Pergamon

Neural Networks, Vol. 10, No. 3, pp. 447–457, 1997

Printed in Great Britain
0893-6080/97 \$17.00+.00

PII: S0893-6080(96)00073-1

CONTRIBUTED ARTICLE

A Numerical Implementation of Kolmogorov's Superpositions II

DAVID A. SPRECHER

University of California, Santa Barbara

(*Received* 18 *December* 1995; *revised and accepted* 24 *June* 1996)

Abstract—*This paper presents a numerical algorithm for the parallel computations* Φ_q *in the Kolmogorov superpositions* $f(\mathbf{x}) = \sum_{q=0}^{m} \Phi_q \circ \xi(\mathbf{x}_q)$, $\mathbf{x} = (x_1, \ldots, x_n)$ *and* $\mathbf{x}_q = (x_1 + qa, \ldots, x_n + qa)$, *thereby providing the final step in their numerical implementation. The first step consisting of the f-independent computation of the functions* $\xi(\mathbf{x}_q) = \sum_{p=1}^{n} \alpha_p \psi(x_p + qa)$ *with a fixed* ψ *and constants a and* α_p *in a hidden layer in the Hecht-Nielsen feedforward neural network has been accomplished previously. The step taken in this paper is the implementation of the output layer of the network that computes an arbitrary known continuous real-valued function f defined on the unit cube* $\mathcal{E}^n$. *Employed for the purpose is an iterative method which is intended as a basis for the possible development of adaptive methods that build on this approach. Each function* Φ_q *is obtained iteratively through a series* $\sum_r \Phi_q^r$ *which is determined on an f and q dependent subsequence* $\mathbf{d}_{k_1}^q, \mathbf{d}_{k_2}^q, \mathbf{d}_{k_3}^q \ldots$ *of rational coordinates* $\mathbf{d}_{k_r}^q = (d_{k_r1}^q, \ldots, d_{k_rn}^q)$ *such that* Φ_q^r *is determined at the coordinate points* $\xi(\mathbf{d}_{k_r}^q)$. *The paper also includes alternative constructions of the functions* Φ_q^r *and a brief discussion of the differentiability of* $\Phi_q \circ \xi(\mathbf{x}_q)$; *together with a previous result it gives a constructive proof of Kolmogorov's theorem.* © *1997 Elsevier Science Ltd. All Rights Reserved.*

Keywords—Superpositions, Kolmogorov, Computation of multivariate functions, Approximation of multivariate functions, Representation of multivariate functions, Continuity.

1. INTRODUCTION

Let $n \geq 2$, $m \geq 2n$ and $\gamma \geq m+2$ be given integers; let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{x}_q = (x_1 + qa, \ldots, x_n + qa)$, where $a = [\gamma(\gamma-1)]^{-1}$. This paper presents the numerical implementation of the functions Φ_q in Sprecher's (1996a) version of the Kolmogorov (1957) superpositions:

$$\begin{cases} f(\mathbf{x}) = \sum_{q=0}^{m} \Phi_q \circ \xi(\mathbf{x}_q) \\ \xi(\mathbf{x}_q) = \sum_{p=1}^{n} \alpha_p \psi(x_p + qa) \end{cases} \tag{1}$$

with fixed transfer functions $\xi(\mathbf{x}_q)$ in which ψ is monotonic increasing, $\alpha_1 = 1$, $\alpha_p = \sum_{r=1}^{\infty} \gamma^{-(p-1)\beta(r)}$ for $p > 1$, and $\beta(r) = (n^r - 1)/(n-1)$. Implementing the function ψ and its affine translates and linear combinations $\xi(\mathbf{x}_q)$ independently of f is the first step, accomplished previously (Sprecher, 1996a), where a suitable function ψ is defined pointwise on the set of rational numbers $d_k = \sum_{r=1}^{k} i_r \gamma^{-r}$, $i_r = 0, 1, \ldots, \gamma - 1$ and $k = 1, 2, 3, \ldots$. We observe that γ depends on n, as do in turn also the function ψ and constants α_p, and from the arguments given previously (Sprecher, 1965) we can deduce that also the above ψ belongs to class $Lip(ln\,2/ln\,\gamma)$. By sacrificing Lipschitz continuity the dependence on n of ψ and constants α_p can be eliminated (Sprecher, 1993; Katsuura & Sprecher, 1994; Sprecher, 1996b). This paper deals with the second step — the implementation of $m+1$ parallel functions Φ_q that compute an arbitrary continuous real-valued function $f\colon \mathcal{E}^n \to \mathcal{R}$ defined on the unit cube $\mathcal{E}^n$ in n-dimensional Euclidean space. In the context of the feedforward neural network of Hecht-Nielsen (1987), this paper deals with forming the output layer of the network. The iterative method which we employ for this is intended as a basis for the possible development of adaptive methods that build on this approach. The functions Φ_q in eqn (1) are constructed iteratively as functions $\Phi_q(y_q)$ of the single variable y_q with a numerical algorithm which produces for each q a series of functions $\Phi_q^j(y_q)$ such that $lim_{r\to\infty} \sum_{j=1}^{r} \Phi_q^j(y_q) = \Phi_q(y_q)$. Each function $\Phi_q^j(y_q)$ is determined by f at the points $y_q = \xi_q(\mathbf{d}_{k_r}^q)$ of a subsequence $\{\mathbf{d}_{k_1}^q, \mathbf{d}_{k_2}^q, \mathbf{d}_{k_3}^q, \ldots\}$ of coordinate points $\mathbf{d}_{k_r}^q = (d_{k_r,1}^q, \ldots, d_{k_r,n}^q)$ (see Figure 3 below);

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.

the rational numbers $d^q_{k_r p} = d_{k_r p} + q\sum_{s=2}^{k_r}\gamma^{-s}$, $p = 1,2,\ldots,n$, are connected with the expansion $a = [\gamma(\gamma-1)]^{-1} = \gamma^{-2}+\gamma^{-3}+\gamma^{-4}+\cdots$.

While the target function f is assumed to be completely known in the constructions of this paper, it is already shown by Kůrková (1992) that eqn (1) can be used to approximate an arbitrary function f (see also below). The use of $m+1$ instead of $2n+1$ summands in eqn (1) follows the observation in the same paper that increased accuracy (rate of convergence) for the approximation $f(\mathbf{x}) \approx \sum_{q=0}^{m}\sum_{j=0}^{r}\Phi_q^j \circ \xi(\mathbf{x}_q)$ for a given number r of iterations is possible by increasing the number of functions Φ_q from its minimum number of $2n+1$ to $m+1 > 2n+1$. This can give a computational advantage when an objective is accuracy as a measure of speed of computation since non-linear iterations are replaced with parallel computations. The implementable numerical algorithm that we develop gives several alternatives for the approximation of f, leading to eqn (1), and these are discussed in Section 5. In Section 6 we touch on the question of differentiability of $\Phi_q \circ \xi(\mathbf{x}_q)$.

On the mathematical side, this paper and Sprecher (1996a) provide a complete constructive proof of Kolmogorov's theorem.

