Pergamon

Neural Networks, Vol. 7, No. 3, pp. 455–461, 1994

0893-6080/94 $6.00 + .00

0893-6080(93)E0017-2

CONTRIBUTED ARTICLE

Computational Aspects of Kolmogorov's Superposition Theorem

HIDEFUMI KATSUURA[1] AND DAVID A. SPRECHER[2]

[1]San Jose State University and [2]University of California

(Received 14 May 1993; revised and accepted 12 November 1993)

Abstract—*This paper continues the investigation of representations of continuous functions $f(x_1, \dots, x_n)$ with $n \geq 2$ in the form $f(x_1, \dots, x_n) = \sum_{q=0}^{2n} \Phi_q[\sum_{p=1}^{n} \lambda_p \psi(x_p + qe_n)]$ with a predetermined function ψ that is independent of n. The function ψ is defined through its graph that is the limit point of iterated contraction mappings. The functions ψ and Φ_q are the uniform limits of sequences of computable functions constructed with a fixed mapping σ, which itself can be approximated with sigmoid functions.*

Keywords—Superpositions, Kolmogorov, Contraction mapping, Representation of continuous functions of several variables, Approximation of functions.

1. INTRODUCTION

In what follows, $\mathcal{R}$ designates the real line, $\mathcal{E} = [0, 1]$, $\mathcal{D} = [0, 1 + 1/5!]$, and $\mathcal{R}^n$ and $\mathcal{E}^n$ designate the respective n-fold Cartesian products of $\mathcal{R}$ and $\mathcal{E}$. We shall also adhere to the notation $m_k = (1 + 1/5!)k!$ and $I_k = \{i: 0 \leq i \leq m_k\}$. Continuity and convergence are uniform in the Euclidean metric.

A recent version (Sprecher, 1993) of the Superposition Theorem of Kolmogorov (1957) shows that an arbitrary continuous functions $f: \mathcal{E}^n \to \mathcal{R}$ has a representation

$$\begin{cases} f(x_1, \dots, x_n) = \sum_{q=0}^{2n} \Phi_q(\xi_q) \\ \xi_q = \sum_{p=1}^{n} \lambda_p \psi(x_p + qe_n) \end{cases} \tag{1}$$

in which ψ is a predetermined continuous function that is independent of the number of variables n, and e_n is a suitable constant (see below). In this paper we consider computational aspects of the implementation of the functions ψ and Φ_q. The function ψ is obtained through a graph that is the limit point of an iterated composition of contraction mappings on the space of all nonempty compact subsets of $\mathcal{E}^2$. This application of contraction mappings is based on an idea in Katsuura (1991). Using a fixed mapping σ to construct computable approximating sequences converging uniformly to ψ and Φ_q, respectively, we prove:

THEOREM 1. *Let $\sigma: \mathcal{R} \to \mathcal{E}$ be an arbitrary continuous function with norm $\|\sigma\| = 1$, such that $\sigma(x) = 0$ for $x \leq 0$ and $\sigma(x) = 1$ for $x \geq 1$; let $\{\lambda_k\}$ be a sequence of positive integrally independent numbers. There exist λ_k-dependent constants a_{ki}, for which the sequence*

$$\psi_k(x) = \sum_{i=0}^{m_k} a_{ki} \sigma(k!x - i) \tag{2}$$

converges to a continuous function $\psi: \mathcal{D} \to \mathcal{D}$ with the following property: for every real-valued continuous function $f: \mathcal{E}^n \to \mathcal{R}$ with $n \geq 2$ there are continuous functions Φ_q such that eqn (1) holds with $e_n = \sum_{r=2n+2}^{\infty} 1/r!$. Furthermore, there are constants c^s and $\xi^s_{i_{q1} \dots i_{qn}}$ depending on $\lambda_1, \dots, \lambda_n$ and ψ, and constants $b^s_{i_{q1} \dots i_{qn}}$ depending on f, such that the series of functions

$$\Phi^r_q(\xi_q) = \sum_{s=1}^{r} \sum_{i_{q1} \dots i_{qn}} b^s_{i_{q1} \dots i_{qn}} \sigma[c^s(\xi_q - \xi^s_{i_{q1} \dots i_{qn}}) + 1] \tag{3}$$

converge uniformly to $\Phi_q(\xi_q)$ for $q = 0, 1, 2, \dots, 2n$.

Clearly, given a number $\varepsilon > 0$, we can replace eqn (1) by the approximate formula

$$\begin{cases} \left| f(x_1, \dots, x_n) - \sum_{q=0}^{2n} \Phi^r_q(\xi^k_q) \right| < \varepsilon \\ \xi^k_q = \sum_{p=1}^{n} \lambda_p \psi_k(x_p + qe_n) \end{cases}$$

for suitable integers r and k. Furthermore, if $\tilde{\sigma}: \mathcal{R} \to \mathcal{E}$ is a sigmoidal function, and $S(\tilde{\sigma})$ is the set of functions

Requests for reprints should be sent to David A. Sprecher, Department of Mathematics, University of California, Santa Barbara, CA 93106.

