Pergamon

Neural Networks, Vol. 9, No. 5, pp. 765–772, 1996

Printed in Great Britain
0893–6080/96 $15.00+.00

0893–6080(95)00081–X

CONTRIBUTED ARTICLE

A Numerical Implementation of Kolmogorov's Superpositions

DAVID A. SPRECHER

University of California

(*Received* 21 *November* 1994; *revised and accepted* 9 *May* 1995)

Abstract—*Hecht-Nielsen proposed a feedforward neural network based on Kolmogorov's superpositions*

$$f(x_1,\dots,x_n)=\sum_{q=0}^{2n}\Phi_q(y_q)$$

that apply to all real valued continuous functions $f(x_1,\dots,x_n)$ *defined on a Euclidean unit cube of dimension* $n\geq 2$. *This network has a hidden layer that is independent of f and that transforms the n-tuples* $(x_1,\dots,x_n)$ *into the* $2n+1$ *variables* y_q, *and an output layer in which f is computed. Kůrková has shown that such a network has an approximate implementation with arbitrary activation functions of sigmoidal type. Actual implementation is, however, impeded by the lack of numerical algorithms for the hidden layer which contain continuous functions of the form* $y_q=\sum_{p=1}^{n}\alpha_p\psi(x_p+qa)$ *with constants a and* α_p. *This paper gives an explicit numerical implementation of the hidden layer that also enables the implementation of the output layer.*

Keywords—Superpositions, Kolmogorov, Representation of functions of several variables, Feedforward neural networks, Continuity.

1. INTRODUCTION

Kolmogorov's superposition representation tells us that every continuous function $f:\mathscr{E}^n\to\mathscr{R}$ defined on the n-dimensional Euclidean unit cube $\mathscr{E}^n$ and with range on the real line $\mathscr{R}$ can be represented as a sum of continuous functions:

$$f(x_1,\dots,x_n)=\sum_{q=0}^{2n}\Phi_q(y_q).$$

In this representation, the variables $x_1,\dots,x_n$ are the parameters of an embedding of $\mathscr{E}^n$ into $\mathscr{R}^{2n+1}$ of the form

$$y_q=\sum_{p=1}^{n}\alpha_p\psi(x_p+qa)$$

with a continuous function ψ and suitable constants α_p and a. This embedding is independent of f, and it enables the complete separation of variables y_q in the representation of f, with the result that the simultaneous n-variable computation of f in terms of n-tuples $(x_1,\dots,x_n)$ is replaced by $2n+1$ parallel one-variable computations in $y_0,\dots,y_{2n}$, with addition accounting for dimensionality. Hecht-Nielsen (1987a, b) recognized that the specific format in Sprecher (1965) of the form

$$f(x_1,\dots,x_n)=\sum_{q=0}^{2n}\Phi_q\left[\sum_{p=1}^{n}\alpha_p\psi(x_p+qa)\right]$$

of Kolmogorov's superpositions (1957) can be interpreted as a feedforward neural network with a hidden layer that computes the variables y_q and therefore involves the embedding mapping only, and a single output layer in which f is computed by means of the functions $\Phi_q(y_q)$ (see Figure 1 and formula (2) below). What makes Hecht-Nielsen's network particularly attractive as a computational device is that the hidden layers are fixed independently of any function f, so that in theory this part of the neural network is trained once for n. [It was demonstrated recently by Sprecher (1993) (see also Katsuura & Sprecher (1994) that there are universal hidden layers that are independent even of n.]

Requests for reprints should be sent to: David A. Sprecher, Department of Mathematics, University of California, Santa Barbara, California 93106, USA; e-mail: sprecher@math.ucsb.edu

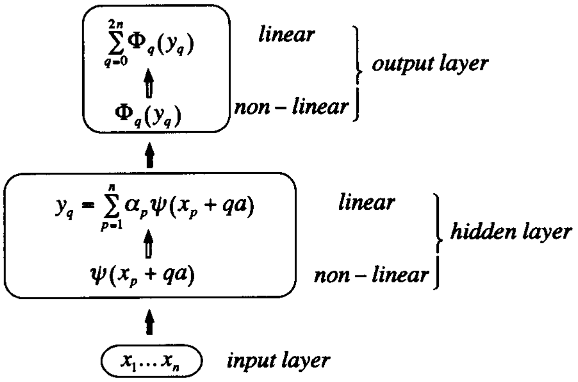

FIGURE 1. Schematic representation of formula (2).