2. COMPUTATIONS OF ψ

Referring to Sprecher (1996a), let $\langle i_1\rangle = [i_1] = 0$, and for $r > 1$ let

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

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

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

$$\text{for } r = 1,2,\ldots,k$$

$$\beta(r) = \frac{n^r-1}{n-1}.$$

Then for $k = 1,2,3,\ldots,$

$$\psi(d_k) = \sum_{r=1}^{k}\tilde{i}_r 2^{-m_r}\gamma^{-\beta(r-m_r)} \qquad (2)$$

where $d_k = \sum_{r=1}^{k} i_r\gamma^{-r}$ and $\tilde{i}_r = i_r - (\gamma-2)\langle i_r\rangle$. This uniquely determines a continuous function ψ: $\mathcal{E}\to\mathcal{E}$ that is extended beyond the unit interval $\mathcal{E}$ through the definition $\psi(x+1) = \psi(x)+1$. From the defining eqn (2) we derive the following simplified procedure for finding the values $\psi(d_k)$: We note that $\psi(d_1) = i_1\gamma^{-1}$, and when $i_r \le \gamma-2$ for $r > 1$ then $\langle i_r\rangle = 0$, so that $\tilde{i}_r = i_r$, $m_r = 0$. Consequently

$$\psi(d_k) = \sum_{r=1}^{k} i_r\gamma^{-\beta(r)}$$

when no digits $i_r = \gamma-1$ for $r > 1$ are present in the sequence $(i_1,i_2,\ldots,i_k)$ defining the rational number d_k. When digits $i_r = \gamma-1$ for $r > 1$ are present in such a sequence, we compute $\psi(d_k)$ by applying appropriate cases of the basic pattern of consecutive clusters of digits i_r

$$(u_1 \textit{ digits } i_r = \gamma-2)(u_2 \textit{ digits } i_r = \gamma-1)$$
$$\times (u_3 \textit{ digits } i_r = \gamma-2)(u_4 \textit{ digits } i_r = \gamma-1)$$

according to the following rule. Let i_u be a digit such that $i_u \le \gamma-1$ if $u = 1$, or $i_u \le \gamma-3$ if $u > 1$. Let $k = u+u_1+u_2+u_3+u_4$, then with the convention $\sum_{r=1}^{0} \equiv 0$ we have

$$\begin{aligned}\psi(d_k) = {} & \psi(d_u) + (\gamma-2)\sum_{r=1}^{u_1}\gamma^{-\beta(u+r)} \\ & + \gamma^{-\beta(u)}\sum_{r=1}^{u_2}2^{-u_1-r} + (\gamma-2)\sum_{r=1}^{u_3}\gamma^{-\beta(u+u_1+u_2+r)} \\ & + \gamma^{-\beta(u)}\sum_{r=1}^{u_4}2^{-u_1-u_2-u_3-r}.\end{aligned}$$

Table 2 of Sprecher (1996a) gives the 10,000 values of $\psi(d_k)$ corresponding to $k = 4$. The rational numbers $d_k^q = d_k + q\sum_{r=2}^{k}\gamma^{-r}$ can be used to simplify the computations of the translates $\psi(d_k+qa)$ as follows: Let

$$\epsilon_k = \sum_{r=k+1}^{\infty}\gamma^{-\beta(r)}.$$

Because $\gamma-2 \ge m \ge q$ we have from Lemma 1 of Sprecher (1996a)

$$\psi(d_k+qa) = \psi(d_k^q) + q\epsilon_k$$

and if we set

$$b_k = \epsilon_k\sum_{p=1}^{n}\alpha_p$$

then

$$\begin{aligned}\xi(\mathbf{d}_k+q\mathbf{a}) &= \sum_{p=1}^{n}\alpha_p\psi(d_{kp}+qa) \\ &= \sum_{p=1}^{n}\alpha_p\psi(d_{kp}^q) + qb_k = \xi(\mathbf{d}_k^q) + qb_k.\end{aligned}$$

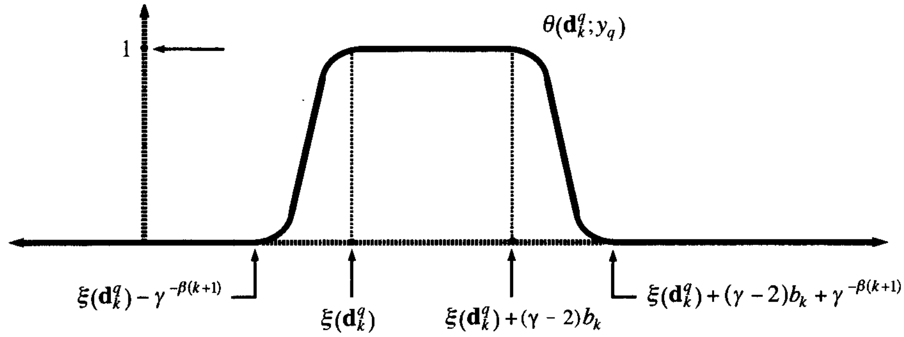

FIGURE 1. The function $\theta(d_k^q; y_q)$.

Clearly $lim_{k\to\infty} b_k = 0$ and given functions Φ_q, we have the following approximate formula for rational numbers:

$$f(\mathbf{d}_k) = \sum_{q=0}^{m} \Phi_q \circ [\xi(\mathbf{d}_k^q) + qb_k] \approx \sum_{q=0}^{m} \Phi_q \circ \xi(\mathbf{d}_k^q).$$

3. THE IMPLEMENTATION OF Φ_q

The functions Φ_q are implementable in a number of ways, of which we present here one, using only the constructions that are necessary for the purpose. Additional detail and justification are contained in Section 4, and alternative implementations are given in Section 5.

DEFINITION 1

Let $\sigma: \mathcal{R} \to \mathcal{E}$ be an arbitrary continuous function with $\sigma(x) \equiv 0$ when $x \leq 0$, and $\sigma(x) \equiv 1$ when $x \geq 1$. For each number $\xi(\mathbf{d}_k^q)$, $\mathbf{d}_k^q = (d_{k,1}^q, \ldots, d_{k,n}^q)$,

