorentz' Lemma 3.1 and induction we obtain

$$\left|\left|f - \sum_{i=0}^{r} h_{g_i}\right|\right| < \theta^{r+1}||f|| \text{ and } ||g_r|| < \frac{1}{n}\theta^r||f||.$$

These inequalities imply that the series $g := \sum_{i=0}^{\infty} g_i$ and $\sum_{i=0}^{\infty} h_{g_i}$ converge uniformly and effectively. By continuity and linearity of h we obtain

$$f = \sum_{i=0}^{\infty} h_{g_i} = h_{\sum_{i=0}^{\infty} g_i} = h_g = \sum_{q=0}^{2n} g \circ \xi_q.$$

We still have to show that the procedure which maps f to g defines a computable operation $K : \mathcal{C}^n \rightrightarrows \mathcal{C}$. This follows by applying some recursive closure schemes for multi-valued operations (cf. [Bra99]). To make this a little bit more precise, we first note that

$$N := \{(f, k, r) \in \mathcal{C}^n \times \mathbb{N} \times \mathbb{N} : \theta^{r+1}||f|| < 2^{-k}\}$$

is an r.e. open set, since the supremum norm is a computable function. This does imply that there is a computable modulus operation $m : \mathcal{C}^n \times \mathbb{N} \rightrightarrows \mathbb{N}$ such that $\text{graph}(m) \subseteq N$. Now we define an operation $T : \mathcal{C}^n \rightrightarrows \mathcal{C}^{\mathbb{N}}$ by

$$T(f) := \left\{(g_i)_{i \in \mathbb{N}} : (\forall j)\ g_j \in L\left(f - \sum_{i=0}^{j-1} h_{g_i}\right)\right\}.$$

Then T is a computable multi-valued operation, which follows by applying Lemma 3.3 to the function $F :\mathcal{C}^n \times \mathcal{C} \rightrightarrows \mathcal{C}^n \times \mathcal{C}$ with $F(f,g) := (f - h_g, L(f - h_g))$. Now we define another operation $S : \mathcal{C}^n \times \mathcal{C}^{\mathbb{N}} \times \mathbb{N} \rightrightarrows \mathcal{C}$ by $S(f,(g_i)_{i\in\mathbb{N}},k) := \sum_{i=0}^{m(f,k)} h_{g_i}$. Then S is computable too and so is $[S] : \mathcal{C}^n \times \mathcal{C}^{\mathbb{N}} \rightrightarrows \mathcal{C}^{\mathbb{N}}$, defined by

$$[S](f,(g_i)_{i\in\mathbb{N}}) := \{(h_i)_{i\in\mathbb{N}} : (\forall k)\ h_k \in S(f,(g_i)_{i\in\mathbb{N}},k)\}.$$

Finally, $K := \mathrm{Lim} \circ [S] \circ (\mathrm{id}, T) : \mathcal{C}^n \rightrightarrows \mathcal{C}$ is computable, where $\mathrm{Lim} :\subseteq \mathcal{C}^{\mathbb{N}} \to \mathcal{C}, (f_n)_{n\in\mathbb{N}} \mapsto \lim_{n\to\infty} f_n$ denotes the limit operation (w.r.t. to the maximum metric) which is defined only for rapidly converging sequences, i.e. $\mathrm{dom}(\mathrm{Lim}) := \{(f_n) \in \mathcal{C}^{\mathbb{N}} : (\forall i > k)||f_i - f_k|| \leq 2^{-k}\}$. $\square$

Especially, the proof shows that $\xi = (\xi_0, ..., \xi_{2n}) : [0,1]^n \to \mathbb{R}^{2n+1}$ is an embedding of $[0,1]^n$ into $\mathbb{R}^{2n+1}$. Thus, we are in the situation of the following commutative diagram:

$$\begin{tikzcd}
{[0,1]^n} \arrow[r, "f"] \arrow[d, "\xi"'] & \mathbb{R} \\
{[0,1]^{2n+1}} \arrow[r, "\times_{i=0}^{2n} g"'] & \mathbb{R}^{2n+1} \arrow[u, "\Sigma"']
\end{tikzcd}$$

Figure 5: Kolmogorov superposition

4 Representation of Multivariate Functions

In this section we will briefly discuss another possibility to express the computable Kolmogorov Superposition Theorem as an equivalence of representations. For two representations δ, δ' of a set X we write $\delta \leq \delta'$ and we say that δ is *reducible* to δ', if there exists some computable $F :\subseteq \Sigma^\omega \to \Sigma^\omega$ such that $\delta(p) = \delta' F(p)$ for all $p \in \mathrm{dom}(\delta)$. Furthermore, δ is called *equivalent* to δ' and we write $\delta \equiv \delta'$, if and only if $\delta \leq \delta'$ and $\delta' \leq \delta$.