The actual implementation of such a network as well as other applications of superpositions are impeded, however, by the lack of numerical algorithms [see Nakamura et al., (1993) and Nees (1993) for recent advances in this area]. Another factor, attributed to a perceived incompatibility between the algorithms governing the constructions of superpositions and the elements of neural networks was resolved by Kůrková (1992), who noted that Hecht-Nielsen's network does provide for an implementation by demonstrating the existence of approximate superposition representations within the constraints of neural networks, in which the functions ψ and Φ_q are approximated with functions of the form $\Sigma a_r\sigma(b_r x + c_r)$ where σ is an arbitrary activation sigmoidal function: $\sigma : \mathscr{R} \to \mathscr{E}$ such that

$$\lim_{x\to-\infty}\sigma(x) = 0 \text{ and } \lim_{x\to\infty}\sigma(x) = 1.$$

The present paper gives an explicit numerical algorithm for superpositions based on a fixed continuous function ψ , thereby enabling also the numerical implementation of the continuous functions Φ_q. The structure of the function ψ affords computational economy in the numerical implementation of an artibrary function $f : \mathscr{E}^n \to \mathscr{R}$ in that the discrete mappings in formula (1) below and Lemma 2 are sufficient for the numerical implementation of formula (2). That is to say, we require only the discrete mapping $\psi : \mathscr{D} \to \mathscr{E}$ defined on the set $\mathscr{D}$ of terminating rational numbers

$$d_k = \sum_{r=1}^{k} i_r\gamma^{-r}$$

for a suitable integer γ and a lower bound of the minimal distance beteen their discrete images under ψ as a function of k, and with this we can carry out the numerical algorithms as a finite process and obtain the approximate equality

$$f(d_{k,1},\dots,d_{k,n}) \approx \sum_{q=0}^{2n}\Phi_q\left[\sum_{p=1}^{n}\alpha_p\psi(d_{k,p}+qa)\right].$$

Central to the construction of the functions Φ_q and the establishment of formula (2) is the existence of $2n+1$ families of pairwise disjoint n-dimensional cubes for each value of q and k, whose union covers $\mathscr{E}^n$ in a prescribed manner; furthermore, the images on $\mathscr{R}$ of each family must likewise be pairwise disjoint. In the original formulation of Kolmogorov's (1957) and in subsequent work (see Nees, 1993; Sprecher, 1965 and elsewhere) these conditions are stated as part of the formulation of the superposition theorem, and the fixed inner functions $\psi_{pq}(x_p) = \alpha_p\psi(x_p+qa)$ are constructed there to meet these specifications. In the present paper, these properties are derived instead from the function ψ, and therefore no explicit mention of these is included in the statement of Theorem 1. These derivations are included in the Proofs section together with the proofs of Lemmas 1 and 2.

2. MAIN RESULTS

Let $n \leq 2$ and $\gamma \leq 2n+2$ be given integers. Consider the set $\mathscr{D}$ of terminating rational numbers

$$d_k = \sum_{r=1}^{k} i_r\gamma^{-r}$$

where $i_r = 0, 1, \dots, \gamma-1$ and $k \in N$. Let $\langle i_1\rangle = 0$ and for $r \geq 2$ let:

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

let $[i_1] = 0$ and for $r \geq 2$ let

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

We prove:

THEOREM 1. *Define the function $\psi : \mathscr{D} \to \mathscr{E}$ such that for each integer $k \in N$,*

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

where

$$\tilde{i}_r = i_r - (\gamma-2)\langle i_r\rangle$$

and

$$m_r = \langle i_r \rangle \left(1 + \sum_{s=1}^{r-1} [i_s] \times \cdots \times [i_{r-1}]\right)$$

for $r = 1, 2, \ldots, k$.

This specification uniquely determines a monotonically increasing continuous function $\psi : [0, 1] \to \mathscr{E}$ with the following property:

Let $\alpha_1 = 1$ and

$$\alpha_p = \sum_{r=1}^{\infty} \gamma^{-(p-1)\frac{n^r-1}{n-1}}$$

for $p = 2, 3, \ldots, n$; let $\psi(x+1) = \psi(x) + 1$ for $x \in [0, 1)$. Then every continuous function $f : \mathscr{E}^n \to \mathscr{R}$ has a representation

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \Phi_q \left[\sum_{p=1}^{n} \alpha_p \psi\left(x_p + \frac{q}{\gamma(\gamma-1)}\right)\right] \quad (2)$$

with continuous functions Φ_q.

