Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Computational and Applied Mathematics 188 (2006) 283–308

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

Constructive approximate interpolation by neural networks

B. Llanas*, F.J. Sainz

Departamento de Matemática Aplicada, ETSI de Caminos, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain

Received 1 April 2004; received in revised form 7 March 2005

Abstract

We present a type of single-hidden layer feedforward neural networks with sigmoidal nondecreasing activation function. We call them ai-nets. They can approximately interpolate, with arbitrary precision, any set of distinct data in one or several dimensions. They can uniformly approximate any continuous function of one variable and can be used for constructing uniform approximants of continuous functions of several variables. All these capabilities are based on a closed expression of the networks.

Keywords: Neural networks; Approximate interpolation; Uniform approximation

1. Introduction

An interpolation problem is given by a set of ordered pairs

$$(\mathbf{x}_0, f_0), (\mathbf{x}_1, f_1), \ldots, (\mathbf{x}_n, f_n), \tag{1}$$

where $S = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset R^d$ is a set of distinct vectors and $\{f_i / i = 0, \ldots, n\}$ is a set of real numbers.

We say that the function $f : R^d \to R$ is an *exact interpolant* of (1) if

$$f(\mathbf{x}_i) = f_i \quad (i = 0, 1, \ldots, n).$$

* Corresponding author.
E-mail address: ma07@caminos.upm.es (B. Llanas).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2005.04.019

Consider the following set of functions

$$\mathcal{N}^d_{n+1,\phi} = \left\{ N(\mathbf{x}) = \sum_{j=0}^{n} c_j \phi(\mathbf{w}_j.\mathbf{x} + b_j)/\mathbf{w}_j \in R^d, c_j, b_j \in R \right\},$$

where $\mathbf{w}.\mathbf{x}$ denotes the usual inner product of R^d and ϕ is a function from R to R. We call

- single layer feedforward networks to the elements of $\mathcal{N}^d_{n+1,\phi}$,
- activation function to ϕ,
- inner weights to the coefficients $\mathbf{w}_j$ and b_j,
- outer weights to the coefficients c_j.

When ϕ is a sigmoidal function, that is, it verifies $\lim_{t\to-\infty}\phi(t) = 0$ and $\lim_{t\to\infty}\phi(t) = 1$, we call

- single layer feedforward neural networks to the elements of $\mathcal{N}^d_{n+1,\phi}$,
- neuron to each addend of $N(\mathbf{x})$.

It is well known that single layer feedforward networks with at most $n+1$ addends (elements of $\mathcal{N}^d_{n+1,\phi}$) can learn $n+1$ distinct samples $(\mathbf{x}_i, f_i)$ with zero error (exact interpolants) and the inner weights can be choosen "almost" arbitrarily.

Several proofs of this fact have been proposed

- Proofs of analytical type are given in [12–14,20,24]. Ito and Saito [14] prove that if the activation function is sigmoidal, continuous and nondecreasing, then the interpolation can be made with inner weights $\mathbf{w}_j \in S^{d-1}$ ($S^{d-1} \equiv \{\mathbf{x} \in R^d / x_1^2 + x_2^2 + \cdots + x_d^2 = 1\}$).
  In [20] Pinkus proves the same result but ϕ only needs to be continuous in R and not a polynomial.
- Algebraic proofs which have a more constructive character can be found in [11,22,23,26]. Shrivastava and Dasgupta [23] give a proof for logistic sigmoidal activation functions and then, they prove the theorem (by means of an asymptotic procedure) for activations functions ϕ, such that $\phi(0) \neq 0$ and ϕ^{-1} exists.
  Sartori and Antsaklis [22] consider that ϕ is any nonlinear function. Huang and Babri [11] criticize the above proof because they consider that it has a less general validity. They give a proof for arbitrary bounded nonlinear activation functions which have a limit at one infinity, that is, there exists $\lim_{t\to\infty}\phi(t)$ or $\lim_{t\to-\infty}\phi(t)$.

With regard to the problem of finding the weights of the network, all the algebraic proofs fix the inner weights in a more or less "arbitrary way" and the outer weights are obtained by solving a $(n+1)\times(n+1)$ linear system.

Other direct approaches for obtaining the weights are more cumbersome, for example, Barhen et al. [3] connect a virtual input layer to the nominal input layer by a special nonlinear transfer function and to the first layer by regular (linear) synapses. A sequence of alternating direction singular value decompositions is then used to determine precisely the interlayer weights.

Li [17], considers that ϕ is m-times continuously differentiable on an interval containing θ and $\phi^{(k)}(\theta) \neq 0$ $(0 \leqslant k \leqslant m-1)$. He gives an algorithm to approximately interpolate n samples that consists of the following steps:

- Solve a $(n+1) \times (n+1)$ linear system.
- Find the inverse of a $(n+1) \times (n+1)$ Vandermonde matrix.
- Estimate the maximum of a function on an interval.

The procedures cited above can be scarcely effective, especially, when the number of neurons is large. Even the most simple (algebraic) methods require to invert an $(n+1) \times (n+1)$ dense matrix. Therefore, we can state the following problem:

Problem I. *Is there a way to find the weights of an exact neural interpolant without training?*

Let ε be a positive real number, we say that the function $g : R^d \to R$ is an *ε-approximate interpolant* of (1) if

$$|g(\mathbf{x}_i) - f_i| < \varepsilon \quad (i = 0, 1, \ldots, n).$$

Given an arbitrary function g, we define the *interpolation error* of g relating to problem (1) as

$$\max_{i=0,1,\ldots,n} |g(\mathbf{x}_i) - f_i|.$$

Approximate interpolants are used in [24] as a tool for studying exact interpolation. In this reference it is proved that if arbitrary precision approximate interpolants exist in a linear space of functions, then an exact interpolant can be obtained in that space (Lemma 9.1). Furthermore, Sontag gives a proof of the following fact:

"If ϕ is sigmoidal, continuous and there exits a point c such that $\phi'(c) \neq 0$, then an interpolation problem with $2n+1$ samples can be approximated with arbitrary precision by a net with $n+1$ neurons" (Proposition 9.4).

Although this proof (based on the induction principle) can be made constructive, the determination of the weights requires to solve n times a set of one variable equations, in the most favorable case.

On the other hand, most of the methods of approximation by neural networks are reduced to the search of approximate interpolants. These are usually obtained by

- local optimization techniques (gradient descent, second order training methods,...),
- global optimization procedures (stochastic optimization, methods based on homotopy,...).

The connection between approximate interpolants and uniform approximation in the context of Probabilistic (Computational) Learning Theory can be found in [1,2].

We can state the following problem:

Problem II. *Existence and construction of neural ε-approximate interpolants for arbitrary values of ε.*

It is clear that a solution to this problem gives a procedure for solving the Problem I in practice.

In this paper we give a solution to Problem II, that is, for any interpolation problem (1) and for any precision ε, we show a constructive method for obtaining a family of ε-approximate interpolants in $\mathscr{N}^d_{n+1,\phi}$ (if ϕ is sigmoidal and nondecreasing).