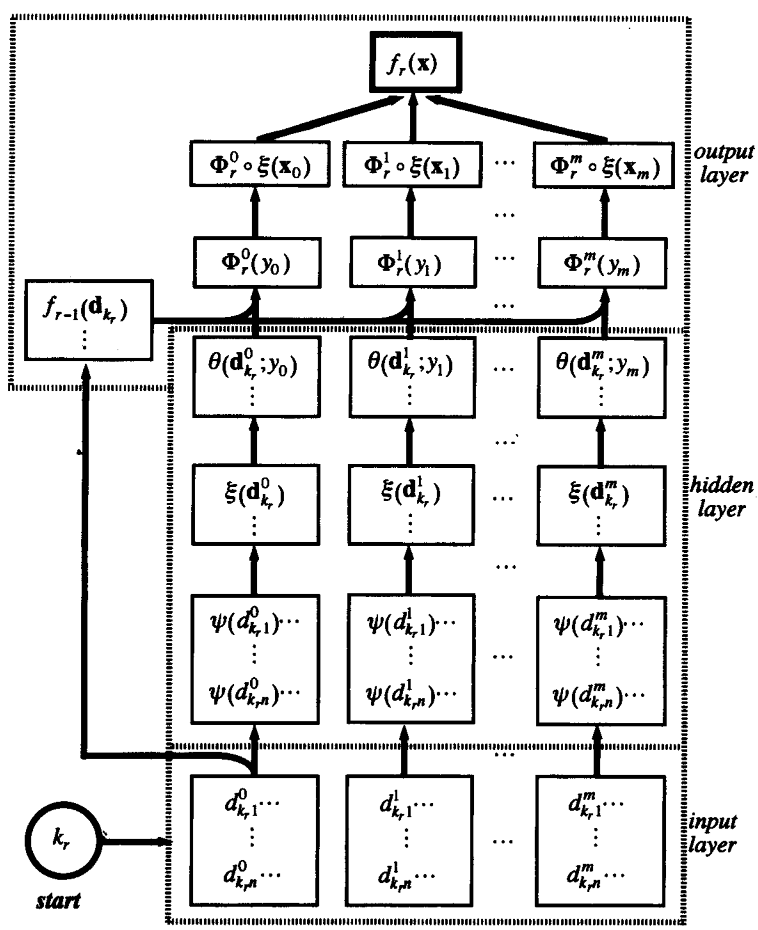

FIGURE 2. Schematic representation of the implementation algorithm.

we set

$$\theta(\mathbf{d}_k^q; y_q) = \sigma(\gamma^{\beta(k+1)}(y_q - \xi(\mathbf{d}_k^q)) + 1)$$
$$- \sigma(\gamma^{\beta(k+1)}(y_q - \xi(\mathbf{d}_k^q) - (\gamma - 2)b_k)).$$

Clearly $0 \le \theta(\mathbf{d}_k^q; y_q) \le 1$ and $\theta(\mathbf{d}_k^q; y_q) = 1$ for $\xi(\mathbf{d}_k^q) \le y_q \le \xi(\mathbf{d}_k^q) + (\gamma - 2)b_k$ (see Figure 1). We remark that the numbers $\gamma^{\beta(k+1)}$ appearing in this definition are not uniquely determined. They can be replaced for each k by any number $0 < \Gamma_k \le \gamma^{\beta(k+1)}$.

NOTE

For the remainder of this paper, let $f: \mathcal{E}^n \to \mathcal{R}$ be a given continuous function with known uniform norm $\|f\|$, and η and ϵ numbers such that $0 < ((m - n + 1)/(m + 1))\epsilon + (2n/(m + 1)) \le \eta < 1$. This implies that $\epsilon < 1 - (n/(m - n + 1)$ and fixing η and ϵ in this way is sufficient to guarantee the convergence of the following algorithm:

The Implementation Algorithm

Starting with $f_0 \equiv f$, iterate the following steps for $r = 1, 2, 3, \ldots$ (consult Figure 2).

I. **Input layer**
Given the function $f_{r-1}(\mathbf{x})$, determine an integer k_r such that $|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{x}')| \le \epsilon \|f_{r-1}\|$ when $|x_p - x'_p| \le \gamma^{-k_r}$ for $p = 1, \ldots, n$. This determines the rational coordinate points $\mathbf{d}_{k_r}^q = (d_{k_r1}^q, \ldots, d_{k_rn}^q)$, where

$$d_{k_rp}^q = d_{k_rp} + q \sum_{r=2}^{k_r} \gamma^{-r}.$$

II. **Hidden layer**
For $q = 0, 1, \ldots, m$:
II-1 Compile the values $\psi(d_{k_r}^q)$
II-2 Compute the linear combinations $\xi(\mathbf{d}_{k_r}^q) = \sum_{p=1}^n \alpha_p \psi(\mathbf{d}_{k_r}^q)$
II-3 Compute the one-variable functions $\theta(\mathbf{d}_{k_r}^q; y_q)$.

III. **Output layer**
III-1 Compute the one variable functions

$$\Phi_q^r(y_q) = \frac{1}{m+1} \sum_{\mathbf{d}_{k_r}^q} f_{r-1}(\mathbf{d}_{k_r})\theta(\mathbf{d}_{k_r}^q; y_q),$$
$$q = 0, 1, \ldots, m.$$

III-2 Substitute the transfer functions $\xi(\mathbf{x}_q)$ to compute the multi-variable functions

$$\Phi_q^r \circ \xi(\mathbf{x}_q) = \frac{1}{m+1} \sum_{\mathbf{d}_{k_r}^q} f_{r-1}(\mathbf{d}_{k_r})\theta(\mathbf{d}_{k_r}^q; \xi(\mathbf{x}_q)),$$
$$q = 0, 1, \ldots, m.$$

III-3 Compute the function

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

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. ■

To each r there correspond γ^{k_r} rational numbers d_{k_r} from which we compute in the input layer $m + 1$ lists each of $n \times \gamma^{k_r}$ rational numbers $d_{k_r}^q$. As seen in Figure 2, we compile from these in the hidden layer $m + 1$ lists each of $n \times \gamma^{k_r}$ entries $\psi(d_{k_rp}^q)$, and from these we compile, in turn, $m + 1$ tables each with $\gamma^{n \times k_r}$ entries $\xi(\mathbf{d}_{k_r}^q)$ and $m + 1$ tables each with $\gamma^{n \times k_r}$ entries $\theta(\mathbf{d}_{k_r}^q; y_q)$. We note in passing that the sets $\{d_{k_r}^q\}$ of rational numbers are not mutually exclusive for fixed k_r and variable q, $q = 0, 1, \ldots, m$, and neither are the corresponding lists of compiled values $\{\psi(d_{k_r}^q)\}$. In the output layer we use these tables together with the tables of $\gamma^{n \times k_r}$ values $f_{r-1}(\mathbf{d}_{k_r})$ to compute the output functions Φ_q^r and $f_r(\mathbf{x})$. The prescribed iteration loops producing these tables are possible because the functions $f_0 = f, f_1, f_2, \ldots$ and $\Phi_q^r(y_q)$ are continuous, and formulae III-2 and III-3 can be obtained with direct computations since the function ψ and as well as all constants are given numerically. Section 5 includes further comments concerning the implementation of the functions $\Phi_q^r(y_q)$.