According to its definition, $m_r = m_r(i_1, \ldots, i_r)$ is a function of the indices $i_1, \ldots, i_r$ with range $\{0, 1, 2, \ldots, r-1\}$, and specifically $m_r = 0$ when $i_r \neq \gamma - 1$. We note that $\tilde{i} = i_1$ and

$$\tilde{i}_r = \begin{cases} i_r \text{ when } i_r = 0, 1, \ldots, \gamma-2 \\ 1 \text{ when } \quad i_r = \gamma - 1 \end{cases} \text{for } r \geq 2$$

so that

$$\psi\left(\sum_{r=1}^{k} i_r \gamma^{-r}\right) = \sum_{r=1}^{k} i_r \gamma^{-\frac{n^r-1}{n-1}}$$

TABLE 1
The Values Generated by $\tilde{i}_r 2^{-m_r} \gamma^{-\frac{n^{r-m_r}-1}{n-1}}$, $i_r = 0, 1, \ldots, \gamma - 2$

$r \setminus m_r$	0	1	2	3	4	$\cdots$	$k-1$
1	$i_1\gamma^{-1}$						
2	$i_2\gamma^{-\frac{n^2-1}{n-1}}$	$2^{-1}\gamma^{-1}$					
3	$i_3\gamma^{-\frac{n^3-1}{n-1}}$	$2^{-1}\gamma^{-\frac{n^2-1}{n-1}}$	$2^{-2}\gamma^{-1}$				
4	$i_4\gamma^{-\frac{n^4-1}{n-1}}$	$2^{-1}\gamma^{-\frac{n^3-1}{n-1}}$	$2^{-2}\gamma^{-\frac{n^2-1}{n-1}}$	$2^{-3}\gamma^{-1}$			
5	$i_5\gamma^{-\frac{n^5-1}{n-1}}$	$2^{-1}\gamma^{-\frac{n^4-1}{n-1}}$	$2^{-2}\gamma^{-\frac{n^3-1}{n-1}}$	$2^{-3}\gamma^{-\frac{n^2-1}{n-1}}$	$2^{-4}\gamma^{-1}$		
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	
k	$i_k\gamma^{-\frac{n^k-1}{n-1}}$	$2^{-1}\gamma^{-\frac{n^{k-1}-1}{n-1}}$	$2^{-2}\gamma^{-\frac{n^{k-2}-1}{n-1}}$	$2^{-3}\gamma^{-\frac{n^{k-3}-1}{n-1}}$	$2^{-4}\gamma^{-\frac{n^{k-4}-1}{n-1}}$	$\cdots$	$2^{-(k-1)}\gamma^{-1}$

when $i_1 = 0, 1, \ldots, \gamma - 1$ and $i_2 = 0, 1, \ldots, \gamma - 2$ for $r \geq 2$.

The computation of the terms in the right side of formula (1) is given in Table 1 with the help of these observations.

Formula (1) is represented schematically in Figure 2 for $k = 4$.

The implementation of formula (2) requires the computation of values

$$\psi\left(d_k + \frac{q}{(\gamma-1)\cdot\gamma}\right)$$

for $q = 0, 1, \ldots, 2n$. This computation is facilitated through the following lemma which shows that these functional values can be obtained through the values $\psi(d_k)$ given in formula (1):

LEMMA 1. *Let an integer $k \in N$ be given. For each integer $q = 0, 1, 2, \ldots, 2n$ let the mapping $\omega_q : \mathscr{D} \to \mathscr{D}$ be defined as*

$$\omega_q(d_k) = \frac{i_1}{\gamma} + \sum_{r=2}^{k} \frac{i_r + q}{\gamma^r}.$$