$\sum_{s=1}^{r} a_s\tilde{\sigma}(b_s x + c_s)$ with constants a_s, b_s, c_s, then we can approximate respectively the functions ψ_k and Φ_q^r with functions $\tilde{\psi}_k$ and $\tilde{\Phi}_q^r$ in $S(\tilde{\sigma})$ within a predetermined error. The resulting approximate formula is that obtained in Kůrková (1992a) within the constraints imposed by the application to neural networks. Hence, the algorithm developed below for the construction of the functions ψ_k and Φ_q can be adapted to implement in computable form the approximate version of Kůrková (see also Kůrková, 1992b).

Numerical implementation can be enhanced through an adaptation of the constructions in Nees (to appear).

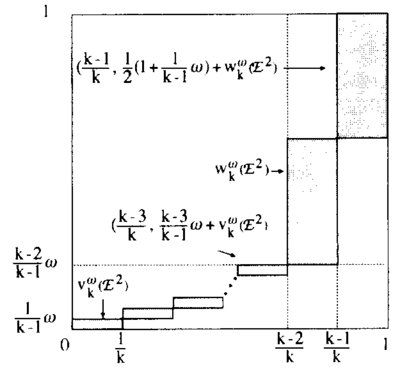

FIGURE 1. The contraction mappings $g_k^\omega(A)$.

2. THE CONTRACTION MAPPINGS

The function ψ appearing in formula (1) must be defined on each interval $[0, 1 + 2ne_n]$ for $n \geq 2$. Because of the inequality $2ne_n < 1/5!$, this condition will be satisfied when ψ has domain $\mathcal{D}$. For $i = 0, 1, 2, \ldots, 5!$, let $u_i:\mathcal{E}^2 \to \mathcal{D}^2$ be mappings defined by

$$u_i(x, y) = \left(\frac{x+i}{5!}, \frac{y+i}{5!}\right)$$

for every point $(x, y) \in \mathcal{E}^2$. For each value $i > 0$ each of these mappings can be regarded as a translation of u_0 along the diagonal of $\mathcal{D}^2$. For the respective collections $G(\mathcal{E}^2)$ and $G(\mathcal{D}^2)$ of nonempty closed subsets of $\mathcal{E}^2$ and $\mathcal{D}^2$, let $U:G(\mathcal{E}^2) \to G(\mathcal{D}^2)$ be a function defined by

$$U(A) = \bigcup_{i=0}^{5!} u_i(A)$$

for every $A \in G(\mathcal{E}^2)$.

Let $k \geq 6$ be a given integer. With an arbitrary rational number $0 < \omega < 1$ we define the mappings $v_k^\omega:\mathcal{E}^2 \to \mathcal{E}^2$ and $w_k^\omega:\mathcal{E}^2 \to \mathcal{E}^2$

$$v_k^\omega(x, y) = \left(\frac{x}{k}, \frac{\omega y}{k-1}\right)$$

$$w_k^\omega(x, y) = \left[\frac{k-2}{k} + \frac{x}{k}, \frac{k-2}{k-1}\omega + \left(1 - \frac{k-2}{k-1}\omega\right)\frac{y}{2}\right].$$

Consider the collection $G(\mathcal{E}^2)$ of nonempty compact subsets of $\mathcal{E}^2$, and let $g_k^\omega:G(\mathcal{E}^2) \to G(\mathcal{E}^2)$ be a mapping defined by

$$g_k^\omega(A) = \left\{\bigcup_{t=0}^{k-3}\left[\left(\frac{t}{k}, \frac{t\omega}{k-1}\right) + v_k^\omega(A)\right]\right\} \cup w_k^\omega(A) \cup \left\{\left[\frac{k-1}{k}, \frac{1}{2}\left(1 + \frac{(k-2)\omega}{k-1}\right)\right] + w_k^\omega(A)\right\} \quad (4)$$

for every set $A \in G(\mathcal{E}^2)$ (Figure 1). From the Hausdorff metric d_H (see the proof of Corollary 1 below) it follows at once that $d_H(g_k^\omega(A), g_k^\omega(B)) \leq \frac{1}{2} d_H(A, B)$ for every $A, B \in G(\mathcal{E}^2)$.

The convergence arguments based on these contraction mappings utilize an analog of the Banach Contraction mapping theorem and a corollary. These are given later in this paper.

