Pergamon

Appl. Math. Lett. Vol. 11, No. 3, pp. 45–49, 1998

0893-9659/98 $19.00 + 0.00
PII: S0893-9659(98)00032-9

Approximation of Continuous Functions of Several Variables by an Arbitrary Nonlinear Continuous Function of One Variable, Linear Functions, and Their Superpositions

A. N. GORBAN'
Krasnoyarsk Computing Center, Russian Academy of Sciences, Siberian Branch
Krasnoyarsk-36, 660036, Russian Federation
gorban@cc.krascience.rssi.ru

(Received and accepted July 1997)

Communicated by A. Alekseev

Abstract—Linear spaces of continuous functions of real variables closed under the superposition operation are considered. It has been proved that when such a space contains constants, linear functions, and at least one nonlinear function, it is dense in the space of all continuous functions in the topology of uniform convergence on compact sets.

So, the approximation of continuous functions of several variables by an arbitrary nonlinear continuous function of one variable, linear functions, and their superpositions is possible.

Keywords—Approximation, Superposition, Neural networks, Weierstrass theorem.

1. INTRODUCTION

From a mathematical standpoint, this work extends the classical Weierstrass and Stone theorems on density of polynomials and continuous function rings separating points [1] to semigroups of functions. The product operation under which the function rings are closed is replaced by the operation of function superposition.

The main applied issue of the results obtained is that every nonlinearity makes it possible to compute any function: if we have at least one device computing some (arbitrary) nonlinear continuous function, and if it is possible to calculate every linear combination and superposition of functions as well as to use constants, then every continuous function can be approximated.

The question of representing continuous functions of several variables by superposition of continuous functions with fewer variables has been the essence of Gilbert's 13^{th} problem. Arnold and Kolmogorov resolved the problem [2,3]: every continuous function of several variables can be represented by the superposition of functions only of two variables.

Specifically, *every continuous function of* n *variables can be represented as a sum of* $3n$ *functions; each summand is a superposition produced by substituting one of the arguments in the function of two variables for a function of* $n-1$ *variables.*

This work was supported by Krasnoyarsk Regional Science Foundation, Grant 6F0124.

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX

Descending by n to 2, we come to a representation of the function of n variables from the function of two variables (by superposition and addition operations).

The question of representing continuous functions of several variables by continuous functions of a single variable has also been solved.

In the paper [4], Kolmogorov has proved an elegant theorem, which follows.

KOLMOGOROV THEOREM. *(See [4].) Every continuous function of n variables defined in the standard n-dimensional cube can be represented in the following form:*

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} h_q \left[\sum_{p=1}^{n} \varphi_q^p(x_p) \right],$$

where the functions $h_q(u)$ and $\varphi_q^p(x_p)$ are continuous, and moreover, the functions $\varphi_q^p(x_p)$ are standard, i.e., they are independent of the function f.

It should be noted that the functions $\varphi_q^p(x_p)$ used here are essentially nonsmooth. Therefore, in applied problems, it is necessary either to increase the number of variables in the functions over which expansion is done, or to consider approximate and not accurate representations.

In this letter, we study approximate representations of functions of several variables by functions of one variable. In contrast to the Kolmogorov theorem, the question we address is: how broad is the class of functions that can be approximated using a single (*arbitrarily taken*, and not specially constructed) nonlinear continuous function?

The answer is: *every continuous function can be arbitrarily accurately approximated by operations of addition, multiplication by a number, superposition, and an arbitrary (one is sufficient) continuous nonlinear function of one variable.*

Renewed interest in the classical question of approximation functions of several variables by superpositions and sums of functions of one variable and a new version of this question (confined to one arbitrarily taken nonlinear function) have been invoked by neurocomputing studies.

In applied problems of approximation, the current trend is to use devices which compute superpositions of simple functions of one variable, and their linear combinations. These devices receive the name "artificial neural networks" [5–7].