This paper is organized in six sections. Section 2 gives a new and quantitative proof of the fact that $n+1$ hidden neurons can learn $n+1$ distinct samples with zero error. Based on this result, Section 3 introduces the approximate interpolation nets (ai-nets) in the unidimensional case. These networks do not require training and can approximately interpolate an arbitrary set of distinct samples. We give a rigorous upper bound of the interpolation error and show that it basically depends on a real parameter (A) and the number of neurons $n+1$ of the network.

The remaining sections show some applications of ai-nets.

Section 4 gives a constructive proof of the uniform convergence of ai-nets to any continuous function of one variable (if ϕ is the logistic sigmoidal function). Lemma 8 and Theorem 3 describe such a neural network whose weights are explicitly known. Although Cardaliaguet et al. [4] have also proposed approximants with similar features, our approach presents some advantages:

- It provides a genuine neural network, that is, the activation function is sigmoidal. In [4] the activation functions are "bell-shaped" and even with the transformation proposed in Section 3.1.4 the result is a neural network with additional terms that depend on the number of neurons.
- Obtaining the number of neurons n as a function of the precision ε is an easy task in Lemma 8, but it is much more difficult in [4].

Although other constructive proofs of the Cybenko's theorem, with more general activation functions, have been proposed [5,8,19], we give a more general approximation operator obtained from ai-nets. In this way, we show a connection between interpolation and approximation in the context of neural networks.

Corollary 3 provides a procedure of uniform approximation of functions of several variables, based on ai-nets.

Section 5 extends the concept of ai-net to multidimensional domains.

Section 6 provides some concluding remarks.

2. Existence of exact neural interpolants (ei-nets)

Let $\mathscr{P}=\{x_0=a, x_1, \ldots, x_n=b\}$ be any partition of the finite interval $[a,b]\subset R$. We want to find a neural net N that exactly interpolates the data

$$(x_0, f_0), (x_1, f_1), \ldots, (x_n, f_n),$$

where $f_i \in R\ (i=0,1,\ldots,n)$, that is

$$N(x_i)=f_i \quad (i=0,1,\ldots,n).$$

Define the following quantities

$$
\begin{aligned}
&\Delta_0 \equiv 0,\\
&\Delta_i \equiv x_i - x_{i-1} \quad (i = 1, \dots, n),\\
&M \equiv \max_{i=1,\dots,n} \Delta_i,\\
&m \equiv \min_{i=1,\dots,n} \Delta_i,\\
&r \equiv m/M,\\
&\Delta_{ij} \equiv \Delta_i + \Delta_{i+1} + \cdots + \Delta_j \quad (\text{if } i < j),\\
&\Delta_{ii} \equiv \Delta_i \quad (i = 0, 1, \dots, n).
\end{aligned}
$$

From the above definitions, we have

$$
\begin{aligned}
&0 < r \leqslant 1,\\
&\frac{\Delta_i}{\Delta_j} \geqslant r \quad (i, j = 1, \dots, n).
\end{aligned}
\tag{2}
$$

From now on, we consider that $\sigma(t)$ is a sigmoidal nondecreasing function, that is

1. $\lim_{t\to\infty} \sigma(t) = 1$.
2. $\lim_{t\to-\infty} \sigma(t) = 0$.
3. If $t_1 \leqslant t_2$, then $\sigma(t_1) \leqslant \sigma(t_2)$.

The third condition seems scarcely restrictive in practice, in fact [9] gives seven examples of sigmoidal functions and all of them are nondecreasing.

We look for a function of type

$$N(x) = \sum_{j=0}^{n} c_j \sigma(w_j x + b_j), \tag{3}$$

such that

$$\sum_{j=0}^{n} c_j \sigma(w_j x_i + b_j) = f_i \quad (i = 0, 1, \dots, n). \tag{4}$$

Or, in vectorial form

$$M\mathbf{c} = \mathbf{f}.$$

The inner weights of the network are determined by the following procedure:

Let A be a positive real variable. Let α and β be positive and negative constants, respectively. For $j = 0, \dots, n-1$ we consider the following systems

$$
\begin{aligned}
&w_j x_j + b_j = \alpha A,\\
&w_j x_{j+1} + b_j = \beta A,
\end{aligned}
$$

and

$$\begin{aligned} w_n x_{n-1} + b_n &= (2\alpha - \beta)A, \\ w_n x_n + b_n &= \alpha A. \end{aligned}$$

That is, for each neuron j (except the last one) we can choose its value in x_j near c_j and its value in x_{j+1} near 0 when A is large enough. The solution of the above systems is as follows:

$$\begin{aligned} w_j &= \frac{\beta - \alpha}{x_{j+1} - x_j} A \quad (j = 0, 1, \ldots, n-1), \\ b_j &= \frac{\alpha x_{j+1} - \beta x_j}{x_{j+1} - x_j} A \quad (j = 0, 1, \ldots, n-1), \\ w_n &= \frac{\beta - \alpha}{x_n - x_{n-1}} A, \\ b_n &= \frac{(2\alpha - \beta) x_n - \alpha x_{n-1}}{x_n - x_{n-1}} A. \end{aligned}$$

A neural network of type (3), that satisfies (4) and with inner weights given by the above expressions, will be called *ei-net* (exact interpolation net) for the points $((x_0, f_0), (x_1, f_1), \ldots, (x_n, f_n))$. From now on, we denote it by $N_e(x, A)$.

If we replace each x_j by its concrete value we have

$$\begin{aligned} w_j &= \frac{\beta - \alpha}{\Delta_{j+1}} A \quad (j = 0, 1, \ldots, n-1), \\ b_j &= \left(\alpha + \frac{(\alpha - \beta)(a + \Delta_0 + \Delta_1 + \cdots + \Delta_j)}{\Delta_{j+1}} \right) A \quad (j = 0, 1, \ldots, n-1), \\ w_n &= \frac{\beta - \alpha}{\Delta_n} A, \\ b_n &= \left(\alpha + \frac{(\alpha - \beta) b}{\Delta_n} \right) A. \end{aligned}$$

Hence, the entries of the matrix M are

$$\begin{aligned} M_{in} &= \sigma\left(\left(\alpha + (\alpha - \beta) \frac{\Delta_{i+1\,n}}{\Delta_n} \right) A \right) \quad (i = 0, 1, \ldots, n-1), \\ M_{ij} &= \sigma\left(\left(\alpha + (\alpha - \beta) \frac{\Delta_{i+1\,j}}{\Delta_{j+1}} \right) A \right) \quad (i < j < n), \\ M_{ii} &= \sigma(\alpha A) \quad (i = 0, 1, \ldots, n), \\ M_{i+1\,i} &= \sigma(\beta A) \quad (i = 0, 1, \ldots, n-1), \\ M_{ij} &= \sigma\left(\left(\beta + (\beta - \alpha) \frac{\Delta_{j+2\,i}}{\Delta_{j+1}} \right) A \right) \quad (j \leqslant i - 2, \ i = 2, \ldots, n). \end{aligned}$$

Let us now introduce some previous definitions and lemmas.

If S is a square matrix we denote by $\det(S)$ its determinant and by S^t its transpose.

