# Kolmogorov's Mapping Neural Network Existence Theorem

Robert Hecht-Nielsen
Hecht-Nielsen Neurocomputer Corporation
5893 Oberlin Drive
San Diego, CA 92121
619-546-8877

Dedicated to Andrei Nikolaevic Kolmogorov

Abstract

An improved version of Kolmogorov's powerful 1957 theorem concerning the representation of arbitrary continuous functions from the n-dimensional cube to the real numbers in terms of one dimensional continuous functions is reinterpreted to yield an existence theorem for mapping neural networks.

## 1 Introduction

In 1957 Soviet mathematician A.N. Kolmogorov published an astounding theorem concerning the representation of arbitrary continuous functions from the n-dimensional cube to the real numbers in terms of one dimensional functions [2]. This theorem intrigued a number of mathematicians and over the next twenty years several improvements to it were discovered, notably by G. G. Lorentz [3,5].

The Kolmogorov theorem was discovered during a friendly mathematical duel between Kolmogorov and fellow Soviet mathematician V. I. Arnol'd in which they each tried to be the first to put to rest the remaining questions surrounding the 13th problem of Hilbert (a prominent mathematician who, at the turn of the century, announced a list of difficult problems for 20th century mathematicians to solve). In a series of papers in the mid to late 1950's Kolmogorov and Arnol'd fought their battle, each trying to one-up the other in successive papers. Kolmogorov won. His result was a mathematical supernova.

Although Kolmogorov's theorem was both powerful and shocking (many mathematicians do not believe it can be true when they first see it), it has not been found to be of much utility in terms of its use in proving other important theorems. In mathematical terms, no one has found a significant use for it. The point of this paper is that this is *not* the case in neurocomputing!

## 2 Mapping Networks

The most popular neural network architecture in use today is the Rumelhart backpropagation network [4]. This is an example of a ***mapping neural network***, so named because the information processing operation carried out is the approximation of a mathematical mapping or function $\phi$ from vectors $\mathbf{x}$ to vectors $\mathbf{y}$, with $\mathbf{y} = \phi(\mathbf{x})$. Another type of mapping network is the Counterpropagation network [1]. Mapping networks play an important role in neurocomputing applications because of their ability to self-organize new information processing algorithms that (at least in some cases) could not otherwise be affordably developed.

## 3 Kolmogorov's Theorem

In this section an improved version of Kolmogorov's theorem due to Sprecher is reexpressed as a result concerning the existence of mapping neural networks. The theorem follows:

**Kolmogorov's Mapping Neural Network Existence Theorem:** Given any continuous function
$\phi : I^n \longrightarrow \mathbf{R}^m$ , $\phi(\mathbf{x}) = \mathbf{y}$, where $I$ is the closed unit interval $[0,1]$ (and therefore $I^n$ is the $n$-dimensional unit cube), $\phi$ can be implemented ***exactly*** by a three-layer neural network having n processing elements in the first (x - input) layer, (2n+1) processing elements in the middle layer, and m processing elements in the top (y - output) layer (see Figure 1).

The processing elements on the bottom layer are fanout units that simply distribute the input x-vector components to the processing elements of the second layer.

The processing elements of the second layer implement the following transfer function:

$$z_k = \sum_{j=1}^{n} \lambda^k \psi(x_j + \epsilon k) + k$$

where the real constant $\lambda$ and the continuous real monotonic increasing function $\psi$ are independent of $\phi$ (although they do depend on $n$) and the constant $\epsilon$ is a rational number $0 < \epsilon \leq \delta$, where $\delta$ is an arbitrarily chosen positive constant. Further, it can be shown that $\psi$ can be chosen to satisfy a Lipschitz condition $|\psi(x) - \psi(y)| \leq c|x - y|^\alpha$ for any $0 < \alpha \leq 1$.

The $m$ top layer processing elements have the following transfer functions:

$$y_i = \sum_{k=1}^{2n+1} g_i(z_k)$$

where the functions $g_i$ $i = 1, 2, \ldots, m$ are real and continuous (and depend on $\phi$ and $\epsilon$.