4. MATHEMATICAL ARGUMENTS AND PROOFS

The central theorem assuring that the implementation algorithm produces functions which convergence to produce eqn (1) is Theorem 1 below. The proof of this theorem relies on certain properties of the functions defined earlier, and we begin by establishing these. The first is noting that the *support* of each function $\theta(\mathbf{d}_k^q; y_q)$ is the open interval

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

i.e., $\theta(\mathbf{d}_k^q; y_q) \equiv 0$ when $y_q \notin U_k^q(\mathbf{d}_k^q)$, as depicted in Figure 1 (see, however, the remark following Definition 1). These intervals are pairwise disjoint for fixed q

and k: If $\xi(\mathbf{d}_k^q) \neq \xi(\mathbf{d}_k'^q)$ then $U_k^q(\mathbf{d}_k^q) \cap U_k^q(\mathbf{d}_k'^q) = \varnothing$ (Lemma 4 in Sprecher, 1996a). From this fact we derive the following estimate for formula III-1:

LEMMA 1
For each value of q and r,

$$\left\|\Phi_q^r(y_q)\right\| \leq \frac{1}{m+1}\left\|f_{r-1}\right\|.$$

Proof
In view of the observations following Definition 1 and the above,

$$\left\|\frac{1}{m+1}\sum_{\mathbf{d}_{k_r}^q} f_{r-1}(\mathbf{d}_{k_r})\theta(\mathbf{d}_{k_r}^q; y_q)\right\|.$$

$$= \frac{1}{m+1}\max_{\mathbf{d}_{k_r}} |f_{r-1}(\mathbf{d}_{k_r})|.$$

and the lemma follows. ■

We conclude that a typical graph of $\Phi_q^r(y_q)$ is as depicted in Figure 3.

The functions f_{r-1} and Φ_q^r as implemented above are connected through the following essential inequality:

THEOREM 1
For $r = 1, 2, 3, \ldots$ we have the inequalities

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

Iterating the inequalities in Lemma 1 and Theorem 1 give at once

COROLLARY 1
For $j = 1, 2, 3, \ldots,$

$$\|\Phi_q^j(y_q)\| \leq \frac{1}{m+1}\eta^{j-1}\|f\| \quad (3)$$

$$\left\|f(\mathbf{x}) - \sum_{q=0}^{m}\sum_{j=1}^{r}\Phi_q^j \circ \xi(\mathbf{x}_q)\right\| < \eta^r\|f|. \quad (4)$$

From this corollary it follows that

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

$$< \frac{1}{m+1}\|f\|\sum_{j=0}^{\infty}\eta^j < \infty. \quad (5)$$

Accordingly, each series of functions $\sum_{j=1}^{r}\Phi_q^j(y_q)$ converges absolutely for each value of q to a continuous function $\Phi_q(y_q)$ as $r \to \infty$; eqn (1) follows from the fact that $\eta^r|f\| \to 0$ as $r \to \infty$. Expressed directly in terms of the functions $\theta(\mathbf{d}_{k_r}^q; \xi(\mathbf{x}_q))$ (and hence the function σ) and the numbers $f(\mathbf{d}_{k_r})$, estimate (4) gives the approximate equality

$$f(\mathbf{x}) \approx \sum_{q=0}^{m}\sum_{j=1}^{r}\sum_{\mathbf{d}_{k_r}^q} f(\mathbf{d}_{k_j})\theta(\mathbf{d}_{k_j}^q; \xi(\mathbf{x}_q))$$

(see in this connection Kůrková, 1992).

Proof of theorem 1
A version of the proof of this theorem can be found in Lorentz (1966). Presented here is a proof that relates to the specific constructions and notation of this paper. To simplify the arguments, we include now the value $d_k = 1$ in the definition of the rational numbers d_k. The proof is based on the following additional

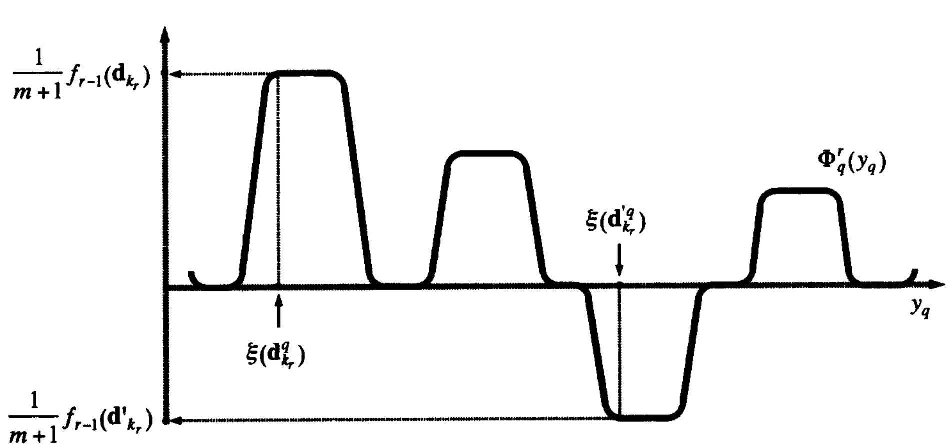

FIGURE 3. The function $\Phi_q^r(y_q)$.

property of the functions ψ and $\xi(\mathbf{x}_q)$ (see Sprecher, 1996a): Consider for each integer q the family of closed intervals

$$E_k^q(d_k^q) = \left[-\frac{q}{\gamma(\gamma-1)} + d_k^q, -\frac{q}{\gamma(\gamma-1)} + \frac{\gamma-2}{\gamma-1}\gamma^{-k} + d_k^q\right] = \left[-\frac{q}{\gamma-1}\gamma^{-k} + d_k, \frac{\gamma-2-q}{\gamma-1}\gamma^{-k} + d_k\right]. \tag{6}$$

It is easily verified that for fixed q and k these intervals are separated by gaps (open intervals) of width $(\gamma-1)^{-1}\gamma^{-k}$. From these intervals we obtain for each k the $m+1$ families of closed (Cartesian product) cubes

$$S_k^q(\mathbf{d}_k^q) = E_k^q(d_{k,1}^q) \times \cdots \times E_k^q(d_{k,n}^q), \quad q = 0, 1, \ldots, m.$$

whose images under $\xi(\mathbf{x}_q)$ are the closed intervals

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

A direct calculation shows that

$$\text{if } \mathbf{x} \in S_k^q(\mathbf{d}_k^q) \text{ then } \xi(\mathbf{x}_q) \in T_k^q(\mathbf{d}_k^q). \tag{7}$$

Let k be fixed. To gain insight into the mechanics of superpositions and the proof, let us examine the effect of the mapping $\xi(\mathbf{x}_q)$ for a given value of q on a single cube $S_k^q(\mathbf{d}_k^q)$. As a first approximation, the surface $y_q = \xi(\mathbf{x}_q)$ for $\mathbf{x} \in S_k^q(\mathbf{d}_k^q)$ can be taken to be a tilted plane as in Figure 4, with its lower left hand corner at the point $(\mathbf{d}_k^q, \xi(\mathbf{d}_k^q))$. Intuitively, the mapping $\xi(\mathbf{x}_q)$ acts like a cookie-cutter, removing $S_k^q(\mathbf{d}_k^q)$ from the coordinate space and giving it a unique image $T_k^q(\mathbf{d}_k^q)$ on the y_q-axis. The images of any two cubes have empty intersections for fixed q (and k) as $\mathbf{d}_k^q$ varies over its domain, and this property enables the local approximation of the target function $f(\mathbf{x})$ on the surfaces $y_q = \xi(\mathbf{x}_q)$ for $\mathbf{x} \in S_k^q(\mathbf{d}_k^q)$. Loosely speaking, we can view $\xi(\mathbf{x}_q)$ as replacing the coordinate system with $\gamma^{n \times k}$ local coordinate systems (surfaces) corresponding to the $\gamma^{n \times k}$ cubes $S_k^q(\mathbf{d}_k^q)$. Continuity requires these surfaces to be separated by gaps, however narrow, in which $f(\mathbf{x})$ cannot be approximated, and it is there that each stage of the implementation iteration introduces q-dependent errors that cannot be made arbitrarily small with $\xi(\mathbf{x}_q)$. Kolmogorov's ingenuity comes to bear on the problem most profoundly here by introducing parallel schemes through the affine translations of the families of cubes $S_k^q(\mathbf{d}_k^q)$ such that each family intersects the gaps of the other families in a prescribed manner.