Lemma 1. *Let $S = (\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_n)$ be a square matrix with column vectors $\mathbf{s}_j$. If we define*

$$S^d \equiv (\mathbf{s}_0 - \mathbf{s}_1, \mathbf{s}_1 - \mathbf{s}_2, \ldots, \mathbf{s}_{n-1} - \mathbf{s}_n, \mathbf{s}_n),$$

we have that

$$\det(S) = \det(S^d).$$

Proof. Apply the properties of the determinants. □

A square matrix $S = (S_{ij})$ is strictly diagonally dominant if

$$|S_{ii}| > \sum_{\substack{j=0 \\ j \neq i}}^{n} |S_{ij}|$$

for $i = 0, 1, \ldots, n$.

Lemma 2 (*Lascaux and Theodor [16, p. 70]*). *Strictly diagonally dominant matrices are invertible.*

Lemma 3. *Let $a_0, a_1, \ldots, a_n$ be $n+1$ arbitrary real numbers. We define $a_{n+1} \equiv 0$ and the finite sequence*

$$a_0 - a_1, a_1 - a_2, \ldots, a_{n-1} - a_n, a_n - a_{n+1}.$$

If $k \in \{0, 1, \ldots, n\}$ we have

$$a_k - a_{k+1} > \sum_{\substack{i=0 \\ i \neq k}}^{n} (a_i - a_{i+1}),$$

if and only if

$$a_k > a_{k+1} + \frac{a_0}{2}. \tag{5}$$

Proof. If $k = 0$,

$$\sum_{i=1}^{n} (a_i - a_{i+1}) = a_1.$$

If $k = n$,

$$\sum_{i=0}^{n-1} (a_i - a_{i+1}) = a_0 - a_n.$$

If $n \neq 0$ and $k \neq 0$,

$$\sum_{\substack{i=0 \\ i \neq k}}^{n} (a_i - a_{i+1}) = a_0 - a_k + a_{k+1}.$$

The second term in the above equations is lower than $a_k - a_{k+1}$ if and only if (5) holds. □

We can now state

Theorem 1. *There exists a real number A^* such that if $A > A^*$, then the matrix M is invertible.*

Proof. From now on, we shall denote $\alpha - \beta$ by ω.

From Lemmas 1 and 2 it is enough to study when $(M^t)^d$ is strictly diagonally dominant.

As the rows of the matrix M^t are strictly decreasing finite sequences we can apply Lemma 3 to determine the conditions on A, α and β that make $(M^t)^d$ strictly diagonally dominant.

The following $n + 1$ conditions must hold

$$\begin{array}{rcl}
\sigma(\alpha A) & > & 2\sigma(\beta A), \\
2\sigma(\alpha A) & > & 2\sigma(\beta A) + \sigma\left(\left(\alpha + \omega\frac{\Delta_1}{\Delta_2}\right) A\right), \\
2\sigma(\alpha A) & > & 2\sigma(\beta A) + \sigma\left(\left(\alpha + \omega\frac{\Delta_{12}}{\Delta_3}\right) A\right), \\
\cdots & \cdots & \cdots, \\
2\sigma(\alpha A) & > & 2\sigma(\beta A) + \sigma\left(\left(\alpha + \omega\frac{\Delta_{1n-1}}{\Delta_n}\right) A\right), \\
2\sigma(\alpha A) & > & \sigma\left(\left(\alpha + \omega\frac{\Delta_{1n}}{\Delta_n}\right) A\right).
\end{array}$$

Define

$$\begin{array}{rcl}
g(A) & \equiv & 2\sigma(\alpha A), \\
g_0(A) & \equiv & 4\sigma(\beta A), \\
g_1(A) & = & 2\sigma(\beta A) + \sigma\left(\left(\alpha + \omega\frac{\Delta_1}{\Delta_2}\right) A\right), \\
\cdots & \cdots & \cdots, \\
g_{n-1}(A) & \equiv & 2\sigma(\beta A) + \sigma\left(\left(\alpha + \omega\frac{\Delta_{1n-1}}{\Delta_n}\right) A\right), \\
g_n(A) & \equiv & \sigma\left(\left(\alpha + \omega\frac{\Delta_{1n}}{\Delta_n}\right) A\right).
\end{array}$$

We have $\lim_{A\to\infty} g(A) = 2$, $\lim_{A\to\infty} g_0(A) = 0$ and $\lim_{A\to\infty} g_k(A) = 1$ $(k = 1, \ldots, n)$.

From the definition of limit with

- $\varepsilon = \frac{1}{2}$, there exists A' such that if $A > A'$, then $g(A) > \frac{3}{2}$.
- $\varepsilon = \frac{3}{2}$, there exists A_0 such that if $A > A_0$, then $g_0(A) < \frac{3}{2}$.
- $\varepsilon = \frac{1}{2}$, there exists A_k such that if $A > A_k$, then $g_k(A) < \frac{3}{2}$ for $k = 1, \ldots, n - 1$.

Hence, if we take $A^* = \max(A', A_0, A_1, \dots, A_{n-1})$, it follows that if $A > A^*$, then

$$g(A) > \tfrac{3}{2} > g_k(A) \quad (k = 0, 1, \dots, n). \qquad \square$$

In the case of the logistic sigmoidal activation function

$$s(x) = \frac{1}{1 + \mathrm{e}^{-x}},$$

a sufficient condition for the existence of ei-nets is:

Corollary 1. *If the activation function is the logistic sigmoidal, and the following inequality*

$$\mathrm{e}^{\alpha A}\left(\frac{1 - \mathrm{e}^{\beta A}}{1 + 3\mathrm{e}^{\beta A}}\right) > 1.$$

holds, then M is invertible.

Proof. The $n + 1$ conditions of Theorem 1 hold if the following inequalities

$$\sigma(\alpha A) > 2\sigma(\beta A), \tag{6}$$
$$2\sigma(\alpha A) > 2\sigma(\beta A) + \sigma((\alpha + \omega v)A) \quad (\text{for all } v > 0), \tag{7}$$
$$2\sigma(\alpha A) > \sigma((\alpha + \omega \tau)A) \quad (\text{for all } \tau \geqslant 1) \tag{8}$$

are satisfied.

If $\sigma(x) = s(x)$, (6) can be written as

$$\frac{1}{1 + \mathrm{e}^{-\alpha A}} > \frac{2}{1 + \mathrm{e}^{-\beta A}},$$

which is equivalent to

$$\mathrm{e}^{\omega A} - \mathrm{e}^{\alpha A} - 2 > 0. \tag{9}$$

Inequality (7) can be written as

$$\frac{2}{1 + \mathrm{e}^{-\alpha A}} > \frac{2}{1 + \mathrm{e}^{-\beta A}} + \frac{1}{1 + \mathrm{e}^{-(\alpha + \omega v)A}},$$

this is equivalent to

$$\mathrm{e}^{\omega A} - \mathrm{e}^{\alpha A} - \mathrm{e}^{-\beta A} + \frac{2}{\mathrm{e}^{\omega v A}}(\mathrm{e}^{-\beta A} - \mathrm{e}^{-\alpha A}) - 3 > 0. \tag{10}$$

If the inequalities

$$\mathrm{e}^{\omega A} - \mathrm{e}^{\alpha A} - \mathrm{e}^{-\beta A} - 3 > 0, \tag{11}$$

and

$$\mathrm{e}^{-\beta A} - \mathrm{e}^{-\alpha A} > 0, \tag{12}$$

both hold, then (10) is satisfied.