3. THE CONSTRUCTION OF ψ_k AND ψ

The functions ψ_k and ψ are subject to certain conditions that are satisfied with a specific selection of a decreasing sequence of rational numbers $\omega_k > 0$ depending on the numbers λ_k. Without loss of generality we shall construct a sequence $\{\lambda_k\}$ for which $\sum \lambda_k < 1$ (Sprecher, 1993). This additional assumption simplifies computations, and this is one of the aims of the constructions in this paper. The numbers λ_k and ω_k will be selected by an inductive boot-strapping procedure based on the procedure described in Sprecher (1993, Lemma 2). In what follows it will be assumed that the numbers λ_k are selected to comprise a sequence of integrally independent numbers. **The λ_k are integrally independent if $\sum r_k\lambda_k \neq 0$ for any finite selection of integers r_k for which $\sum |r_k| \neq 0$.** It is convenient to let λ_1 be rational, and λ_k be transcendental for each integer $k > 1$. In the approximate version developed in Nees (to appear), the transcendental numbers λ_k are approximated with rational numbers to facilitate numerical work. The procedures in Nees can also be applied to the constructions below. We begin with the following lemma, stated for its subsequent application to rationally independent numbers. Integral independence and rational independence, however, are clearly equivalent. The lemma is proved at the end of this paper.

LEMMA 1. *Let $\theta_1 > \theta_2 > \ldots > \theta_m > 0$, $m \geq 2$, be a set of rationally independent numbers; let Q be a set of rational numbers: $Q = \{r_0 = 0 < r_1 < r_2 < r_3 < \ldots < r_k = 1\}$, $k > m$. If*

$$\theta_m(1 + r_1) < \min_{s_p, t_p \in Q}\left|\sum_{p=1}^{m-1} \theta_p(s_p - t_p)\right|,$$

$$\sum_{p=1}^{m-1} |s_p - t_p| \neq 0 \quad (5)$$

then

$$\theta_m r_1 = \min_{s_p, t_p \in Q} \left| \sum_{p=1}^{m} \theta_p (s_p - t_p) \right|, \quad \sum_{p=1}^{m} |s_p - t_p| \neq 0. \quad (6)$$

We now proceed with the construction of the sequence of functions $\psi_k : \mathcal{D} \rightarrow \mathcal{D}$ by induction on k. We begin with the value $k = 5$ because smaller values of k are not useful in the proof of Theorem 1.

Designate by Ω the restriction of the graph of σ to $\mathcal{E}$:

$$\Omega = \{(x, \sigma(x)) : x \in \mathcal{E}\}.$$

Let the initial function $\psi_5 : \mathcal{D} \rightarrow \mathcal{D}$ be

$$\psi_5(x) = \sum_{i=0}^{m_5} a_{5,i} \sigma(5!x - i)$$

with $a_{5,i} = 1/5!$ for each value $i \in I_5$. Then

$$\psi_5\left(\frac{i}{5!}\right) = \frac{i}{5!} \quad \text{for} \quad i \in I_5$$

(Figure 2). Let $\lambda_1 = 1/5!$, and select a transcendental number $\lambda_2 < 1/(5! + 1)$. Then $0 < (1 + 1/5!)\lambda_2 < \lambda_1$, and by Lemma 2

$$\lambda_2 \frac{1}{5!} = \min_{i_p, j_p \in I_5} \left| \sum_{p=1}^{2} \lambda_p \left[\psi_5\left(\frac{i_p}{5!}\right) - \psi_5\left(\frac{j_p}{5!}\right) \right] \right|, \quad \sum_{p=1}^{2} |i_p - j_p| \neq 0.$$

We select a rational number ω_5 such that $0 < \omega_5 < \lambda_2$.

Let $g_6 = g_6^{\omega_5}$, and define $\psi_6 : \mathcal{D} \rightarrow \mathcal{D}$ to be the function having the graph $U \circ g_6(\Omega)$ obtained from eqn (4) through the substitutions $A = \Omega$ and $\omega = \omega_5$ (Figure 3). Notice that the mechanics of this composition is to place a copy of Figure 1 for $k = 6$ into each of the marked squares in Figure 2. With the values $\psi_6(i/6!)$ so obtained, we define

$$a_{6,i} = \psi_6\left(\frac{i+1}{6!}\right) - \psi_6\left(\frac{i}{6!}\right)$$

and have

$$\psi_6(x) = \sum_{i=0}^{m_6} a_{6,i} \sigma(6!x - i).$$

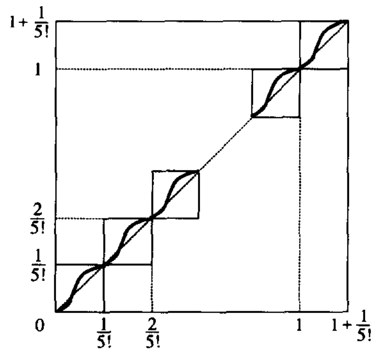

FIGURE 2. The graph of $\psi_5(x)$.