Most applications of artificial neural networks are concerned with estimating and approximating unknown input-output dependency from examples, in order to use this dependency to predict future outputs from the given input values.

The question of what functions they are able to approximate is becoming topical.

Relevant theorems on completeness for several versions of the neural networks have been proved [8–12]. They are distinct in admissible architectures of the networks, in functions that are computed by an individual "neuron", etc. The present work proves the theorem on completeness for arbitrary continuous functions.

For the neural networks, our result states that the function of neuron activation must be nonlinear—and nothing else. Whatever this nonlinearity is, the network of connections can be constructed, and coefficients of linear connections between the neurons can be adjusted in such a way that the neural network will compute any continuous function from its input signals with any given accuracy.

2. SEMIGROUPS OF CONTINUOUS FUNCTIONS OF ONE VARIABLE

Let us consider the space $\mathbf{C}(\mathbf{R})$ of continuous functions on a real axis in the topology of the uniform convergence on compact sets. The space $\mathbf{C}(\mathbf{R})$ with superposition of functions $(f \circ g)(x) = f(g(x))$ on it is a semigroup. Function id $(\mathrm{id}(x) \equiv x)$ is a unit in this semigroup.

THEOREM 1. *Let* $\mathbf{E}$ *be a closed subspace in* $\mathbf{C}(\mathbf{R})$ *which is a semigroup,* $1 \in \mathbf{E}$ *and* $\mathrm{id} \in \mathbf{E}$ *(1 is a function identically equal to 1). Then, either* $E = \mathbf{C}(\mathbf{R})$ *or* $\mathbf{E}$ *is a subspace of linear functions* $(f(x) = ax + b)$.

The proof is based on several lemmas.

LEMMA 1. *Under the conditions of Theorem 1, let there exist function* $f \in \mathbf{E}$ *which is not linear. Then, there exists a twice continuously differentiable function* $g \in \mathbf{E}$ *which is not linear.*

PROOF. Let $v(x) \in \mathbf{C}^{\infty}(\mathbf{R})$, $v(x) = 0$ for $|x| > 1$. Consider the averaging operator

$$J_{\varepsilon} f(x) = \int_{\mathbf{R}} f(x+y) \frac{1}{\varepsilon} v\left(\frac{y}{\varepsilon}\right) dy.$$

The inclusion $J_{\varepsilon} f \in \mathbf{E}$ holds for every $\varepsilon > 0$. Indeed, $f(x+y) \in \mathbf{E}$ for every fixed y (the constants belong to $\mathbf{E}$, and $\mathbf{E}$ is closed under linear operations and superposition of the functions). Integral $J_{\varepsilon} f$ is in $\mathbf{E}$, because $\mathbf{E}$ is a closed linear subspace in $\mathbf{C}(\mathbf{R})$ and this integral is the limit of finite sums.

Function $J_{\varepsilon} f$ is in $\mathbf{C}^{\infty}(\mathbf{R})$ because

$$\int_{\mathbf{R}} f(x+y) \frac{1}{\varepsilon} v\left(\frac{y}{\varepsilon}\right) dy = \int_{\mathbf{R}} f(z) \frac{1}{\varepsilon} v\left(\frac{z-x}{\varepsilon}\right) dz$$

(remember that v is a function with a compact support).

There exists $\varepsilon > 0$ such that the function $g = J_{\varepsilon} f$ is not linear because $J_{\varepsilon} f \to f$ with $\varepsilon \to 0$, the space of linear functions is closed, and f is not a linear function. So, under the assumptions of the lemma, there exists a nonlinear function $g \in \mathbf{E} \cap \mathbf{C}^{\infty}(\mathbf{R})$ which can be taken in form $g = J_{\varepsilon} f$.

LEMMA 2. *Under the conditions of Theorem 1, let there exist a twice differentiable function* $g \in \mathbf{E}$ *which is not linear. Then, the function* $q(x) = x^2$ *is in* $\mathbf{E}$.