Inequality (8) can be written as

$$\frac{2}{1+e^{-\alpha A}} > \frac{1}{1+e^{-(\alpha+\omega\tau)A}}.$$

As $A > 0$, this inequality always holds, in effect

$$2 + 2e^{-\alpha A}e^{-\omega\tau A} > 2 > 1 + e^{-\alpha A}.$$

Now, if

$$e^{\alpha A}\left(\frac{1-e^{\beta A}}{1+3e^{\beta A}}\right) > 1,$$

we have

$$e^{\alpha A}(e^{-\beta A} - 1) > e^{-\beta A} + 3,$$

hence

$$e^{\alpha A}(e^{-\beta A} - 1) - 2 > 1 + e^{-\beta A}.$$

This implies (9) and (11).

Finally, as $A > 0$

$$e^{-\beta A} - e^{-\alpha A} > e^{-\beta A} - 1 > 0,$$

and (12) follows. □

Notes

1. In the common case in that $\alpha = 1$ and $\beta = -1$, the second inequality of Corollary 1 can be written as

$$e^{A}\left(\frac{1+e^{-A}}{1+3e^{-A}}\right) > 1.$$

If we make the change of variable $y = e^{A}$, this is equivalent to

$$y^2 - 2y - 3 > 0.$$

Therefore, if $A > ln(3)$ then, M is invertible.

2. The procedure for determining the inner weights is a modification of the method proposed in [11]. We could obtain a more general formulation considering the systems

$$w_j x_j + b_j = g_1(A),$$
$$w_j x_{j+1} + b_j = g_2(A),$$

for $j = 0, \ldots, n-1$, and

$$w_n x_{n-1} + b_n = 2g_1(A) - g_2(A),$$
$$w_n x_n + b_n = g_1(A),$$

where g_1 and g_2 verify $\lim_{A\to+\infty} g_1(A) = +\infty$ and $\lim_{A\to+\infty} g_2(A) = -\infty$.

As the above conditions guarantee that $g_1(A) > g_2(A)$ in an interval $(A_p, +\infty)$, for some $A_p \geqslant 0$, the proof of Theorem 1 remains valid with slight changes.

3. Unidimensional neural approximate interpolants

Let $\mathscr{P}=\{x_0=a, x_1, \ldots, x_n=b\}$ be any partition of the interval $[a, b] \subset R$. Consider the interpolation problem

$$(x_0, f_0), (x_1, f_1), \ldots, (x_n, f_n).$$

Define the following neural network

$$\begin{aligned} N_a(x, A) \equiv & \sum_{j=0}^{n-1} (f_j - f_{j+1}) \sigma \left(\frac{(\beta-\alpha)Ax}{x_{j+1}-x_j} + \frac{\alpha x_{j+1} - \beta x_j}{x_{j+1}-x_j} A \right) \\ & + f_n \sigma \left(\frac{(\beta-\alpha)Ax}{x_n - x_{n-1}} + \frac{(2\alpha-\beta)x_n - \alpha x_{n-1}}{x_n - x_{n-1}} A \right), \end{aligned} \tag{13}$$

where α and β are positive and negative real constants respectively.

This neural network will be called an *ai-net* for the points $((x_0, f_0), (x_1, f_1), \ldots, (x_n, f_n))$. We prove in this section that ai-nets are approximate interpolants that are arbitrarily near of the corresponding ei-net (same A, α and β) when the parameter A increases. We note that $N_a(x, A)$ and $N_e(x, A)$ differ only in the outer weights, which are given explicitly by (13) in the case of an ai-net.

Let us now introduce some previous definitions and lemmas.

In C^{n+1} we introduce the norm

$$\|\mathbf{z}\|_1 = |z_0| + |z_1| + \cdots + |z_n|,$$

where $|z| = \sqrt{z\bar{z}}$ ($\bar{z}$ is the conjugate of the complex number z).

In the space of complex $(n+1) \times (n+1)$ matrices the induced matrix norm is

$$\|B\|_1 \equiv \max_{j=0,\ldots,n} \left(\sum_{i=0}^{n} |b_{ij}| \right).$$

Lemma 4 (*Kress [15, p. 80]*). *Let B and $B + \Delta B$ be complex, invertible matrices. Let $\mathbf{z}$ and $\mathbf{z} + \Delta\mathbf{z}$ be vectors such that*

$$B\mathbf{z} = \mathbf{b},$$
$$(B + \Delta B)(\mathbf{z} + \Delta\mathbf{z}) = \mathbf{b},$$

then

$$\frac{\|\Delta\mathbf{z}\|_1}{\|\mathbf{z}\|_1} \leqslant \mathrm{cond}_1(B) \frac{\|\Delta B\|_1}{\|B\|_1} \left(\frac{1}{1 - \mathrm{cond}_1(B) \dfrac{\|\Delta B\|_1}{\|B\|_1}} \right),$$

where $\mathrm{cond}_1(B) = \|B\|_1 \|B^{-1}\|_1$ *is the condition number of B.*

Define the functions

$$\delta_1(A) \equiv 1 - \sigma(\alpha A),$$
$$\delta_2(A) \equiv \sigma(\beta A),$$
$$\delta(A) \equiv \max(\delta_1(A), \delta_2(A)).$$

We have that $\lim_{A\to\infty} \delta(A) = 0$.

Theorem 2. *There exists a real number A_1 such that if $A > A_1$, then*

$$|N_e(x, A) - N_a(x, A)| < \frac{2(n+1)\delta(A)}{1 - 2(n+1)\delta(A)} \left(\sum_{j=0}^{n-1} |f_j - f_{j+1}| + |f_n| \right), \tag{14}$$

for all $x \in [a, b]$.

Proof. We shall consider that $A > A^*$ (Theorem 1).