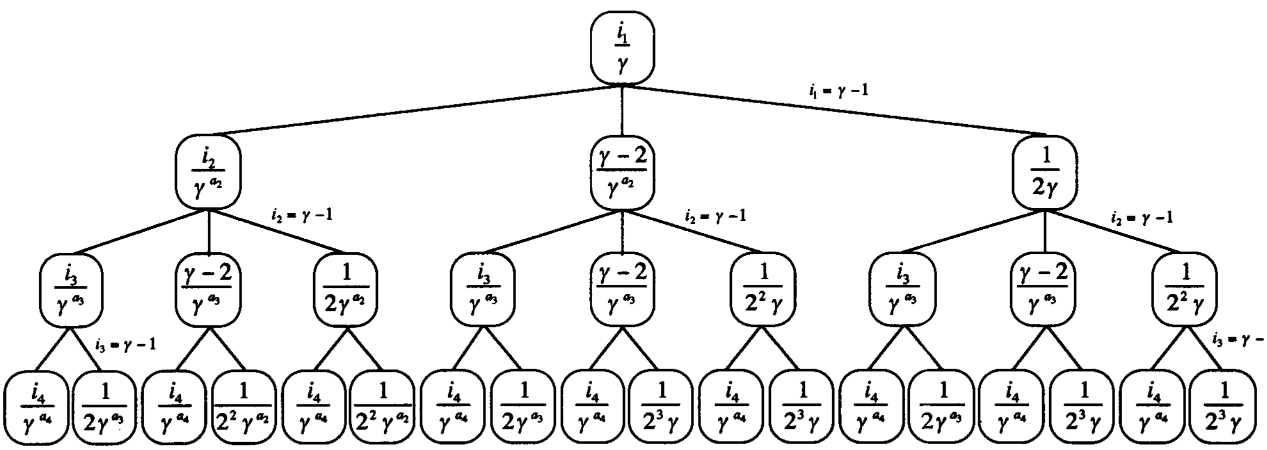

FIGURE 2. Schematic representation of formula (1) for $k = 4$; $a_r = \frac{n^r - 1}{n - 1}$.

Then

$$\psi\left(d_k + \frac{q}{(\gamma-1)\cdot\gamma}\right) = \psi(\omega_q(d_k)) + q \sum_{r=k+1}^{\infty} \gamma^{-\frac{n^r-1}{n-1}}. \quad (3)$$

Now, for each fixed value of k, let $d_{k,p}$ vary independently for $p = 1, 2, \ldots, n$ over $\mathscr{D}$, and consider the mapping

$$\sum_{p=1}^{n} \alpha_p \psi(d_{k,p}) : \mathscr{D} \times \cdots \times \mathscr{D} \to \mathscr{R} \quad (4)$$

of the rational grid-points $(d_{k,1}, \ldots, d_{k,n}) \in \mathscr{E}^n$ into the real line $\mathscr{R}$. The key to the numerical implementation of formula (2) is the computation of the minimal separation of the images of these grid-points as a function of k under the mapping (4). Clearly, this separation is determined by the minimal linear distance between consecutive points in the range of this mapping and this is the subject of the following lemma:

LEMMA 2. *For each integer $k \in N$, let $\mathscr{D}_k$ be the set of terminating rational numbers d_k set*

$$\mu_k = \sum_{p=1}^{n} \alpha_p [(d_{k,p}) - \psi(d'_{k,p})] \quad (5)$$

where d_k, $d'_k \in \mathscr{D}_k$. *Then*

$$\min_{D_k} |\mu_k| \geq \gamma^{-n\frac{n^k-1}{n-1}}$$

when

$$\sum_{p=1}^{n} |d_{k,p} - d'_{k,p}| \neq 0. \quad (6)$$

3. COMPUTATION OF $\psi(d_k)$ FOR $n = 2$

We consider now the case $n = 2$ and $\gamma = 10$. Specifically, we construct in this section the function $\psi : \mathscr{D} \to \mathscr{E}$ for the set $\mathscr{D}$ of terminating decimals such that for each integer $k \in N$,

$$\psi\left(\sum_{r=1}^{k} \frac{i_r}{10^r}\right) = \sum_{r=1}^{k} \frac{\tilde{i}_r}{2^{m_r}\, 10^{2^{r-m_r}} - 1} \quad (7)$$

where $\tilde{i}_r = i_r - 8\langle i_r \rangle$ and m_r, $r = 1, 2, \ldots, k$, is as defined in Theorem 1 above.

From Figure 1 we can compute the values of ψ. These are compiled in Table 2 in decimal notation for $k = 4$.

TABLE 2
The Values $\psi(d_k)$ for $k = 4$

$\psi(.i_1 i_2 i_3 i_4) = .i_1 0 i_2 000 i_3 0000000 i_4$	[1,2,3]
$\psi(.i_1 9 i_3 i_4) = .i_1 50000 i_3 0000000 i_4$	[2,3]
$\psi(.i_1 i_2 9 i_4) = .i_1 0 i_2 5000000000 i_4$	[3,4]
$\psi(.i_1 8 9 i_4) = .i_1 330000000000 i_4$	[3]
$\psi(.i_1 9 9 i_4) = .i_1 750000000000 i_4$	[3]
$\psi(.i_1 i_2 i_3 9) = .i_1 0 i_2 000 i_3 50000000$	[1,5]
$\psi(.i_1 i_2 8 9) = .i_1 0 i_2 250800000000$	[4]
$\psi(.i_1 i_2 9 9) = .i_1 0 i_2 750000000000$	[4]
$\psi(.i_1 88 9) = .i_1 20500800000000$	
$\psi(.i_1 89 9) = .i_1 45500000000000$	
$\psi(.i_1 9 i_3 9) = .i_1 50000 i_3 50000000$	[5]
$\psi(.i_1 98 9) = .i_1 62500800000000$	
$\psi(.i_1 99 9) = .i_1 87500000000000$	