The Kolmogorov Superposition Theorem gives rise to another representation of the space of multivariate continuous functions $\mathcal{C}^n$. We define $\delta^n_{\mathrm{Kolmogorov}} :\subseteq \Sigma^\omega \to \mathcal{C}^n$ by

$$\delta^n_{\mathrm{Kolmogorov}}(p) := \sum_{q=0}^{2n} \delta^1(p)\xi_q$$

for all $p \in \Sigma^\omega$ and $n \geq 2$, where $\xi_q : [0,1]^n \to \mathbb{R}$ are the functions as defined in Definition 2.5. The classical Kolmogorov Superposition Theorem 1.1 implies that $\delta^n_{\mathrm{Kolmogorov}}$

is surjective, i.e. a representation of $\mathcal{C}^n$. Now the computable Kolmogorov Superposition Theorem 3.4 can be rephrased as follows:

Corollary 4.1 $\delta^n \equiv \delta^n_{\text{Kolmogorov}}$ *for all* $n \geq 2$.

More precisely, the reducibility "$\geq$" follows from the computability of the function $h : \mathcal{C}^1 \to \mathcal{C}^n, g \mapsto h_g$, as defined in the previous section, and the reducibility "$\leq$" follows from the existence of a computable multi-valued operation $K : \mathcal{C}^n \rightrightarrows \mathcal{C}$, as expressed by Theorem 3.4.

5 Conclusion

We have presented a computable version of Kolmogorov's Superposition Theorem. An interesting line of continuation could be to study the computational complexity of this problem as well as further applications to the theory of neural networks.

Acknowledgement

The investigation presented in this paper has been kindly suggested and supported by Jiří Wiedermann. Moreover, I would like to thank Věra Kůrková for sending me her helpful notices on the classical proof of Kolmogorov's Superposition Theorem and Vladik Kreinovich for sending me copies of a number of related papers.

References

[Bra99] Vasco Brattka. Recursive and computable operations over topological structures. Informatik Berichte 255, FernUniversität Hagen, Fachbereich Informatik, Hagen, July 1999. Dissertation.

[Hil00] David Hilbert. Mathematische Probleme. *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, pages 253–297, 1900. Vortrag, gehalten auf dem internationalen Mathematiker-Kongreß zu Paris 1900.

[Hil27] David Hilbert. Über die Gleichung neunten Grades. *Mathematische Annalen*, 97:243–250, 1927.

[HN87] Robert Hecht-Nielsen. Kolmogorov's mapping neural network existence theorem. In *Proceedings IEEE International Conference On Neural Networks*, volume II, pages 11–13, New York, 1987. IEEE Press.

[HN90] Robert Hecht-Nielsen. *Neurocomputing*. Addison-Wesley, Reading, 1990.

[Ko91] Ker-I Ko. *Complexity Theory of Real Functions*. Progress in Theoretical Computer Science. Birkhäuser, Boston, 1991.

[Kol57] A.N. Kolmogorov. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Dokl. Akad. Nauk USSR*, 114:953–956, 1957. [translated in: American Mathematical Society Translations 28 (1963) 55–59].

[KS94] Hidefume Katsuura and David A. Sprecher. Computational aspects of Kolmogorov's superposition theorem. *Neural Networks*, 7(3):455–461, 1994.

[Kur91] Věra Kurková. 13th Hilbert's problem and neural networks. In M. Novák and E. Pelikán, editors, *Theoretical Aspects of Neurocomputing*, pages 213–216, Singapore, 1991. World Scientific. Symposium on Neural Networks and Neurocomputing, NEURONET, Prague, 1990.

[Kur92] Věra Kurková. Kolmogorov's theorem and multilayer neural networks. *Neural Networks*, 5:501–506, 1992.

[Lor66] G.G. Lorentz. *Approximation of Functions.* Athena Series, Selected Topics in Mathematics. Holt, Rinehart and Winston, Inc., New York, 1966.

[Lor76] G.G. Lorentz. The 13-th problem of Hilbert. In Felix E. Browder, editor, *Mathematical developments arising from Hilbert problems*, volume 28 of *Proceedings of the Symposium in Pure Mathematics of the AMS*, pages 419–430, Providence, 1976. American Mathematical Society. Northern Illinois Univiversity 1974.

[NMK93] Mutsumi Nakamura, Ray Mines, and Vladik Kreinovich. Guaranteed intervals for Kolmogorov's theorem (and their possible realtion to neural networks). *Interval Computations*, 3:183–199, 1993. Proceedings of the International Conference on Numerical Analysis with Automatic Result Verification (Lafayette, LA, 1993).

[PER89] Marian B. Pour-El and J. Ian Richards. *Computability in Analysis and Physics.* Perspectives in Mathematical Logic. Springer, Berlin, 1989.

[Spr65] David A. Sprecher. On the structure of continuous functions of several variables. *Transactions American Mathematical Society*, 115(3):340–355, 1965.

[Spr93] David A. Sprecher. A universal mapping for Kolmogorov's superposition theorem. *Neural Networks*, 6:1089–1094, 1993.

[Spr96] David A. Sprecher. A numerical implementation of Kolmogorov's superpositions. *Neural Networks*, 9(5):765–772, 1996.

[Spr97] David A. Sprecher. A numerical implementation of Kolmogorov's superpositions II. *Neural Networks*, 10(3):447–457, 1997.

[Wei00] Klaus Weihrauch. *Computable Analysis.* Springer, Berlin, 2000.