The nets can abridgedly be written as

$$N_a(x, A) = \sum_{j=0}^{n} c'_j \sigma(w_j x + b_j),$$

and

$$N_e(x, A) = \sum_{j=0}^{n} c_j \sigma(w_j x + b_j).$$

The coefficients c'_j and c_j are, respectively, solutions of the systems

$$U\mathbf{c}' = \mathbf{f} \quad \text{and} \quad M\mathbf{c} = \mathbf{f},$$

where

$$U = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 & 1 \\ 0 & 1 & 1 & \dots & 1 & 1 & 1 \\ 0 & 0 & 1 & \dots & 1 & 1 & 1 \\ .. & .. & .. & \dots & .. & .. & .. \\ 0 & 0 & 0 & \dots & 1 & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix},$$

and

$$\mathbf{c}' = \begin{pmatrix} c'_0 \\ c'_1 \\ .. \\ c'_n \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_0 \\ c_1 \\ .. \\ c_n \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f_0 \\ f_1 \\ .. \\ f_n \end{pmatrix}.$$

By considering in Lemma 4

$$B = U, \quad \mathbf{z} = \mathbf{c}', \quad M = B + \Delta B, \quad \mathbf{z} + \Delta\mathbf{z} = \mathbf{c}, \quad \mathbf{b} = \mathbf{f},$$

we have

$$\frac{\|\mathbf{c}-\mathbf{c}'\|_1}{\|\mathbf{c}'\|_1} \leqslant \|U^{-1}\|_1\|M-U\|_1\left(\frac{1}{1-\|U^{-1}\|_1\|M-U\|_1}\right). \tag{15}$$

We need an upper bound of the expression

$$\|U^{-1}\|_1\|M-U\|_1.$$

From the explicit values of M_{ij} (Section 2), we have that if $A>0$ and $i=0,1,\ldots,n-1$

$$|M_{in}-1| = 1-\sigma\left(\left(\alpha+\frac{\omega\Delta_{i+1n}}{\Delta_n}\right)A\right) < 1-\sigma(\alpha A) \leqslant \delta(A),$$

if $i<j<n$

$$|M_{ij}-1| = 1-\sigma\left(\left(\alpha+\frac{\omega\Delta_{i+1\,j}}{\Delta_{j+1}}\right)A\right) < 1-\sigma(\alpha A) \leqslant \delta(A),$$

if $i=0,1,\ldots,n$

$$|M_{ii}-1| = 1-\sigma(\alpha A) \leqslant \delta(A),$$

if $i=0,1,\ldots,n-1$

$$|M_{i+1\,i}| = \sigma(\beta A) \leqslant \delta(A).$$

Finally, if $j\leqslant i-2$ and $i=2,\ldots,n$

$$|M_{ij}| = \sigma\left(\left(\beta-\frac{\omega\Delta_{j+2\,i}}{\Delta_{j+1}}\right)A\right) < \sigma(\beta A) \leqslant \delta(A).$$

Hence

$$\|M-U\|_1 < (n+1)\delta(A).$$

On the other hand

$$U^{-1} = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -1 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 & 0 \\ .. & .. & .. & \ldots & .. & .. & .. \\ 0 & 0 & 0 & \ldots & 1 & -1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 & -1 \\ 0 & 0 & 0 & \ldots & 0 & 0 & 1 \end{pmatrix},$$

hence

$$\|U^{-1}\|_1 = 2,$$

so that

$$\|U^{-1}\|_1\|M-U\|_1 < 2(n+1)\delta(A). \tag{16}$$

There exists a number A' such that if $A > A'$, then $2(n+1)\delta(A) < 1$. Then, from (15), (16) and considering that the function $g(x) = x/(1-x)$ is strictly increasing on $(-\infty, 1)$, it follows that

$$\|\mathbf{c} - \mathbf{c}'\|_1 < \frac{2(n+1)\delta(A)}{1 - 2(n+1)\delta(A)} \|\mathbf{c}'\|_1. \tag{17}$$

On the other hand

$$\begin{aligned} |N_e(x, A) - N_a(x, A)| &= \left| \sum_{j=0}^{n} (c_j - c'_j)\sigma(w_j x + b_j) \right| \\ &\leqslant \sum_{j=0}^{n} |c_j - c'_j| = \|\mathbf{c} - \mathbf{c}'\|_1. \end{aligned}$$

Finally, from (17) we have that if $A > A_1 = \max(A', A^*)$, then

$$|N_e(x, A) - N_a(x, A)| < \frac{2(n+1)\delta(A)}{1 - 2(n+1)\delta(A)} \left(\sum_{j=0}^{n-1} |f_j - f_{j+1}| + |f_n| \right). \qquad \square$$

If the values f_i are the images of x_i under a real function f, the second factor in (14) can be bounded as follows

- If f is of bounded variation on $[a, b]$

$$V_a^b(f) + |f(b)|.$$

- If f is continuously differentiable on $[a, b]$

$$\left(\max_{x \in [a,b]} |f'(x)| \right) (b - a) + |f(b)|.$$

- If f is Lipschitz on $[a, b]$ and L is a Lipschitz constant (Section 4)

$$L(b - a) + |f(b)|.$$

In the case of the logistic sigmoidal activation function we can find a more sharp bound. If we define $\gamma \equiv \min(|\alpha|, |\beta|)$ we have:

Corollary 2. *If $\sigma(x) = s(x)$, there exists a real number A_2 such that if $A > A_2$, then*

$$|N_e(x, A) - N_a(x, A)| < \frac{4\mathrm{e}^{-\gamma A}}{1 - \mathrm{e}^{-\omega r A} - 4\mathrm{e}^{-\gamma A}} \left(\sum_{j=0}^{n-1} |f_j - f_{j+1}| + |f_n| \right) \tag{18}$$

for all $x \in [a, b]$.

Proof. Consider that $A > A^* > 0$. By (2), $\frac{\Delta_i}{\Delta_j} \geqslant r$ for $i, j = 1, \dots, n$, and therefore, the sum of terms of the kth column of the matrix whose entries are the absolute value of the entries of $M - U$, $\mathscr{S}_k$ $(1 \leqslant k < n)$, can be bounded as follows

$$
\begin{aligned}
1 - s\left(\left(\alpha + \tfrac{\omega(\Delta_1 + \cdots + \Delta_k)}{\Delta_{k+1}}\right) A\right) &< \mathrm{e}^{-(\alpha + k\omega r)A},\\
1 - s\left(\left(\alpha + \tfrac{\omega(\Delta_2 + \cdots + \Delta_k)}{\Delta_{k+1}}\right) A\right) &< \mathrm{e}^{-(\alpha + (k-1)\omega r)A},\\
&\dots\dots,\\
1 - s\left(\left(\alpha + \tfrac{\omega\Delta_k}{\Delta_{k+1}}\right) A\right) &< \mathrm{e}^{-(\alpha + \omega r)A},\\
1 - s(\alpha A) &< \mathrm{e}^{-\alpha A},\\
s(\beta A) &< \mathrm{e}^{\beta A},\\
s\left(\left(\beta - \tfrac{\omega\Delta_{k+2}}{\Delta_{k+1}}\right) A\right) &< \mathrm{e}^{(\beta - \omega r)A},\\
s\left(\left(\beta - \tfrac{\omega(\Delta_{k+2} + \Delta_{k+3})}{\Delta_{k+1}}\right) A\right) &< \mathrm{e}^{(\beta - 2\omega r)A},\\
&\dots\dots,\\
s\left(\left(\beta - \tfrac{\omega(\Delta_{k+2} + \Delta_{k+3} + \cdots + \Delta_n)}{\Delta_{k+1}}\right) A\right) &< \mathrm{e}^{(\beta - (n-k-1)\omega r)A}.
\end{aligned}
$$

Hence

$$\mathscr{S}_k < \frac{\mathrm{e}^{-\alpha A}(1 - \mathrm{e}^{-(k+1)\omega r A})}{1 - \mathrm{e}^{-\omega r A}} + \frac{\mathrm{e}^{\beta A}(1 - \mathrm{e}^{-(n-k)\omega r A})}{1 - \mathrm{e}^{-\omega r A}} < \frac{\mathrm{e}^{-\alpha A} + \mathrm{e}^{\beta A}}{1 - \mathrm{e}^{-\omega r A}}.$$