$i_1 = 0, 1, \ldots, 9$
[1] $i_2 = 0, 1, \ldots, 8$ [4] $i_2 = 0, 1, \ldots, 7$
[2] $i_3 = 0, 1, \ldots, 8$ [5] $i_3 = 0, 1, \ldots, 7$
[3] $i_4 = 0, 1, \ldots, 8$

The values $\psi(d_{k,1}) + \alpha_2\psi(d_{k,2})$ are now easily computed for any desired values of k, and with the help of Lemma 1 we also obtain from these the values

$$\Psi\left(d_{k,1} + \frac{q}{90}\right) + \alpha_2 \psi\left(d_{k,2} + \frac{q}{90}\right)$$

that are necessary for the numerical implementation of the formula

$$f(d_{k,1}, d_{k,2}) \approx \sum_{q=0}^{4} \Phi_q \left[\psi\left(d_{k,1} + \frac{q}{90}\right) + \alpha_2 \psi\, d_{k,2} + \frac{q}{90}\right)\right]. \quad (8)$$

Following along the branches of Figure 1 (see also Table 1 in this connection) we observe that the terms in (7) corresponding respectively to $r - m_r = c$ for $c = 0, 1, 2, \ldots, k-1$ can be represented as

$$a_r^s = \sum_{u=0}^{s} \langle i_r \rangle \times \cdots \times \langle i_{r+u} \rangle \frac{1}{2^u},\ 0 \leq s \leq k - r,$$

$$r = 1, 2, \ldots, k-1;\ a_k^k = 0. \quad (9)$$

By collecting the terms $r - m_r = c$, we deduce that the right side of formula (7) can also be expressed as the sum of terms of the form

$$\frac{1}{10^{2^r-1}} (\tilde{i}_r + a_r^s)$$

for a suitable choices for r and s. With this notation, formula (7) can be written in the equivalent form

$$\psi(d_k) = \frac{1}{10}\left(\tilde{i}_1 + a_1^{s_1}\right) + \frac{1}{10^{2^{1+s_1}-1}}\left(\tilde{i}_{1+s_1} + a_{1+s_1}^{s_2}\right) + \frac{1}{10^{2^{1+s_1+s_2}-1}}\left(\tilde{i}_{1+s_1+s_2} + a_{1+s_1+s_2}^{s_3} + \cdots\right). \tag{10}$$

For the remainder of this paper we shall discuss Theorem 1 for arbitrary $n \geq 2$.

4. CONTINUITY OF ψ

By construction, ψ is a monotonically increasing function on the everywhere dense set $\mathscr{D}$ of terminating rational numbers in $\mathscr{E}$, and a direct estimate shows that if d_k, $d'_k \in \mathscr{D}$ are consecutive numbers then

$$|\psi(d_k) - \psi(d'_k)| \leq \frac{1}{\gamma \cdot 2^{k-1}}, \; k \in N.$$

These inequalities on an everywhere dense set ensure that the function $\psi : \mathscr{D} \to \mathscr{E}$ has a unique extension to a continuous function $\psi : [0, 1] \to \mathscr{E}$. With the definition $\psi(x+1) = \psi(x) + 1$ for $x \in [0, 1)$, the domain of ψ contains the interval

$$\left[0, \; 1 + \frac{2n}{(\gamma-1)\gamma}\right],$$

as required by formula (2). It will be apparent from the properties of ψ below that we could prove formula (2) with constants

$$\frac{q}{(\gamma-1)\cdot\gamma^b}$$

for any predetermined positive integer b.

5. PROOFS

Proof of Lemma 1. With the observation that

$$\frac{q}{(\gamma-1)\cdot\gamma} = q \cdot \lim_{r\to\infty} \sum_{s=2}^{r} \frac{1}{\gamma^s}$$

we can write

$$d_k + \frac{q}{(\gamma-1)\cdot\gamma} = \frac{i_1}{\gamma} + \sum_{r=2}^{k} \frac{i_r + q}{\gamma^r} + \sum_{r=1}^{\infty} \frac{q}{\gamma^{k+r}} = \omega_q(d_k) + \sum_{r=1}^{\infty} \frac{q}{\gamma^{k+r}},$$

and the lemma follows directly from formula (1).