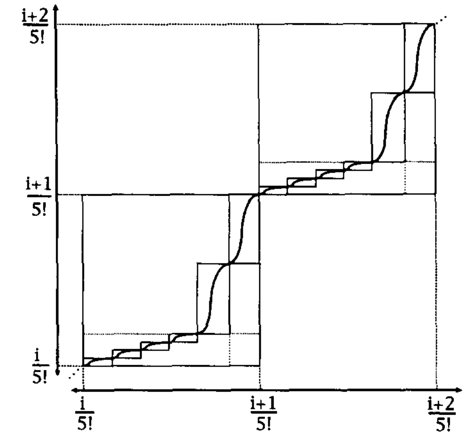

FIGURE 3. Segment of the graph of $\psi_6(x)$.

Select a transcendental number λ_3 such that

$$\left[1 + \psi_6\left(\frac{1}{6!}\right)\right] \lambda_3 < \min_{i_p, j_p \in I_6} \left| \sum_{p=1}^{2} \lambda_p \left[\psi_6\left(\frac{i_p}{6!}\right) - \psi_6\left(\frac{j_p}{6!}\right) \right] \right|,$$

$$\sum_{p=1}^{2} |i_p - j_p| \neq 0.$$

Then

$$\lambda_3 \psi_6\left(\frac{1}{6!}\right) = \min_{i_p, j_p \in I_6} \left| \sum_{p=1}^{3} \lambda_p \left[\psi_6\left(\frac{i_p}{6!}\right) - \psi_6\left(\frac{j_p}{6!}\right) \right] \right|,$$

$$\sum_{p=1}^{3} |i_p - j_p| \neq 0.$$

The values $\psi_6(i/6!)$ are rational, and in particular, $\psi_6(1/6!) = (1/5!) \cdot (\omega_5/5)$, as is deduced from the definition of ψ_5 and eqn (4). This assures that $\lambda_3 \neq 0$, so that we can select a rational number ω_6 such that $0 < \omega_6 < \lambda_3$.

Assume that ψ_{k-1}, λ_{k-4}, and ω_{k-1} are already determined. Let

$$g_k = g_6^{\omega_5} \circ g_7^{\omega_6} \circ \ldots \circ g_k^{\omega_{k-1}},$$

and define the function $\psi_k : \mathcal{D} \rightarrow \mathcal{D}$ through the graph $U \circ g_k(\Omega)$. With the values $\psi_k(i/k!)$ so obtained, we set

$$a_{ki} = \psi_k\left(\frac{i+1}{k!}\right) - \psi_k\left(\frac{i}{k!}\right)$$

and obtain the function $\psi_k(x)$ in eqn (2). We now select a transcendental number λ_{k-3} such that

$$\left[1 + \psi_k\left(\frac{1}{k!}\right)\right] \lambda_{k-3} < \min_{i_p, j_p \in I_k} \left| \sum_{p=1}^{k-4} \lambda_p \left[\psi_k\left(\frac{i_p}{k!}\right) - \psi_k\left(\frac{j_p}{k!}\right) \right] \right|,$$

$$\sum_{p=1}^{k-4} |i_p - j_p| \neq 0,$$

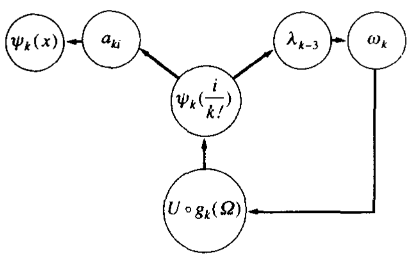

FIGURE 4. The implementation of $\psi_k(x)$.

and then by Lemma 1,

$$\lambda_{k-3}\psi_k\left(\frac{1}{k!}\right) = \min_{i_p, j_p \in I_k} \left|\sum_{p=1}^{k-3} \lambda_p\left[\psi_k\left(\frac{i_p}{k!}\right) - \psi_k\left(\frac{j_p}{k!}\right)\right]\right|,$$

$$\sum_{p=1}^{k-3} |i_p - j_p| \neq 0.$$

Observing that

$$\psi_k\left(\frac{1}{k!}\right) = \frac{1}{5!} \times \frac{\omega_5}{5} \times \frac{\omega_6}{6} \times \ldots \times \frac{\omega_{k-1}}{k-1},$$

we select a rational number ω_k such that $0 < \omega_k < \lambda_{k-3}$. This concludes the implementation of ψ_k. The iteration is presented schematically in Figure 4.