As

$$\mathrm{e}^{-\alpha A} \leqslant \mathrm{e}^{-\gamma A}, \quad \mathrm{e}^{\beta A} \leqslant \mathrm{e}^{-\gamma A}.$$

We have

$$\mathscr{S}_k < \frac{2\mathrm{e}^{-\gamma A}}{1 - \mathrm{e}^{-\omega r A}}.$$

This bound is also valid for the columns 0 and n.
Therefore

$$\|M - U\|_1 < \frac{2\mathrm{e}^{-\gamma A}}{1 - \mathrm{e}^{-\omega r A}},$$

and

$$\|U^{-1}\|_1 \|M - U\|_1 < \frac{4\mathrm{e}^{-\gamma A}}{1 - \mathrm{e}^{-\omega r A}}.$$

There exists a real number A'' such that if $A > A''$, then

$$\frac{4\mathrm{e}^{-\gamma A}}{1 - \mathrm{e}^{-\omega r A}} < 1.$$

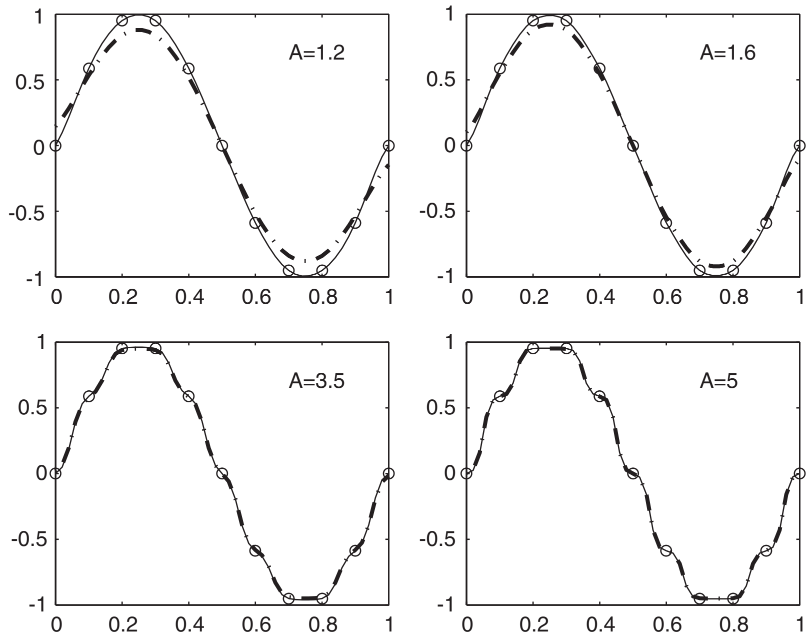

Fig. 1. N_a in the case of $\sin(2\pi x)$.

Therefore, if $A > A''$

$$\|\mathbf{c} - \mathbf{c}'\|_1 < \frac{4\mathrm{e}^{-\gamma A}}{1 - \mathrm{e}^{-\omega r A} - 4\mathrm{e}^{-\gamma A}} \|\mathbf{c}'\|_1,$$

and if $A > A_2 = \max(A'', A^*)$, then

$$|N_e(x, A) - N_a(x, A)| < \frac{4\mathrm{e}^{-\gamma A}}{1 - \mathrm{e}^{-\omega r A} - 4\mathrm{e}^{-\gamma A}} \left(\sum_{j=0}^{n-1} |f_j - f_{j+1}| + |f_n| \right). \qquad \square$$

Fig. 1 shows the graph of $N_e(x, A)$ (continuous line) and $N_a(x, A)$ (discontinuous line) as A increases. We have considered the interpolation problem (circles) of the function $\sin(2\pi x)$ on 11 points, the logistic sigmoidal activation function, $\alpha = 1$ and $\beta = -1$.

4. Uniform approximation by means of ai-nets

In this section we use the results of the previous one for obtaining constructive methods for the uniform approximation of continuous functions. To simplify matters, we restrict ourselves to the logistic sigmoidal activation function. Let us now introduce some definitions and auxiliary results.

Let f be a function defined on $[a, b]$. The *modulus of continuity* of f on $[a, b]$, is defined for $\delta > 0$ by

$$\omega(f, \delta) = \sup_{\substack{x, y \in [a,b] \\ |x-y| \leqslant \delta}} |f(x) - f(y)|.$$

Lemma 5 (*Rivlin [21, p. 14]*). *f is continuous on $[a, b]$ if and only if*

$$\lim_{\delta \to 0} \omega(f, \delta) = 0.$$

Lemma 6. *Let f be a continuous function on $[a, b]$. For each $\varepsilon > 0$ there exists a piecewise linear function l_ε such that*

$$|f(x) - l_\varepsilon(x)| < \varepsilon \quad \textit{for all } x \in [a, b].$$

Proof. From Lemma 5, for each $\varepsilon > 0$ there exists $\eta > 0$ such that $\omega(f, \delta) < \varepsilon$ if $\delta < \eta$. Consider a partition $\mathscr{P} = \{x_0, \ldots, x_n\}$ of $[a, b]$ such that $|x_{i+1} - x_i| < \eta$ for $i = 0, \ldots, n-1$ and define

$$l_\varepsilon((1-\theta)x_i + \theta x_{i+1}) = (1-\theta) f(x_i) + \theta f(x_{i+1}) \quad \text{for } \theta \in [0, 1].$$

Then, if $x = (1-\theta)x_i + \theta x_{i+1}$, we have

$$|f(x) - l_\varepsilon(x)| \leqslant (1-\theta)|f(x) - f(x_i)| + \theta |f(x) - f(x_{i+1})| < \varepsilon. \qquad \square$$

We consider in this section ai-nets, with the following features

- The partition of $[a, b]$ is uniform, that is, $x_j = a + \frac{b-a}{n} j \;\; (j = 0, 1, \ldots, n)$.
- A depends on n, that is, $A = A(n)$.
- The real numbers f_j are the images of x_j under a function f

$$f_j = f(x_j) = f\left(a + \frac{b-a}{n} j\right).$$

Then, if $\alpha > 0$ and $\beta < 0$ are real constants we can define

$$\begin{aligned} N_a(x, A) \equiv & \sum_{j=0}^{n-1} (f_j - f_{j+1}) s\left(\frac{(\beta-\alpha)A(n)x}{x_{j+1} - x_j} + \frac{\alpha x_{j+1} - \beta x_j}{x_{j+1} - x_j} A(n)\right) \\ & + f_n s\left(\frac{(\beta-\alpha)A(n)x}{x_n - x_{n-1}} + \frac{(2\alpha - \beta)x_n - \alpha x_{n-1}}{x_n - x_{n-1}} A(n)\right). \end{aligned}$$

Lemma 7. *Let $s(x)$ be the logistic function, we have*:

$$\textit{If } v > 0, \textit{ then } \lim_{x \to \infty} x s(vx)(1 - s(vx)) = 0, \tag{19}$$

$$s(x)(1 - s(x)) \leqslant \tfrac{1}{4} \quad \textit{for all } x \in R, \tag{20}$$

$$\lim_{x \to \infty} \frac{ln(x)}{x} = 0. \tag{21}$$

Proof. All these results can easily be obtained using techniques of differential calculus. □

A function *f* is Lipschitz on $[a, b]$ if there exists a constant $L > 0$ such that

$$|f(x) - f(y)| \leqslant L|x - y|$$

for all $x, y \in [a, b]$. We say that L is a Lipschitz constant for *f*.