Returning to eqn (6), we note that the gaps separating the intervals do not intersect for fixed k and variable q. Therefore, any point $x \in \mathcal{E}$ can be contained in at most one gap and consequently can be excluded from at most one of the $m+1$ intervals $E_k^q(d_k^q)$, and so must be contained in at least m of them. Thus, if $\mathbf{x} \in \mathcal{E}^n$ is an arbitrary point, then we deduce that there are at least $m-n+1$ values of q for which $\mathbf{x} \in S_k^q(\mathbf{d}_k^q)$. We see at once from eqn (6) that $d_k \in E_k^q(d_k^q)$ for each q, and consequently $\mathbf{d}_k \in \bigcap_{q=0}^m S_k^q(\mathbf{d}_k^q)$. It therefore follows that there are at least $m-n+1$ cubes containing both $\mathbf{x}$ and some grid-points $\mathbf{d}_k$.

Now let k_r be a given integer for which Step I of the Implementation Algorithm holds; let $x \in \mathcal{E}^n$ be an arbitrary point. Let $q_j, j = 1, \ldots, m-n+1$, be values for which $\mathbf{x} \in S_{k_r}^{q_j}(\mathbf{d}_{k_r}^{q_j})$. For the point $\mathbf{d}_{k_r} \in S_{k_r}^{q_j}(\mathbf{d}_{k_r}^{q_j})$ we have

$$|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{d}_{k_r})| \leq \epsilon \|f_{r-1}\| \tag{8}$$

and in view of eqn (7), $\xi(\mathbf{x}_{q_j}) \in T_{k_r}^{q_j}(\mathbf{d}_{k_r}^{q_j})$ when

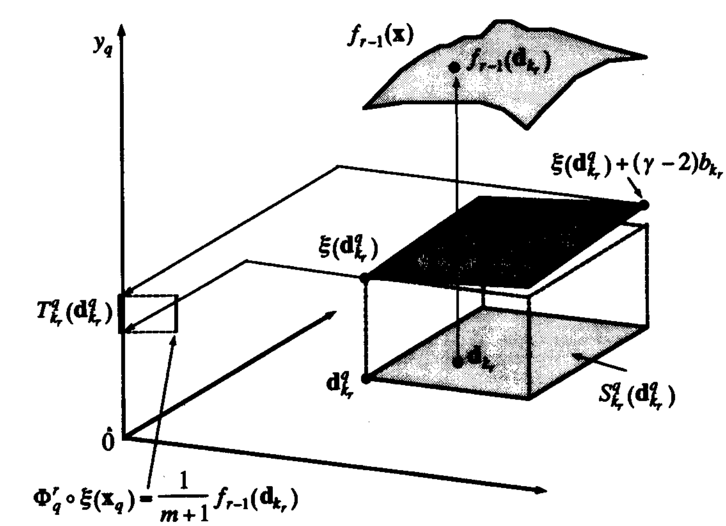

FIGURE 4. The mapping $\xi(\mathbf{x}_q)$.

$\mathbf{x} \in S^{q_j}_{k_r}(\mathbf{d}^{q_j}_{k_r})$, so that III-2 gives

$$\Phi^r_{q_j} \circ \xi(\mathbf{x}_{q_j}) = \frac{1}{m+1} f_{r-1}(\mathbf{d}_{k_r})\theta(\mathbf{d}^{q_j}_{k_r}; \xi(\mathbf{x}_{q_j})) = \frac{1}{m+1} f_{r-1}(\mathbf{d}_{k_r}).$$

Together with eqn (8) this shows that

$$\left|\frac{1}{m+1} f_{r-1}(\mathbf{x}) - \Phi^r_{q_j} \circ \xi(\mathbf{x}_{q_j})\right| \le \frac{1}{m+1} \cdot \epsilon \| f_{r-1} \| \quad (j = 1, \dots, m-n+1).$$

For the remaining values of q we have the estimate in Lemma 1. Thus,

$$\begin{aligned} |f_r(\mathbf{x})| &= \left| f_{r-1}(\mathbf{x}) - \sum_{q=0}^{m} \Phi^r_q \circ \xi(\mathbf{x}_q) \right| \\ &\le \left| \sum_{j=1}^{m-n+1} \left[\frac{1}{m+1} f_{r-1}(\mathbf{x}) - \Phi^r_{q_j} \circ \xi(\mathbf{x}_{q_j})\right] + \frac{n}{m+1} f_{r-1} \right| + \frac{n}{m+1} \| f_{r-1} \| \\ &\le \frac{m-n+1}{m+1} \epsilon \| f_{r-1} \| + \frac{2n}{m+1} \| f_{r-1} \| \\ &\le \eta \| f_{r-1} \| \end{aligned}$$

and the theorem follows. ■

5. NOTES ON THE FUNCTIONS $\Phi_q(y_q)$

i) From the perspective of computing, superpositions can be interpreted as a device which enables the computation of $f(\mathbf{x})$ through $m+1$ iterative parallel computations of which not less than $m-n+1$ approximate f at any point $\mathbf{x} \in \mathcal{E}^n$ using neighbouring points $\mathbf{d}_{k_r} \approx \mathbf{x}$, not necessarily all distinct, and at most n of these introduce an error as specified in eqn (6). The implementation of the functions $\Phi_q(y_q)$ as well as the proofs of convergence are predicated on the knowledge of f on an everywhere dense set in $\mathcal{E}^n$ as well as its uniform norm. For computational efficiency we standardized the sets of points at which the functions in the hidden layer are determined and evaluated, and the evaluation of f in the output layer is limited to the coordinate points $\mathbf{d}_{k_r}$. Prescribing the evaluation of f in this manner may not always be appropriate, however, since actual data arising in applications may not include the values $f(\mathbf{d}_{k_r})$, or these may not be efficiently computable. In such cases the implementation of the output layer can be modified to use arbitrary values $f(\mathbf{x}^q_{k_r})$ as long as $\mathbf{x}^q_{k_r} \in S^q_{k_r}(\mathbf{d}^q_{k_r})$. An examination of the proof of Theorem 1 shows that this produces the same end result. More generally we can, for example, replace the functions $\Phi^r_q(y_q)$ with functions

$$\bar{\Phi}^r_q(y_q) = \frac{1}{m+1}\left[c_r + \sum_{\mathbf{d}^q_{k_r}} \left[f_{r-1}(\mathbf{x}^q_{k_r}) - c_r\right]\theta(\mathbf{d}^q_{k_r}; y_q)\right]$$

where $c_r = \frac{1}{2}[max f_{r-1} + min f_{r-1}]$. These functions have a smaller oscillation than the corresponding functions Φ^r_q, and clearly $\|\bar{\Phi}^r_q\| = \|\Phi^r_q\|$. The functions $\theta(\mathbf{d}^q_{k_r}; y_q)$ are convenient because their supports are localized, and this may be of particular interest with dynamic target functions whose values, rather than being prescribed in advance, may change locally. The disadvantage of the functions $\theta(\mathbf{d}^q_{k_r}; y_q)$ lies in the fact that they may introduce large oscillations into each individual computation Φ^r_q even when we take advantage of functions of the type $\bar{\Phi}^r_q$. Any such oscillations, however, decrease with increasing r, and they are anyway diminished in the sum of the functions Φ^r_q over q.