An examination of the above construction shows that for each integer $k \geq 5$ and $i \in I_k$ we have

$$\psi_k\left(\frac{i}{k!}\right) = \psi_r\left(\frac{i}{k!}\right) \quad \text{for all integers} \quad r \geq k,$$

and from this we deduce that

$$\psi\left(\frac{i}{k!}\right) = \psi_k\left(\frac{i}{k!}\right) \quad \text{for} \quad i \in I_k \quad \text{and} \quad k \geq 5. \tag{7}$$

We claim that $\{\psi_k(x)\}$ is a Cauchy sequence in the supremum norm. This is verified as follows: from the fact that the graphs of both $\psi_r(x)$ and $\psi_k(x)$ are contained in $U \circ g_k(\mathcal{E}^2) = U \circ g_6^{\omega_5} \circ g_7^{\omega_6} \circ \ldots \circ g_k^{\omega_{k-1}}(\mathcal{E}^2)$, we conclude that $\sup_{x \in \mathcal{E}} |\psi_k(x) - \psi_r(x)|$ is bounded by the height of the largest rectangle in $U \circ g_k(\mathcal{E}^2)$. From the definition of the mappings g_k^ω we see that the height of the largest rectangle in $U \circ g_k^\omega(\mathcal{E}^2)$ is $1/5! \cdot \frac{1}{2}[1 - ((k-2)/(k-1))\rho]$ (Figure 1). Therefore, setting

$$\rho_k = \frac{k}{5!} \cdot \frac{\omega_5}{5} \cdot \frac{\omega_6}{6} \cdot \ldots \cdot \frac{\omega_k}{k},$$

we see that the height of the largest rectangle in $U \circ g_k(\mathcal{E}^2)$ is given by

$$\frac{1}{5!} \cdot \frac{1}{2}\left(1 - \frac{4}{5}\rho_5\right)\frac{1}{2}\left(1 - \frac{5}{6}\rho_6\right) \ldots \frac{1}{2}\left(1 - \frac{k-2}{k-1}\rho_{k-1}\right)$$
$$= \frac{1}{5!2^{k-5}} \prod_{s=5}^{k-1}\left(1 - \frac{s-1}{s}\rho_s\right). \tag{8}$$

Because $0 < [1 - ((s-1)/s)\rho_s] < 1$ for all values $s \geq 5$ we have

$$\sup_{x \in \mathcal{D}} |\psi_k(x) - \psi_r(x)| < \frac{1}{5!2^{k-5}}. \tag{9}$$

It follows that $\{\psi_k(x)\}$ is a Cauchy sequence. Its uniform limit $\psi(x)$ is a continuous monotonically increasing function. Using eqn (8) and letting $r \to \infty$ in eqn (9) gives the following estimate on the rate of convergence of the sequence $\{\psi_k\}$:

$$\sup_{x \in \mathcal{D}} |\psi(x) - \psi_k(x)| < \frac{1}{5!2^{k-5}} \prod_{r=5}^{k-1}\left(1 - \frac{r-1}{r}\rho_r\right).$$

This concludes the construction of $\psi(x)$.

REMARK. We began the construction with the value $k = 5$, but it is evident that we could have started with any integer $k \geq 5$. The implication of this is the following: for any given number $\varepsilon > 0$ we can construct a function $\psi: \mathcal{D} \to \mathcal{D}$ meeting the conditions of Theorem 1 such that $\sup_{x \in \mathcal{D}} |\psi(x) - x| < \varepsilon$.

4. CONSTRUCTION OF THE FUNCTIONS Φ_q

Based on the strategy in Kolmogorov (1957), we select by induction on r a subsequence of integers k_r of the sequence $\{k\}$, and construct corresponding uniformly convergent sequences of continuous functions $\Phi_q^r(\xi_q) = \sum_{s=1}^{r} \varphi_q^s(\xi_q)$ for $q = 0, 1, 2, \ldots, 2n$, each with uniform limit Φ_q, with Φ_q being the functions in formula (1). In what follows, we shall assume that $n \geq 2$ is a given integer, and that $k_1 \geq 2n + 1$.

We begin with the following important property of the function ψ (Sprecher, 1993, Lemma 1):

LEMMA 2. *For each integer $k \geq 5$ and $i \in I_k$ we have*

$$\psi\left(\frac{i}{k!} + \delta_k\right) = \psi\left(\frac{i}{k!}\right) + \nu_k$$

where

$$\delta_k = \frac{1}{k!} - \sum_{r=k+1}^{\infty} \frac{1}{r!}$$

and

$$\nu_k = \sum_{r=k}^{\infty} \frac{r-1}{r}\rho_r.$$

We mention in passing that this lemma determines the smoothness characteristics of ψ, and with appropriate modifications one can also derive conclusions similar to those in Sprecher (1966). With it, we define for each value $q = 0, 1, 2, \ldots, 2n$ the constants

$$\xi_{i_{q1}\ldots i_{qn}}^r = \sum_{p=1}^{n} \lambda_p \psi\left(\frac{i_{qp}}{k_r!} + qe_n\right),$$

and the constants

$$\gamma_r = \nu_{k_r} \sum_{p=1}^{n} \lambda_p.$$

We note that the range of each index i_{qp} depends on the domain of ψ. These ranges, however, are easily computed. For each n-tuple $(i_{q1}, \ldots, i_{qn}) \neq (0, \ldots, 0)$ we designate by $(i_{\bar{q}1}, \ldots, i_{\bar{q}n})$ that n-tuple for which $\xi^r_{i_{\bar{q}1}\ldots i_{\bar{q}n}}$ is the immediate predecessor of $\xi^r_{i_{q1}\ldots i_{qn}}$ (Figure 6). Clearly, the points $\xi^r_{0\ldots 0}$ have no predecessors. Setting