Lemma 8. *Let f be a Lipschitz function on* $[a, b]$. *For all* $\varepsilon > 0$, *there exists a function* $A(n)$ *and a natural number N such that*

$$|f(x) - N_a(x, A(n))| < \varepsilon$$

for all $n > N$ *and for all* $x \in [a, b]$.

Proof. Let $x \in [a, b]$, and suppose that $x \in [x_J, x_{J+1}]$, then

$$\begin{aligned} |f(x) - N_a(x, A)| &= |f(x) - f(x_J) + f(x_J) - N_a(x_J, A) + N_a(x_J, A) - N_a(x, A)| \\ &\leqslant |f(x) - f(x_J)| + |f(x_J) - N_a(x_J, A)| + |N_a(x_J, A) - N_a(x, A)| \\ &= |f(x) - f(x_J)| + |N_e(x_J, A) - N_a(x_J, A)| + |N_a(x_J, A) - N_a(x, A)| \\ &\leqslant \frac{L(b-a)}{n} + |N_e(x_J, A) - N_a(x_J, A)| + |N_a(x_J, A) - N_a(x, A)|. \end{aligned} \tag{22}$$

From (18), if $A > A_2$

$$|N_e(x, A) - N_a(x, A)| < \frac{4e^{-\gamma A}}{1 - e^{-\omega A} - 4e^{-\gamma A}}(L(b-a) + |f(b)|). \tag{23}$$

The third addend in (22), can be bounded if we compute the derivative of $N_a(x, A)$ for $x \in (x_J, x_{J+1})$. If we use the notation

$$s^d(w_j x + b_j) \equiv s(w_j x + b_j)(1 - s(w_j x + b_j)),$$

and $0 < J < n - 1$, we have

$$\begin{aligned} |N_a'(x, A)| &= \left| \sum_{j=0}^{J-1} \frac{\omega A(f_j - f_{j+1})}{x_j - x_{j+1}} s^d(w_j x + b_j) + \frac{\omega A(f_J - f_{J+1})}{x_J - x_{J+1}} s^d(w_J x + b_J) \right. \\ &\quad \left. + \sum_{j=J+1}^{n-1} \frac{\omega A(f_j - f_{j+1})}{x_j - x_{j+1}} s^d(w_j x + b_j) - \frac{f(b)\omega A n}{b-a} s^d(w_n x + b_n) \right| \\ &\leqslant \sum_{j=0}^{J-1} \left| \frac{f_j - f_{j+1}}{x_j - x_{j+1}} \right| \omega A s^d(w_j x + b_j) + \left| \frac{f_J - f_{J+1}}{x_J - x_{J+1}} \right| \omega A s^d(w_J x + b_J) \\ &\quad + \sum_{j=J+1}^{n-1} \left| \frac{f_j - f_{j+1}}{x_j - x_{j+1}} \right| \omega A s^d(w_j x + b_j) + \frac{|f(b)|\omega A n}{b-a} s^d(w_n x + b_n) \\ &\leqslant \omega n A L s(\gamma A)(1 - s(\gamma A)) + \frac{\omega A L}{4} + \frac{|f(b)|\omega A n}{b-a} s(\gamma A)(1 - s(\gamma A)). \end{aligned}$$

This inequality is also valid for $J = 0$ and $n - 1$ and is based on the following facts

- If $x \notin [\beta A, \alpha A]$ then $s(x)(1 - s(x)) \leqslant s(\gamma A)(1 - s(\gamma A))$.
- (20).
- Lipschitz property of *f*.

Now we can give a bound of the third addend in (22). We have for some $x^* \in (x_J, x_{J+1})$

$$\begin{aligned}|N_a(x, A) - N_a(x_J, A)| &= |N'_a(x^*, A)||x - x_J| \\ &\leqslant (\omega n A L s(\gamma A)(1 - s(\gamma A)) + \frac{\omega A L}{4} \\ &\quad + \frac{|f(b)|\omega A n}{b-a} s(\gamma A)(1 - s(\gamma A)))\frac{b-a}{n} \\ &\leqslant \omega(L(b-a) + |f(b)|)A s(\gamma A)(1 - s(\gamma A)) + \frac{\omega A L(b-a)}{4n}.\end{aligned}$$

We must find a number N such that if $n > N$, (23) and the following inequalities

$$\frac{4e^{-\gamma A}}{1 - e^{-\omega A} - 4e^{-\gamma A}}(L(b-a) + |f(b)|) < \frac{\varepsilon}{4}, \tag{24}$$

$$\frac{\omega A L(b-a)}{4n} < \frac{\varepsilon}{4}, \tag{25}$$

$$\omega(L(b-a) + |f(b)|)A s(\gamma A)(1 - s(\gamma A)) < \frac{\varepsilon}{4}, \tag{26}$$

$$\frac{L(b-a)}{n} < \frac{\varepsilon}{4}, \tag{27}$$

hold.

If we take $A = ln(n^p)$ with $p > 0$ there exists a natural number $N_A(p)$ such that if $n > N_A(p)$, then $A = ln(n^p) > A_2$.

Eq. (24) changes into

$$\frac{4n^{-p\gamma}}{1 - n^{-p\omega} - 4n^{-p\gamma}}(L(b-a) + |f(b)|) < \frac{\varepsilon}{4},$$

and there exists a natural number $N_1(p)$ such that if $n > N_1(p)$, this inequality holds.

Eq. (25) can be expressed as

$$\omega p L(b-a)\frac{ln(n)}{n} < \varepsilon.$$

From (21) there exists a natural number $N_2(p)$ such that if $n > N_2(p)$, this inequality is satisfied.

Eq. (26) changes into

$$\omega(L(b-a) + |f(b)|)p\, ln(n) s(\gamma p\, ln(n))(1 - s(\gamma p\, ln(n))) < \frac{\varepsilon}{4}.$$

From (19) there exists a natural number $N_3(p)$ such that if $n > N_3(p)$, this inequality holds.

Finally, there exists a natural number N_4 such that if $n > N_4$, (27) is satisfied.