ii) We note that the function ψ is strictly monotonic increasing, and that the functions $\xi(\mathbf{x}_q)$ map the n-dimensional cubes $S^q_{k_r}(\mathbf{d}^q_{k_r})$ onto the non-degenerate intervals $T^q_{k_r}(\mathbf{d}^q_{k_r})$ which are used to construct local approximations to $f(\mathbf{x})$. An alternative implementation algorithm can be developed by approximating each $\xi(\mathbf{x}_q)$ with a sequence $\{\xi^q_{k_r}(\mathbf{x}_q)\}$ of continuous functions with the property that for each coordinate point $\mathbf{d}^q_{k_r}$ the image of the cube $S^q_{k_r}(\mathbf{d}^q_{k_r})$ is the number $\xi(\mathbf{d}^q_{k_r})$ instead of the interval $T^q_{k_r}(\mathbf{d}^q_{k_r})$. In this case, rather than composing $\Phi^r_q(y_q)$ with the functions $\theta(\mathbf{d}^q_{k_r}; y_q)$, we can use for each q interpolating functions G^r_q obtained by passing arbitrary interpolating curves through the points $(\xi(\mathbf{d}^q_{k_r}), (1/(m+1))f_{r-1}(\mathbf{d}_{k_r}))$ or through the points $(\xi(\mathbf{d}^q_{k_r}), (1/(m+1))f_{r-1}(\mathbf{x}^q_{k_r}))$ for $\mathbf{x}^q_{k_r} \in S^q_{k_r}(\mathbf{d}^q_{k_r})$, subject only to the condition $\|G^r_q\| \le (1/(m+1))\|f_{r-1}\|$ (see III-2 and Lemma 1). This requires, however, that we know for each q the linear order of the points $\xi(\mathbf{d}^q_k)$ on the y_q-axis for $k = 1, 2, 3, \dots$. An alternative construction of functions Φ^r_q that also requires knowledge of this order can be found in Katsuura and Sprecher (1994). This linear order can be determined from the specific numerical construction of $\xi(\mathbf{d}^q_k)$. We observe in passing that for each value of q, the piecewise linear curves $\Omega^q_k : \xi(\mathbf{d}^q_k) \to \mathbf{d}^q_k$ joining the coordinate points $\mathbf{d}^q_k$ in the order induced by the function $\xi(\mathbf{x}_q)$ converge to a Peano (space-filling) curve as $k \to \infty$. To proceed in this way, we approximate the function ψ with a sequence $\{\psi_k\}$ of continuous monotonic *non-decreasing* functions with the property that the image of the interval $E_k(d_k)$ is the number $\psi(d_k)$. Toward this end we introduce continuous functions $\omega(d_k; x): \mathcal{R} \to \mathcal{E}$ defined as follows (see Figure 5):

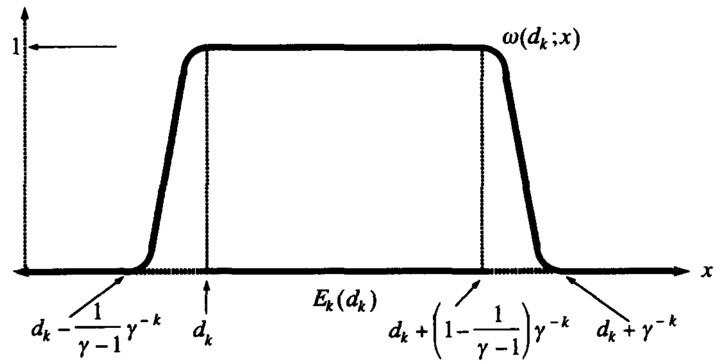

FIGURE 5. The function $\omega(d_k; x)$.

DEFINITION 2
For each rational number d_k and $k = 1, 2, 3, \ldots$,

$$\omega(d_k; x) = \sigma(\gamma^k(\gamma - 1)(x - d_k) + 1) - \sigma(\gamma^k(\gamma - 1)(x - d_k) - \gamma + 2).$$

We observe that $\sum_{d_k} \omega(d_k; x) \equiv 1$ and $\omega(d_k; d_k) = 1$ so that the functions

$$\psi_k(x) = \sum_{d_k} \psi(d_k)\omega(d_k; x)$$

are such that $\psi_k(d_k) = \psi(d_k)$. They have the property that $\psi_k(x) = \psi_k(d_k)$ when $x \in E_k(d_k)$ for each rational number d_k. The next step is to apply the functions $\psi_k(x)$ to the translations of ψ and their linear combinations $\xi(\mathbf{x}_q)$. Since $\gamma \geq 2n + 2 \geq q + 2$, it follows that $\omega(d_k^q; d_k^q) = 1$ and also that

$$\sum_{d_k^q} \omega(d_k^q; x + qa) \equiv 1.$$

for fixed q and k. Defining

$$\psi_k^q(x + qa) = \sum_{\mathbf{d}_k^q} \psi(d_k + qa)\omega(d_k^q; x + qa)$$

we have

LEMMA 2

$$\lim_{k\to\infty} \psi_k^q(x + qa) = \psi(x + qa).$$

Proof
Let k be given. By construction, if d_k and d_k' are consecutive rational numbers, then $|\psi(d_k + qa) - \psi(d_k' + qa)| \leq 2^{-k+1}\gamma^{-1}$ (Sprecher, 1996a). Now, if x is given, then for each k there are consecutive rational numbers $d_k < d_k'$ such that $d_k \leq x \leq d_k'$ and using the fact that $\psi_k^q(d_k + qa) = \psi(d_k + qa)$ for all rational numbers d_k we find that

$$\begin{aligned} &|\psi_k^q(x + qa) - \psi(x + qa)| \\ &\quad \leq |\psi_k^q(x + qa) - \psi_k^q(d_k + qa)| \\ &\qquad + |\psi(x + qa) - \psi(d_k + qa)| \\ &\quad \leq 2 \cdot 2^{-k+1}\gamma^{-1} \end{aligned}$$

and the lemma follows. ■

Now let

$$\psi_k^q(x_p + qa) = \sum_{d_{kp}^q} \psi(d_{k,p} + qa)\omega(d_{kp}^q; x_p + qa).$$

and define the functions

$$\xi_k^q(\mathbf{x}_q) = \sum_{p=1}^{n} \alpha_p \psi_k(x_p + qa).$$

We have

COROLLARY 2
For each value of q we have

$$\lim_{k\to\infty} \xi_k^q(\mathbf{x}_q) = \xi(\mathbf{x}_q).$$

We can now replace the functions $\Phi_q^r(y_q)$ in III-2 with interpolating functions $G_q^r(y_q)$ such that

$$G_q^r \circ \xi(\mathbf{d}_{k_r}^q) = \frac{1}{m+1} f_{r-1}(\mathbf{d}_{k_r})$$

and

$$\|G_q^r\| \leq \frac{1}{m+1} \|f_{r-1}\|$$

and we modify the implementation algorithm accordingly. Alternatively, we could replace the functions