$$c^r = (\mu_{k_r} - \gamma_r)^{-1},$$

we note that for any integer r

$$\sigma[c^r(\xi_q - \xi^r_{i_{q1}\ldots i_{qn}}) + 1] = \begin{cases} 0 & \text{if} \quad \xi_q \le \xi^r_{i_{q1}\ldots i_{qn}} - (\mu_{k_r} - \gamma_r) \\ 1 & \text{if} \quad \xi_q \ge \xi^r_{i_{q1}\ldots i_{qn}} \end{cases}$$

Now let k_1 be an integer such that

$$|f(x_1, \ldots, x_n) - f(y_1, \ldots, y_n)| \le \frac{1}{2n+2} \|f\| \quad \text{for} \quad |x_p - y_p| \le \frac{1}{k_1!}.$$

Let

$$b^1_{0\ldots 0} = \frac{1}{n+1} f(0, \ldots, 0)$$

and for each value of q, $q = 0, 1, 2, \ldots, 2n$, for which $(i_{q1}, \ldots, i_{qn}) \neq (0, \ldots, 0)$ let

$$b^1_{i_{q1}\ldots i_{qn}} = \frac{1}{n+1}\left[f\left(\frac{i_{q1}}{k_1!}, \ldots, \frac{i_{qn}}{k_1!}\right) - f\left(\frac{i_{\bar{q}1}}{k_1!}, \ldots, \frac{i_{\bar{q}n}}{k_1!}\right)\right];$$

define the functions

$$\Phi^1_q(\xi_q) = \varphi^1_q(\xi_q) = \sum_{i_{q1}, \ldots, i_{qn}} b^1_{i_{q1}\ldots i_{qn}} \sigma[c^1(\xi_q - \xi^1_{i_{q1}\ldots i_{qn}}) + 1]$$

and set

$$f_1(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) - \sum_{q=0}^{2n} \Phi^1_q(\xi_q).$$

Assuming that an integer k_{r-1} and the functions $\varphi^{r-1}_q(\xi_q)$ are already determined, we set

$$f_{r-1}(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) - \sum_{q=0}^{2n} \Phi^{r-1}_q(\xi_q)$$

where

$$\Phi^{r-1}_q(\xi_q) = \sum_{s=1}^{r-1} \varphi^s_q(\xi_q)$$

and select a constant k_r such that

$$|f_{r-1}(x_1, \ldots, x_n) - f_{r-1}(y_1, \ldots, y_n)| \le \frac{1}{2n+2} |f_{r-1}| \quad \text{when} \quad |x_p - y_p| \le \frac{1}{k_r!}.$$

We now define the constants

$$b^r_{0\ldots 0} = \frac{1}{n+1} f_{r-1}(0, \ldots, 0)$$

and

$$b^r_{i_{q1}\ldots i_{qn}} = \frac{1}{n+1}\left[f_{r-1}\left(\frac{i_{q1}}{k_r!}, \ldots, \frac{i_{qn}}{k_r!}\right) - f_{r-1}\left(\frac{i_{\bar{q}1}}{k_r!}, \ldots, \frac{i_{\bar{q}n}}{k_r!}\right)\right]$$

and with these we define

$$\varphi^r_q(\xi_q) = \sum_{i_{q1}, \ldots, i_{qn}} b^r_{i_{q1}\ldots i_{qn}} \sigma[c^r(\xi_q - \xi^r_{i_{q1}\ldots i_{qn}}) + 1].$$

We note in passing that

$$\varphi^r_q(\xi_q) = \frac{1}{n+1}\left[f_{r-1}\left(\frac{i_{\bar{q}1}}{k_r!}, \ldots, \frac{i_{\bar{q}n}}{k_r!}\right)(1 - \sigma) + f_{r-1}\left(\frac{i_{q1}}{k_r!}, \ldots, \frac{i_{qn}}{k_r!}\right)\sigma\right],$$

where we have used the abbreviated notation $\sigma = \sigma[c^r(\xi_q - \xi^r_{i_{q1}\ldots i_{qn}}) + 1]$, and specifically

$$\varphi^r_q(\xi_q) = \frac{1}{n+1} f_{r-1}\left(\frac{i_{q1}}{k_r!}, \ldots, \frac{i_{qn}}{k_r!}\right) \quad \text{when} \quad \xi_q \ge \xi^r_{i_{q1}\ldots i_{qn}}.$$

The implementation of these iterative approximations completes the construction of the functions Φ_q. The process is presented in Figure 5.