Proof of Lemma 2. We begin the proof with the elementary statement that a unique minimum does exist. For each $k \in N$, let $d_{k,p}, d'_{k,p} \in \mathscr{D}$, $p = 1, 2, \ldots, n$, and let $A_{k,p} = \psi(d_{k,p})$. Because ψ is strictly monotonic we know that $A_{k,p} \neq 0$ for all admissible values of p and it follows from the construction that

$$\min_{D_k} |A_{k,p}| = \gamma^{-\frac{n^k-1}{n-1}}, \tag{11}$$

each minimum being taken over the decimals for which $|d_{k,p} - d'_{k,p}| \neq 0$; the upper bounds

$$\min_{D_k} |\mu_k| \leq \alpha_n \cdot \gamma^{-\frac{n^k-1}{n-1}}, \tag{12}$$

for each $k \in N$ are immediately evident from the definition of μ_k; throughout this proof we shall assume that

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

In view of (11) and (12), a necessary condition for a minimum of $|\mu_k|$ to occur is that $A_{k,T} \neq 0$ for some integer $2 \leq T \leq n$. Let us now designate the kth remainder of α_p by

$$\varepsilon_{k,p} = \sum_{r=k+1}^{\infty} \gamma^{-(p-1)\frac{n^r-1}{n-1}} \tag{13}$$

so that

$$\alpha_p - \varepsilon_{k,p} = \sum_{r=1}^{k} \gamma^{-(p-1)\frac{n^r-1}{n-1}} \tag{14}$$

and consider the expression

$$A_{k,1} + \sum_{p=2}^{T} (\alpha_p - \varepsilon_{k,p}) A_{k,p}. \tag{15}$$

We claim that

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

To show this, we observe that

$$\alpha_T - \varepsilon_{k,T} = \gamma^{-(T-1)} + \gamma^{-(T-1)\frac{n^2-1}{n-1}} + \cdots + \gamma^{-(T-1)\frac{n^k-1}{n-1}},$$

and that the largest possible term in the expansion of $|A_{k,T}|$ in powers of$1/\gamma$ is

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

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

$$0 < \tau \leq \gamma^{-(T-1)\frac{n^k-1}{n-1}} \cdot \frac{\gamma - 1}{\gamma},$$

yet the smallest possible term of $(\alpha_p - \varepsilon_{k,p})|A_{k,p}|$ for $p < T$ is

$$\gamma^{-(T-2)\frac{n^k-1}{n-1}} \cdot \gamma^{-\frac{n^k-1}{n-1}} = \gamma^{-(T-1)\frac{n^k-1}{n-1}}$$

according to (11) and (14), so that the assertion holds and (15) does indeed not vanish.

Next, we note that the expansion of (15) in powers of $1/\gamma$ contains the term

$$\gamma^{-(T-1)\frac{n^k-1}{n-1}} \cdot \gamma^{-\frac{n^k-1}{n-1}} = \gamma^{-T\frac{n^k-1}{n-1}}$$

when $|i_{k,T} - i'_{k,T}| = 1$, and a direct examination of the remaining terms in (15) shows that this is the smallest possible such term, and hence it cannot be annihilated by any other terms in (15). Accordingly,

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

But this implies that also

$$\left| A_{k,1} + \sum_{p=2}^{T} \alpha_p A_{k,p} \right| \geq \gamma^{-T\frac{n^k-1}{n-1}}$$

since all possible terms in the expansion of

$$\sum_{p=2}^{T} \varepsilon_{k,p} A_{k,p}$$

in powers of $1/\gamma$ are too small to annihilate

$$\gamma^{-T\frac{n^k-1}{n-1}},$$

and the lemma follows.

Kolmogorov (1957) formulated a number of sufficient conditions for the establishment of formula (2). These sufficient conditions are a direct consequence of the definition of ψ and they are derived from it with the two lemmas and corollary below.

Lemma 3. *For each integer $k \in N$, let*

$$\delta_k = \frac{\gamma - 2}{(\gamma - 1) \cdot \gamma^k}. \tag{16}$$

Then

$$\psi(d_k + \delta_k) = \psi(d_k) + (\gamma - 2)\varepsilon_{k,2} \tag{17}$$

for all $d_k \in \mathscr{D}$, where $\varepsilon_{k,2}$ is as defined in (13).