$\theta(\mathbf{d}_k^q; y_q)$ in the implementation algorithm with the functions

$$\hat{\theta}(\mathbf{d}_k^q; y_q) = \sigma(\gamma^{\beta(k+1)}(y_q - \xi_k^q(\mathbf{d}_k^q)) + 1) - \sigma(\gamma^{\beta(k+1)}(y_q - \xi_k^q(\mathbf{d}_k^q)))$$

whose supports are the open intervals $(\xi_k^q(\mathbf{d}_k^q) - \gamma^{-\beta(k+1)}, \xi_k^q(\mathbf{d}_k^q) + \gamma^{-\beta(k+1)})$ (see, however, the remark following Definition 1). It is easy to see that Theorem 1 and Corollary 1 remain valid with these new constructions.

iii) Consider a function $f(\mathbf{x})$ given as an aggregate of values

$$(\mathbf{x}; f(\mathbf{x}))$$

and the functions $\Phi_0(y_0), \Phi_1(y_1), \ldots, \Phi_m(y_m)$ representing it also as aggregates of values

$$(y_0; \Phi_0(y_0)), (y_1; \Phi_1(y_1)), \ldots, (y_m; \Phi_m(y_m)).$$

These functions, related to f through eqn (1), give an alternative system for the listing of the values $y = f(\mathbf{x})$ in the following sense. A point a in the range of f determines the level-set $F_a = \{\mathbf{x}: f(\mathbf{x}) = a\}$ in the domain of f; that is, the different solutions $\mathbf{x}$ of the equation $f(\mathbf{x}) = a$ have the same image in the range of f, so that the inverse image of a does not distinguish between the points of its level-set. The transfer functions $\xi(\mathbf{x}_q)$, however, do distinguish between the sets $F_a \cap S_{k_r}^q(\mathbf{d}_{k_r}^q)$, so that different solutions of the equation $f(\mathbf{x}) = a$ may have different images in the ranges of $\Phi_q^r \circ \xi(\mathbf{x}_q)$ for each q and k_r. At the same time, however, the values of f corresponding to the solution set $\Xi_b = \{\mathbf{x}_q: \xi(\mathbf{x}_q) = b\}$ give the same image in the range of $\Phi_q^r \circ \xi(\mathbf{x}_q)$.

iv) A given target function $f(\mathbf{x})$ arising in applications may not be specified on an everywhere dense set, but interpolating functions can still yield useful results, even with incomplete data that may also not be distributed uniformly. Interpolating functions $G_q^r(y_q)$ of a single variable, however, do not necessarily provide useful interpolated values $f(\mathbf{x})$ because neighbouring points $\xi(\mathbf{d}_{k_r}^q)$ on the y_q-axis do not always correspond to neighbouring points $\mathbf{d}_{k_r}^q$ in $\mathcal{E}^n$.

The following specific question is posed by Hecht-Nielsen (1987):

> Suppose that a function $f: \mathcal{E}^n \to \mathcal{R}$ is specified through data that is mostly concentrated in an n-dimensional region $D^n \subset E^n$ and sparsely distributed in $E^n - D^n$. We wish to interpolate values $f(\mathbf{x})$ subject to certain prescribed conditions imposed on f: Let N designate the total number of given data points in the region D^n. Can interpolating functions $G_q^N(y_q)$ be constructed such that $G_q^N(y_q) \to \Phi_q(y_q)$ as $N \to \infty$ and for which:
> a) Equation (1) with the given transfer functions holds at the points at which f is defined.
> b) The functions $G_N(\mathbf{x}) = \sum_{q=0}^m G_q^N \circ \xi(\mathbf{x}_q)$ interpolate values of f subject to its prescribed properties.

The measure of the reliability of such interpolations can be a comparative one that relates to alternative (multi-variable) interpolation methods.

6. NOTES ON SUPERPOSITIONS WITH DIFFERENTIABLE $\Phi_q \circ \xi(\mathbf{x}_q)$

The function ψ is monotonic, and the Lebesgue theorem tells us that it is differentiable almost everywhere (except for a set of measure zero). Like the corresponding functions in other versions of Kolmogorov's superpositions, ψ is singular: $\psi'(x) = 0$ almost everywhere, and $\psi'(x) = +\infty$ or no derivative exists elsewhere (see below). In this section we make a first attempt to determine classes of functions f that are obtainable with differentiable functions $\Phi_q(y_q)$ and differentiable composite functions $\Phi_q \circ \xi(\mathbf{x}_q)$. The character of ψ tells us that these are distinct problems, since it is clear that for the composite function $\Phi_q \circ \xi(\mathbf{x}_q)$ to have partial derivatives there must exist a certain symmetry (complementary relationship) between the differentiability properties of $\Phi_q(y_q)$ and that of $\xi(\mathbf{x}_q)$, as determined by ψ. The complete differentiability profile of ψ is obtained as follows (Sprecher, 1966): The number

$$D\psi(x) = \lim_{h \to 0} \frac{\psi(x+h) - \psi(x)}{h}$$

is the *derived number of ψ at x* when the limit exists; we also use the notation $D\psi(x-)$ and $D\psi(x+)$ when the left or the right limits exist, respectively.

We now set

$$\delta_k = \frac{\gamma - 2}{\gamma - 1}\gamma^{-k} = (\gamma - 2)\sum_{r=k+1}^{\infty} \gamma^{-r}$$

[see the intervals in eqn (6) above] and have

LEMMA 3
For each rational number d_k,

$$D\psi(d_k-) = +\infty$$
$$D\psi(d_k+) = 0$$
$$D\psi((d_k + \delta_k)-) = 0$$
$$D\psi((d_k + \delta_k)+) = +\infty.$$

Consider the open intervals $\dot{E}_k(d_k) = (d_k, d_k + \delta_k)$

and the gaps $G_k(d_k) = (d_k + \delta_k, d_k + \gamma^{-k})$ separating them, and let

$$\mathcal{U} = \{d_k\} \cup \{d_k + \delta_k\}$$

$$\mathcal{V} = \left\{ x: x \in \bigcap_{r=1}^{\infty} \dot{E}_{k_r}(d_{k_r}) \right\}$$

$$\mathcal{W} = \left\{ x: x \in \bigcap_{r=k}^{\infty} G_r(d_r) \right\}.$$

Then $\mathcal{V}$ has Lebesgue measure 1, $\mathcal{U}$ and $\mathcal{W}$ have measure 0, $\mathcal{E} = \mathcal{U} \cup \mathcal{V} \cup \mathcal{W}$ and $\mathcal{U} \cap \mathcal{V} = \mathcal{V} \cap \mathcal{W} = \mathcal{U} \cap \mathcal{W} = \varnothing$, and we have

THEOREM 3

$\psi'(x) = 0$ *when* $x \in \mathcal{V}$;

$\psi'(x) = +\infty$ *when* $x \in \mathcal{W}$;

no derivative exists when $x \in \mathcal{U}$.