To complete the proof of Theorem 1, it is necessary to show that $\lim_{r\to\infty} f_r = 0$, and that $\lim_{r\to\infty} \Phi^r_q = \Phi_q$ for $q = 0, 1, 2, \ldots, 2n$. For these convergence arguments, introduce the intervals

$$T^r_q(i_{q1}, \ldots, i_{qn}) = [\xi^r_{i_{q1}\ldots i_{qn}}, \xi^r_{i_{q1}\ldots i_{qn}} + \gamma_r]$$

(Figure 6). From the observation that $\xi^r_{i_{\bar{q}1}\ldots i_{\bar{q}n}} + \gamma_r < \xi^r_{i_{q1}\ldots i_{qn}} - \mu_{k_r} + \gamma_r$, it follows at once that

$$\operatorname{diam}[T^r_q(i_{q1}, \ldots, i_{qn})] \to 0 \quad \text{as} \quad r \to \infty$$

and that these intervals are pairwise disjoint:

$$T^r_q(i_{q1}, \ldots, i_{qn}) \cap T^r_q(j_{q1}, \ldots, j_{qn}) = 0 \quad \text{when} \quad (i_{q1}, \ldots, i_{qn}) \neq (j_{q1}, \ldots, j_{qn}). \tag{10}$$

It was established in Sprecher (1993) that for each point $(x_1, \ldots, x_n) \in \mathscr{E}^n$ there are not less than $n+1$ value of q for which $\xi_q \in T^r_q(i_{q1}, \ldots, i_{qn})$ for each value

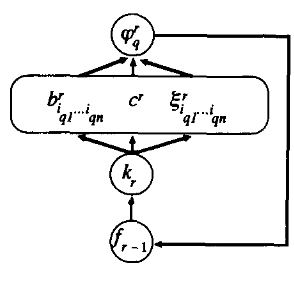

FIGURE 5. The implementation of the iterative approximations to *f*.

of r. With this fact, we can now apply directly the convergence arguments in Kolmogorov (1957) or those in Lorenz (1966) or Sprecher (1965) to complete the proof of Theorem 1. The repetition of these details is omitted here.

5. A SUPPORTING LEMMA AND PROOFS

The basis of the contraction mappings used in this paper is the following:

LEMMA 3. *Let (X, d) be a compact metric space. For every positive integer k, let w_k be a contraction mapping from X into X such that $d[w_k(x), w_k(y)] \le \alpha d(x, y)$ for some constant $0 \le \alpha < 1$ and for all $x, y \in X$. For each k, let $g_k: X \to X$ be a continuous mapping defined by $g_k = w_1 \circ w_2 \circ \ldots \circ w_k$. Then $\lim_{k\to\infty} g_k(x)$ exists for each $x \in X$, and moreover, $\lim_{k\to\infty} g_k(x) = \lim_{k\to\infty} g_k(y)$ for all $x, y \in X$.*

The proof, which is very similar to that of the theorem of Banach (Liusternik & Sobolev, 1961), is omitted here.

COROLLARY 1. *For every positive integer k, let w_k: $G(\mathscr{E}^2) \to G(\mathscr{E}^2)$ be a contraction mapping such that $d_H[w_k(A), w_k(B)] \le \alpha d_H(A, B)$ for some number $0 \le \alpha < 1$ and every $A, B \in G(\mathscr{E}^2)$; let $g_k: G(\mathscr{E}^2) \to G(\mathscr{E}^2)$ be a mapping as in the lemma. Then $\lim_{k\to\infty} g_k(A)$ exists for every $A \in G(\mathscr{E}^2)$, and moreover $\lim_{k\to\infty} g_k(A) = \lim_{k\to\infty} g_k(B)$ for every $A, B \in G(\mathscr{E}^2)$.*

Proof. If $A \in G(\mathscr{E}^2)$, and a number $\varepsilon > 0$ is given, then we define the set

$$N(A;\varepsilon) = \{x \in \mathscr{E}^2 : d(x, y_0) < \varepsilon \quad \text{for some} \quad y_0 \in A\}.$$

For every $A, B \in G(\mathscr{E}^2)$ we define

$$d_H(A, B) = \inf\{\varepsilon > 0 : A \subset N(B;\varepsilon) \quad \text{and} \quad B \subset N(A;\varepsilon)\}.$$

To prove the corollary, we only have to note that $G(\mathscr{E}^2)$ is compact with this metric, as follows from the compactness of $\mathscr{E}^2$. ∎

That the outcome of the constructions leading to ψ does not depend on the initial choice of σ is seen from the following observation: clearly, g_k satisfies Corollary 1 for each integer $k \ge 5$, so that $\lim_{k\to\infty} g_k(A)$ exists for each $A \in G(\mathscr{E}^2)$. Moreover, $\lim_{k\to\infty} g_k(A) = \lim_{k\to\infty} g_k(B)$ for all $A, B \in G(\mathscr{E}^2)$, and in particular, $\lim_{k\to\infty} g_k(A)$ is the graph of ψ. Accordingly, the function ψ constructed here has the same properties as the function ψ obtain in Sprecher (1993), and the first part of Theorem 1 therefore follows.