**Proof:** Apply the result of Sprecher [5] to each of the $m$ coordinates of $\mathbf{y}$ separately.

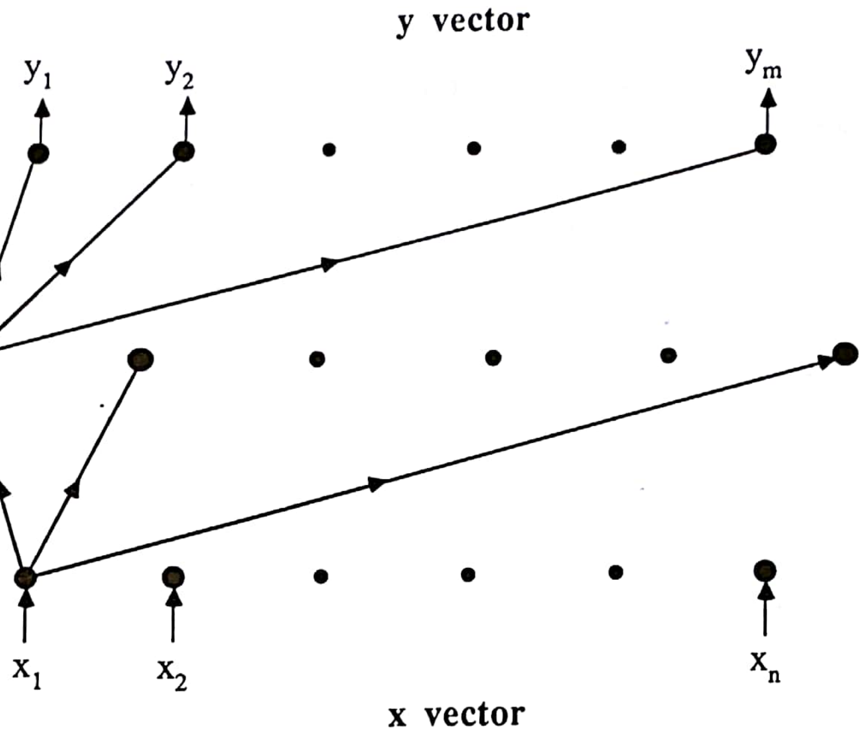


Figure 1: Topology of the Kolmogorov Network

# 4 Implications for Neurocomputing

Kolmogorov's Mapping Neural Network Existence Theorem is a statement that our quest for approximations of functions by networks is, at least in theory, sound. The above form of the Kolmogorov theorem is particularly nice since it clearly identifies all of the ingredients (as opposed to other forms of the theorem in which the existence of the required constants and/or functional expressions is established but their exact form is not specified).

The direct usefulness of this result is doubtful, at least in the near term, because ~~no constructive method for developing the $g_i$ functions is known~~. However, it is likely that more will be learned about this form of mapping network in the years to come. A potentially high-payoff challenge is to discover an adaptive mechanism whereby the $g_i$'s could self-organize themselves in response to incoming example x and y vector pairs.

# References


[1] Hecht-Nielsen, Robert, **Counterpropagation Networks**, *Proc. IEEE International Conference on Neural Networks - 1987.*

[2] Kolmogorov, Andrei Nikolaevic, "On the Representation of Continuous Functions of Many Variables by Superposition of Continuous Functions of One Variable and Addition", Dokl. Akad. Nauk USSR, **114**, 953-956, 1957.

[3] Lorentz, G.G., **The 13-th Problem of Hilbert**, *Proc. of Symposia in Pure Math.*, 28, American Mathematical Society, 1976.

[4] Rumelhart, David E., and McClelland, James L., **Parallel Distributed Processing: Explorations in the Microstructure of Cognition**, Vols. I & II, MIT Press, 1986.

[5] Sprecher, David A., "On the Structure of Continuous Functions of Several Variables", *Trans. Amer. Math. Soc.*, **115**, 340-355, March 1965.