PROOF. There is a point x_0 for which $g''(x_0) \neq 0$. Denote

$$r(x) = \frac{2(g(x+x_0) - g(x_0) - x g'(x_0))}{g''(x_0)}.$$

It is obvious that $r \in \mathbf{E}$, $r(0) = 0$, $r'(0) = 0$, $r''(0) = 2$, $r(x) = x^2 + o(x^2)$. Therefore,

$$\frac{1}{\varepsilon^2} r(\varepsilon x) \to x^2, \qquad \text{with } \varepsilon \to 0.$$

LEMMA 3. *Let, under the conditions of Theorem 1, the function* $q(x) = x^2$ *be in* $\mathbf{E}$. *Then* $\mathbf{E}$ *is a ring: for every* $f, g \in \mathbf{E}$, *their product* $fg \in \mathbf{E}$.

PROOF. Under the conditions of the lemma, $f^2 \in \mathbf{E}$ holds for every $f \in \mathbf{E}$ ($\mathbf{E}$ is a semigroup). What follows is obvious:

$$fg = \frac{1}{2}\left[(f+g)^2 - f^2 - g^2\right].$$

From Lemmas 1–3, it follows that under the conditions of Theorem 1, if $\mathbf{E}$ has even one nonlinear function, then $\mathbf{E}$ is a ring, and contains, in particular, all polynomials. Hence, by the Weierstrass theorem, it follows that $\mathbf{E} = \mathbf{C}(\mathbf{R})$.

3. APPROXIMATION OF FUNCTIONS OF SEVERAL VARIABLES

Most important for applications are the functions of several variables. Let $\mathbf{C}(\mathbf{R}^n)$ be a space of continuous functions on $\mathbf{R}^n$ in the topology of uniform convergence on compact sets.

For every natural n, let there be set a closed subspace $\mathbf{E}_n \subseteq \mathbf{C}(\mathbf{R}^n)$ $(i = 1, \dots, n)$. Let us assume that this set $\{\mathbf{E}_n\}$ has the following properties:

(1) $1 \in \mathbf{E}_n$, for all n;
(2) $\mathbf{E}_n$ contains all linear functions on $\mathbf{R}^n$;
(3) if $f_1, f_2, \dots, f_k \in \mathbf{E}_n$ and $f \in \mathbf{E}_k$, then $f(f_1, f_2, \dots, f_k) \in \mathbf{E}_n$;
(4) there exists a function $f \in \mathbf{E}_1$ which is not linear.

THEOREM 2. *Under the conditions (1)–(4),* $\mathbf{E}_n = \mathbf{C}(\mathbf{R}^n)$, *for all* n.

PROOF. By Theorem 1, $\mathbf{E}_1 = \mathbf{C}(\mathbf{R})$. By analogy with Lemma 3, all $\mathbf{E}_n$ are rings. Hence, and with the Weierstrass theorem, the required equality follows: $\mathbf{E}_n = \mathbf{C}(\mathbf{R}^n)$, for all n.

COROLLARY. *Let the conditions (1)–(3) of Theorem 2 as well as the following condition* (4′) *hold:*

(4′) *for some* $n \geq 1$, *there exists a function* $f \in \mathbf{E}_n$ *which is not linear.*

Then $\mathbf{E}_n = \mathbf{C}(\mathbf{R}^n)$, *for all* n.

PROOF. The function $f \in \mathbf{E}_n$ from condition (4′) is not linear with restriction to at least one straight line in $\mathbf{R}^n$. Using the conditions (1)–(3) produce from it a nonlinear function in $\mathbf{R}$. The rest follows from Theorem 2.

The Stone-Weierstrass theorem deals with continuous functions on general compact spaces. The next theorem gives its generalization.

THEOREM 3. *Let* X *be a compact space,* $\mathbf{C}(X)$ *be a space of continuous functions on* X, $\mathbf{E}$ *be a closed subspace in* $\mathbf{C}(X)$, $1 \in \mathbf{E}$, *functions from* $\mathbf{E}$ *separate points in* X, *and* $f(g(x)) \in \mathbf{E}$ *for some nonlinear continuous function of one variable* f *and every* $g \in \mathbf{E}$. *Then* $\mathbf{E} = \mathbf{C}(X)$.