We conclude this section with proofs of Lemmas 1 and 2:

Proof of Lemma 1. Because $\theta_m r_1$ is one of the possible terms in the summation (6), it is clear that

$$\min_{s_p, t_p \in Q} \left| \sum_{p=1}^{m} \theta_p(s_p - t_p) \right| \le \theta_m r_1,$$

and so the lemma will be proven if we can demonstrate that strict inequality is not possible. Suppose that there is a given selection of numbers s'_p, $t'_p \in Q$ with $\sum_{p=1}^{m} |s'_p - t'_p| \ne 0$, for which a strict inequality is possible. Without loss of generality we may then assume that $\sum_{p=1}^{m} \theta_p(s'_p - t'_p) > 0$, and because $|s'_m - t'_m| < 1$, we obtain from eqn (6) the inequalities

$$0 < \sum_{p=1}^{m} \theta_p(s'_p - t'_p) < \sum_{p=1}^{m-1} \theta_p(s'_p - t'_p) + \theta_m < \theta_m(1 + r_1).$$

Designating with μ the right side of the inequality in eqn (5), it follows that, in particular, $\mu < \theta_m(1 + r_1)$. This, however, is excluded by that inequality, and the lemma follows. ∎

Abbreviated proof of Lemma 2. Consider a fixed value of k and an arbitrary integer $N > k$. From the mapping

$$U \circ g_N = U \circ g_6^{\omega_5} \circ g_7^{\omega_6} \circ \ldots \circ g_N^{\omega_{N-1}}$$

and Figure 1 it is readily derived that

$$\psi\left(\frac{i}{k!} + \sum_{r=k+1}^{N} \frac{r-2}{r!}\right) = \psi\left(\frac{i}{k!}\right) + \sum_{r=k+1}^{N} \frac{r-2}{r-1} \rho_{r-1}$$
$$= \psi\left(\frac{i}{k!}\right) + \sum_{r=k}^{N-1} \frac{r-1}{r} \rho_r.$$

But

$$\lim_{N\to\infty} \sum_{r=k+1}^{N} \frac{r-2}{r!} = \lim_{N\to\infty} \left(\frac{1}{k!} - \sum_{r=k+1}^{N} \frac{1}{r!} - \frac{1}{N!}\right)$$
$$= \frac{1}{k!} - \sum_{r=k+1}^{\infty} \frac{1}{r!} = \delta_k,$$

and

$$\lim_{N\to\infty} \sum_{r=k}^{N-1} \frac{r-2}{r} = \nu_k,$$

and the lemma follows. ∎

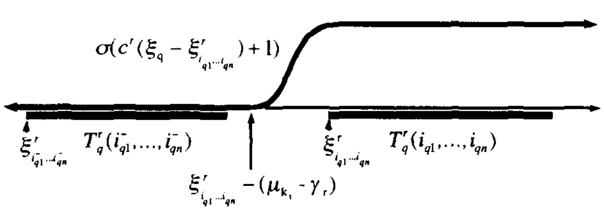

FIGURE 6. The function $\sigma[c'(\xi_q - \xi'_{i_{q1} \ldots i_{qn}}) + 1]$ and intervals $T'_q(i_{q1}, \ldots, i_{qn})$.

REFERENCES

Katsuura, H. (1991). Continuous nowhere-differentiable functions—an application of contraction mappings. *American Mathematical Monthly*, **98**(5), 411–416.

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,* 1963, **2**(28), 55–59.]

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

Kůrková, V. (1992b). Kolmogorov's theorem is relevant. *Neural Computation,* **3,** 617–622.

Lorentz, G. G. (1996). *Approximation of functions.* New York: Holt, Rinehart and Winston.

Liusternik, L. A., & Sobolev, V. J. (1961). *Elements of functional analysis.* New York: Frederick Ungar Publishing Co.

Nees, M. (to appear). Approximative versions of Kolmogorov's superposition theorem, proved constructively.

Sprecher, D. A. (1965). On the structure of continuous functions of several variables. *Transactions American Mathematical Society,* **115**(3), 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 American Mathematical Society,* **17**(1), 98–105.

Sprecher, D. A. (1993). A universal mapping for Kolmogorov's superposition theorem. *Neural Networks,* **6,** 1089–1093.

NOMENCLATURE

$\mathcal{R}$	real line
$\mathcal{R}^n$	n-dimensional Euclidean space
$\mathcal{E}$	interval $[0,1]$
$\mathcal{E}^n$	n-dimensional unit cube
$\mathcal{D}$	interval $[0, 1 + 1/5!]$
$\mathcal{D}^2$	The Cartesian product $\mathcal{D} \times \mathcal{D}$
m_k	integer $(1 + 1/5!)k!$
I_k	set of integers $0 \leq i \leq m_k$
$\mid \ \mid$	supremum norm