We note in passing that

$$\varepsilon_{k,2} = \psi\left(\sum_{r=k+1}^{\infty} \frac{1}{\gamma^r}\right) = \psi\left(\frac{1}{(\gamma-1)\gamma^k}\right) = \psi\left(\frac{\delta_k}{\gamma - 2}\right);$$

$\varepsilon_{k,2}$ is also the remainder of α_2. Formally we can express (17) as follows:

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

Proof of Lemma 3. The proof is via a direct calculation: expressing δ_k as the limit

$$\delta_k = \lim_{N \to \infty} \sum_{r=1}^{N} \frac{\gamma - 2}{\gamma^{k+r}}$$

we conclude from the definition and continuity of ψ that

$$\begin{aligned}
\psi(d_k + \delta_k) &= \lim_{N\to\infty} \psi\left(d_k + \sum_{r=1}^{N} \frac{\gamma-2}{\gamma^{k+r}}\right) \\
&= \lim_{N\to\infty} \left(\psi(d_k) + \sum_{r=k+1}^{N} (\gamma - 2) \cdot 2^{-m_r} \gamma^{-\frac{n^{r-m_r}-1}{n-1}}\right) \\
&= \psi(d_k) + \lim_{N\to\infty} \sum_{r=k+1}^{N} (\gamma-2)\gamma^{-\frac{n^r-1}{n-1}} \\
&= \psi(d_k) + (\gamma - 2)\varepsilon_{k,2}.
\end{aligned}$$

Corollary 1. For each integer $k \in N$ the pairwise disjoint intervals

$$E_k(d_k) = [d_k, d_k + \delta_k], \; d_k \in \mathscr{D} \tag{18}$$

are mapped by ψ into the pairwise disjoint image intervals

$$H_k(d_k) = [\psi(d_k), \psi(d_k) + (\gamma - 2)\varepsilon_{k,2}]. \tag{19}$$

That the intervals $E_k(d_k)$ are pairwise disjoint follows directly from their definition, and the corollary follows from Lemma 2 with a direct calcualtion.

Lemma 4. *For each fixed integer $k \in N$, the pairwise disjoint cubes*

$$S_k(d_{k,1}, \ldots, d_{k,n}) = \prod_{p=1}^{n} E_k(d_{k,p}), \quad d_{k,p} \in \mathscr{D}, \tag{20}$$

in $\mathscr{E}^n$ are mapped by $\Sigma \alpha_p \psi(d_{k,p})$ into the pairwise disjoint intervals

$$T_k(d_{k,1}, \ldots, d_{k,n}) = \left[\sum_{p=1}^{n} \alpha_p \psi(d_{k,p}), \right.$$
$$\left. \sum_{p=1}^{n} \alpha_p \psi(d_{k,p}) + \left(\sum_{p=1}^{n} \alpha_p\right)(\gamma - 2)\varepsilon_{k,2}\right]. \tag{21}$$

Proof of Lemma 4. To establish this lemma, we need to demonstrate that

$$\left(\sum_{p=1}^{n} \alpha_p\right)(\gamma - 2)\varepsilon_{k,2} \le |\mu_k|$$

for each integer $k \in N$. We begin by showing that

$$(\gamma - 2)\sum_{p=1}^{n} \alpha_p < \gamma - 1.$$

To this end, we note that because the series α_p converge absolutely we may arbitrarily rearrange terms in the summation below, and we find that

$$\sum_{p=1}^{n} \alpha_p = 1 + \sum_{p=2}^{n} \sum_{r=1}^{\infty} \gamma^{-(p-1)\frac{n^r-1}{n-1}} = 1 + \sum_{r=1}^{\infty} \sum_{p=2}^{n} \gamma^{-(p-1)\frac{n^r-1}{n-1}}$$
$$< 1 + \frac{1}{\gamma - 1} + \frac{1}{\gamma^{n+1} - 1} + \frac{1}{\gamma^{n^2+n+1} - 1} + \cdots$$
$$\le 1 + \frac{1}{\gamma - 1} + \frac{1}{\gamma^3 - 1} + \frac{1}{\gamma^7 - 1} + \cdots$$
$$< 1 + \frac{1}{\gamma - 1} + \frac{2}{\gamma^3 - 1}$$

and a simple estimate now shows that

$$(\gamma - 2)\sum_{p=1}^{n} \alpha_p < (\gamma - 2)\left(1 + \frac{1}{\gamma - 1} + \frac{2}{\gamma^3 - 1}\right) < \gamma - 1.$$

Hence, according to Lemma 2:

$$(\gamma - 2)\left(\sum_{p=1}^{n} \alpha_p\right)\varepsilon_{k,2} < (\gamma - 1)\varepsilon_{k,2} = (\gamma - 1)\sum_{r=k+1}^{N} \gamma^{-\frac{n^r-1}{n-1}}$$
$$< \gamma^{-\frac{n^{k+1}-1}{n-1}} - 1 = \gamma^{-n\frac{n^k-1}{n-1}} \le |\mu_k|$$

and this concludes the proof of the lemma.