PROOF. Let $\mathbf{E}_1 = \{f \in \mathbf{C}(\mathbf{R}) \mid f(g(x)) \in \mathbf{E}, \text{ for all } g \in \mathbf{E}\}$. For $\mathbf{E}_1$, conditions of Theorem 1 hold. According to Theorem 1, $\mathbf{E}_1 = \mathbf{C}(\mathbf{R})$. In particular, $f(x) = x^2 \in \mathbf{E}_1$, and $\mathbf{E}$ is a ring. By the Stone theorem [1], $\mathbf{E} = \mathbf{C}(X)$.

4. DISCUSSION

Investigation of neural networks has complemented the Weierstrass and Stone theorems: if a semigroup of continuous functions of one variable is closed under linear operations and limit transitions, and contains constants, linear functions, and at least one nonlinear function, it contains all continuous functions.

In addition, the theorem on approximation of functions of several variables is valid: every continuous function of several variables can be approximated arbitrarily accurately using linear functions, superposition operation, and an arbitrary function of one variable.

When we can use superposition of functions, linear functions, and at least one arbitrary nonlinear continuous function of one variable, we can approximate every continuous function of several variables.

These theorems can be interpreted as statements about universal approximation properties of every nonlinearity: linear operations and cascade combinations can be used to produce from arbitrary nonlinear elements every required result with preassigned accuracy.

The proofs given are based on the Weierstrass theorem. It would be of interest to derive "purely semigroup" proofs without this theorem, and what's more—without support of the fact that a linearly closed and topologically closed semigroup of continuous functions is a ring.

REFERENCES

1. M.N. Stone, The generalized Weierstrass approximation theorem, *Math. Mag.* **21**, 167–183, 237–254 (1948).
2. A.N. Kolmogorov, On representation of continuous functions of several variables by superpositions of continuous functions of smaller number of variables, *Dokl. AN SSSR* **108** (2), 179–182 (1956).
3. V.I. Arnold, On the functions of three variables, *Dokl. AN SSSR* **114** (4), 679–681 (1957).
4. A.N. Kolmogorov, On representation of continuous functions of several variables as superpositions of continuous functions of one variable, *Dokl. AN SSSR* **114** (5), 953–956 (1957).
5. R. Pfeifer, Z. Schreter, F. Fogelman-Soulie and L. Steels, Editors, *Connectionism in Perspective*, North-Holland, (1989).
6. A.N. Gorban, *Training Neural Networks*, USSR-USA JV "ParaGraph", Moscow, (1990); English translation, *AMSE Transaction, Scientific Siberian,* **A**, Volume 6, Neurocomputing, pp. 1–134, Tassin, France, (1993).
7. J.M. Zurada, *Introduction to Artificial Neural Systems*, PWS Publishing, (1992).
8. G. Cybenko, Approximation by superposition of a sigmoidal function, *Mathematics of Control, Signals, and Systems* **2**, 303–314 (1989).
9. K. Hornik, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks* **2**, 359–366 (1989).
10. D.A. Kochenov and D.A. Rossiev, Approximations of functions of $C[A, B]$ class by neural-net predictors (architectures and results), *AMSE Transaction, Scientific Siberian,* **A**, Volume 6, Neurocomputing, pp. 189–203, Tassin, France, (1993).
11. J. Wiedermann, On the computation efficiency of symmetric neural networks, *Theor. Comput. Sci.* **80** (2), 337–345 (1991).
12. S.E. Gilev and A.N. Gorban, On completeness of the class of functions computable by neural networks, *Proc. of the World Congress on Neural Networks (WCNN'96)*, Sept. 15–18, 1996, San Diego, CA, pp. 984–991, Lawrens Erlbaum Accociates, (1996).