Lemma 3 and Theorem 3 are easily verified with direct calculations, or they can be derived with minor modifications from Sprecher (1966). The observation made at the beginning of this section suggest that smooth functions $\Phi_q(y_q)$ are not expected to be very useful and this is confirmed in the following.

THEOREM 4

If the function $\Phi_q(y_q)$ *is differentiable with bounded derivatives throughout its domain, then* $\Phi_q \circ \xi(\mathbf{x}_q)$ *has vanishing first order partial derivatives almost everywhere.*

The proof is a direct consequence of the chain rule which applies to each composite function $\Phi_q \circ \xi(\mathbf{x}_q)$ almost everywhere: setting $\xi_q = \xi(\mathbf{x}_q)$ we have

$$\frac{\partial \Phi_q}{\partial x_p} = \frac{\partial \Phi_q}{\partial \xi_q} \cdot \frac{d}{dx_p} \psi(x_p + qa) = \frac{\partial \Phi_q}{\partial \xi_q} \cdot 0 = 0$$

almost everywhere, and hence also $\partial f / \partial x_p = 0$ almost everywhere. Consequently, no continuously differentiable function other than $f = const.$ is representable with differentiable functions $\Phi_q(y_q)$ having bounded derivatives.

To determine classes of functions $\Phi_q(y_q)$ for which $\Phi_q \circ \xi(\mathbf{x}_q)$ is differentiable we have to introduce the following concepts (for further discussion see Bruckner, 1994, chapter 5).

DEFINITIONS

Let I be a closed interval, and $\Phi: I \to \mathcal{R}$ a continuous function. Then:

3-1 Φ is of *bounded variation* if $sup \sum |f(b_i) - f(a_i)| < \infty$ where the supremum is taken over all sequences of non overlapping intervals in I.

3-2 Φ is of *generalized bounded variation* if I is the finite or countable union of sets on each of which Φ has bounded variation.

3-3 The point x is a *point of varying monotonicity* of Φ if x has no neighbourhood in which Φ is either constant or monotonic.

The reader is reminded that every function of bounded variation can be written as the difference of two monotonic increasing continuous functions.

THEOREMS

5-1 *If* $\Phi_q \circ \xi(\mathbf{x}_q)$ *has first order partial derivatives then* Φ_q *is of generalized bounded variation.*

5-2 *If* $\Phi_q \circ \xi(\mathbf{x}_q)$ *has bounded first order partial derivatives then* Φ_q *is of bounded variation.*

5-3 *If* $\Phi_q \circ \xi(\mathbf{x}_q)$ *has continuous first order partial derivatives then* Φ_q *is of bounded variation and in addition* $m[\Phi(K)] = 0$, *where* K *is the set of points of varying monotonicity of* Φ_q.

These theorems give necessary conditions for the stated differentiability of the composite functions $\Phi_q \circ \xi(\mathbf{x}_q)$ and they specify the classes from which the functions Φ_q must be drawn, but for the reasons noted above these conditions are not sufficient. Further discussion of these statements entails considerations and arguments quite alien to the setting of this paper, and they offer no insight into our understanding of superpositions. They are therefore omitted. Proofs can be found in Bruckner (1994), Chapter 5, Section 4. The converse question, that of characterizing classes of functions Φ_q for which f has given differentiation properties, remains completely open. It is clear that the results of this section could have been stated in the more general setting of arbitrary monotonic singular functions $\psi_{pq}(x_p)$ instead of translates $\psi(x_p + qa)$.

REFERENCES

Bruckner, A. (1994). *Differentiation of real functions* (CRM Monograph Series). Providence, RI. American Mathematical Society.

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

Katsuura, H. & Sprecher, D. A. (1994). Computational aspects of Kolmogorov's superposition theorem, *Neural Networks* **7**, 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. (1963). *Translations American Mathematical Society* **2**, 55–59.

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

Lorentz, G. G. (1966). *Approximation of Functions.* NY: Holt, Rinehart and Winston.

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

Sprecher, D. A. (1966). On the structure of representations of continuous functions of several variables as finite sums of continuous functions of one variable. *Proceedings of the American Mathematical Society* **17,** 98–105.
Sprecher, D. A. (1993). A universal mapping for Kolmogorov's superposition theorem. *Neural Networks* **6,** 1089–1094.
Sprecher, D. A. (1996a). A numerical implementation of Kolmogorov's superpositions. *Neural Networks* **9,** 765–772.
Sprecher, D. A. (1996b). A numerical construction of a universal function for Kolmogorov's superpositions. *Neural World* **6,** 711–718.

MATHEMATICAL SYMBOLS

$\mathbf{x}$	vector $(x_1, \ldots, x_n)$
$\mathbf{x}_q$	vector $(x_1 + qa, \ldots, x_n + qa)$ where $a = [\gamma(\gamma - 1)]^{-1}$
$\Phi \circ \xi$	composite function $\Phi(\xi)$
$\mathcal{R}$	real time
$\mathcal{E}$	unit interval [0, 1]
$\mathcal{E}^n$	n-dimensional unit cube
d_k	rational numbers $\sum_{r=1}^{k} i_r \gamma^{-r}$, $i_r = 0, 1, \ldots, \gamma - 1$
d_k^q	rational numbers $d_k + q \sum_{r=2}^{k} \gamma^{-r}$
$\mathbf{d}_k$	vector $(d_{k,1}, \ldots, d_{k,n})$
$\mathbf{d}_k^q$	vector $(d_{k,1}^q, \ldots, d_{k,n}^q)$
$\langle i_r \rangle$	$\langle i_1 \rangle = 0$ and $\langle i_r \rangle = \begin{cases} 0 & \textit{when} \quad i_r = 0, 1, \ldots, \gamma - 2 \\ 1 & \textit{when} \quad i_r = \gamma - 1 \end{cases}$ for $r > 1$
$[i_r]$	$[i_1] = 0$ and $[i_r] = \begin{cases} 0 & \textit{when} \quad i_r = 0, 1, \ldots, \gamma - 3 \\ 1 & \textit{when} \quad i_r = \gamma - 2, \gamma - 1 \end{cases}$ for $r > 1$
m_r	$m_r = \langle i_r \rangle \left(1 + \sum_{s=1}^{r-1} [i_s] \times \cdots \times [i_{r-1}] \right)$
$\beta(r)$	$\beta(r) = \dfrac{n^r - 1}{n - 1}$
σ	Continuous function $\sigma : \mathcal{R} \to \mathcal{E}$ such that $\sigma(x) \equiv 0$ for $x \leq 0$ and $\sigma(x) \equiv 1$ for $x \geq 1$.
$\lVert \cdot \rVert$	the uniform norm