Now let

$$E_k^q(d_{k,p}) = [d_{k,p}^q - qa, d_{k,p}^q + \delta_k - qa],$$

where

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

set

$$S_k^q(d_{k,1}^q, \ldots, d_{k,n}^q) = \prod_{p=1}^{n} E_k^q(d_{k,p}^q)$$

and

$$T_k^q(d_{k,1}^q, \ldots, d_{k,n}^q) = \left[\sum_{p=1}^{n} \alpha_p \psi(d_{k,p}^q), \right.$$
$$\left. \sum_{p=1}^{n} \alpha_p \psi(d_{k,p}^q) + \left(\sum_{p=1}^{n} \alpha_p\right)(\gamma - 2)\varepsilon_{k,2}\right].$$

It is readily verified that for each q and k,

$$S_k^q(d_{k,1}^q, \ldots, d_{k,n}^q) \cap S_k^q(d_{k,1}'^q, \ldots, d_{k,n}'^q) = \varnothing$$

and

$$T_k^q(d_{k,1}^q, \ldots, d_{k,n}^q) \cap T_k^q(d_{k,1}'^q, \ldots, d_{k,n}'^q) = \varnothing$$

whenever

$$(d_{k,1}^q, \ldots, d_{k,n}^q) \ne (d_{k,1}'^q, \ldots, d_{k,n}'^q),$$

and also that for any point $x \in \mathscr{E}^n$ and integer k there are not less than $n+1$ values of q for which $x \in S_k^q(d_{k,1}^q, \ldots, d_{k,n}^q)$. With this, the proof of Theorem 1 can be completed by applying Sprecher (1965), Nees (1993), or Katsuura & Sprecher (1994) to the present constructions. The necessary modifications and supplementary arguments are self-evident and the detail of this is omitted here.

REFERENCES

Hecht-Nielsen, R. (1987a). Counter propagation networks. *Proceedings IEEE International Conference On Neural Networks II* (pp. 19–32). New York: IEEE Press.

Hecht-Nielsen, R. (1987b). Kolmogorov's mapping neural network existence theorem. *Proceedings IEEE International Conference On Neural Networks* (pp. 11–13). New York: IEEE Press.

Katsuura, H., & Sprecher, D. A. (1994). Computational aspects of Kolmogorov's superposition theorem. *Neural Networks*, **7**(3), 455–461.

Kolmogorov, A. N. (1957). On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Doklady Akademii Nauk SSSR*, **114**, 953–956. *Translations American Mathematical Society* **2**(28) (1963), 55–59.

Kůrková, V. (1992). Kolmogorov's theorem and multilayer neural networks. *Neural Networks*, **5**, 501–506.

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

Nees, M. (1993). Approximate versions of Kolmogorov's superposition theorem, proved constructively. *Journal Computer Applied Math.* (in press).

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

Sprecher, D. A. (1993). A universal mapping for Kolmogorov's superposition theorem. *Neural Networks*, 6, 1089–1094.

NOMENCLATURE

$\mathscr{R}$	real line
$\mathscr{E}$	unit interval [0, 1]
$\mathscr{E}^n$	n-dimensional unit cube
N	set of natural numbers
$\mathscr{D}$	set of terminating rational numbers

$$d_k = \sum_{r=1}^{k} \frac{i_r}{\gamma^r},\ 0 \leq i_r \leq \gamma - 1,\ k \in N$$

$\langle i_r \rangle$ $\quad \langle i_1 \rangle = 0$ and $\langle i_r \rangle =$

$$\begin{cases} 0 & \textit{when } i_r = 0,\ 1, \ldots, \gamma - 2 \\ 1 & \textit{when } i_r = \gamma - 1 \end{cases} \quad \text{for } r \geq 2$$

$[i_r]$ $\quad [i_1] = 0$ and $[i_r] =$

$$\begin{cases} i_r & \textit{when } i_r = 0,\ 1, \ldots, \gamma - 3 \\ 1 & \textit{when } i_r = \gamma - 2,\ \gamma - 1 \end{cases} \quad \text{for } r \geq 2$$

m_r $\quad m_r = \langle i_r \rangle \left(1 + \sum_{s=1}^{r-1} [i_s] \times \cdots \times [i_{r-1}]\right)$