The result follows if we take

$$N = \max(N_A(p), N_1(p), N_2(p), N_3(p), N_4). \qquad \square$$

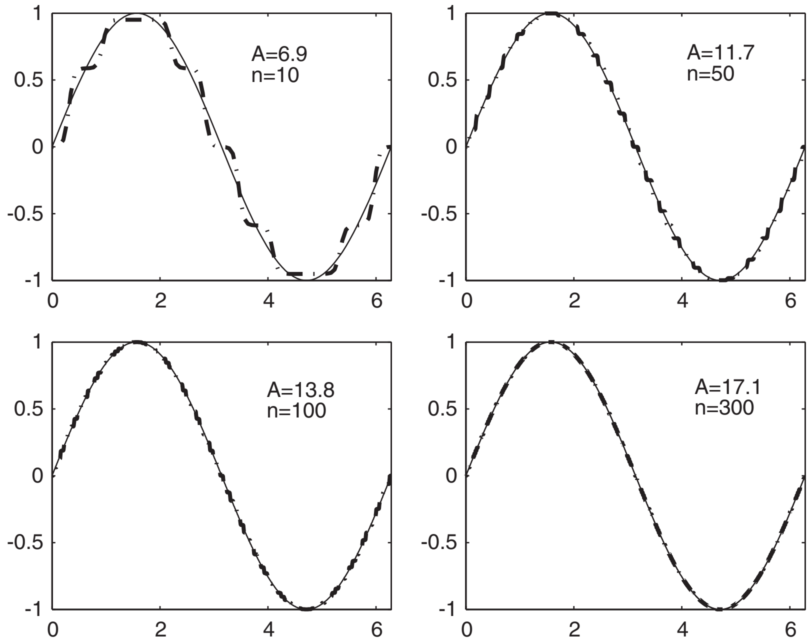

Fig. 2. Uniform convergence of ai-nets: $\sin(x)$.

Theorem 3. *Let f be a continuous function on* $[a, b]$. *For each* $\varepsilon > 0$ *there exists a function* $A(n)$ *and a natural number N such that*

$$|f(x) - N_a(x, A(n))| < \varepsilon$$

for all $n > N$ *and for all* $x \in [a, b]$.

Proof. From Lemma 6 we can find a piecewise linear function $l_{\frac{\varepsilon}{2}}$ such that

$$|f(x) - l_{\frac{\varepsilon}{2}}(x)| < \frac{\varepsilon}{2} \quad \text{for } x \in [a, b].$$

As $l_{\frac{\varepsilon}{2}}$ is Lipschitz on $[a, b]$, we can apply Lemma 8 and find $A(n)$ and N such that

$$|l_{\frac{\varepsilon}{2}}(x) - N_a(x, A(n))| < \frac{\varepsilon}{2} \quad \text{for } x \in [a, b] \text{ and } n > N.$$

The result follows from the above inequalities. □

Figs. 2 and 3 show the uniform convergence of ai-nets when n and A increases. Here we have considered $\alpha = 1$, $\beta = -1$ and $p = 3$.

Theorem 3 can be extended to the multidimensional case according to the procedures given in [5,10]. We follow the first approach, less general but more constructive.

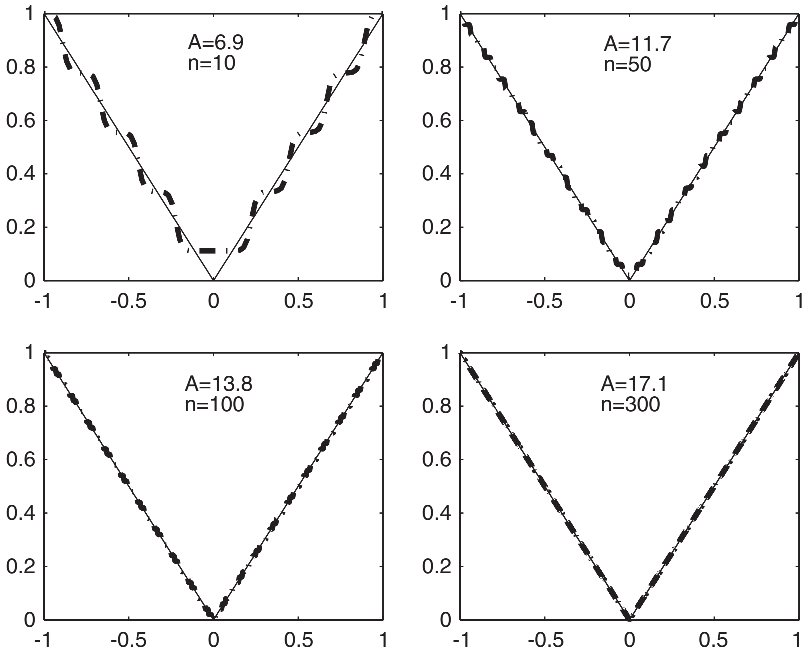

Fig. 3. Uniform convergence of ai-nets: $|x|$.

Corollary 3. *Let $f(\mathbf{x})$ be a continuous function on $I = [0, 1]^d$. For each $\varepsilon > 0$ there exists a neural network*

$$N(\mathbf{x}) = \sum_{j=0}^{n} c_j s(\mathbf{w}_j.\mathbf{x} + b_j),$$

where $\mathbf{w}_j \in R^d$, $c_j, b_j \in R$ such that

$$|N(\mathbf{x}) - f(\mathbf{x})| < \varepsilon \quad \textit{for all } \mathbf{x} \in I.$$

Proof. We only give the main steps of the proof. Details can be consulted in [5].

- Extend $f(\mathbf{x})$ to be a 2-periodic even function with respect to every variable $x_i, i = 1, \ldots, d.$, $g(\mathbf{x})$ defined on $J = [-1, 1]^d$.
- By a result on Bochner–Riesz means [25, p. 256] and by the evenness of $g(\mathbf{x})$, we have that for any $\varepsilon > 0$, there exists $R > 0$ such that for any $\mathbf{x} = (x_1, \ldots, x_d) \in J$

$$\left| \sum_{|\mathbf{m}| \leqslant R} c_{m_1 \ldots m_d} \cos(m_1 x_1 + \cdots + m_d x_d) - g(x_1, \ldots, x_d) \right| < \frac{\varepsilon}{2},$$

  where $c_{m_1 \ldots m_d}$ are real numbers, $\mathbf{m} = (m_1, \ldots, m_d) \in Z^d$ and $|\mathbf{m}|^2 = m_1^2 + \cdots + m_d^2$.
- Define $u = \mathbf{m}.\mathbf{x}$ and approximate $\cos(u)$ by an ai-net. $\square$

5. Multidimensional neural approximate interpolants

Let the function $f : R^d \to R$ and let $S = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n\} \subset R^d$ be a set of distinct vectors. If $\{f_i / i = 0, \dots, n\}$ is a set of real numbers, we search for a neural network N such that

$$N(\mathbf{x}_i) = f_i \quad (i = 0, 1, \dots, n).$$

Lemma 9 (*Huang and Babri [11], Shristava and Dasgupta [23]*). *Let $S = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n\} \subset R^d$ be a set of distinct vectors, then there exists a vector $\mathbf{w}$ such that the $n+1$ inner products*

$$\mathbf{w}.\mathbf{x}_0, \mathbf{w}.\mathbf{x}_1, \dots, \mathbf{w}.\mathbf{x}_n,$$

are different from each other.

The set of $\mathbf{w}$ that verify the above condition is the complement of a finite union of lower dimensional hyperplanes. Then we can pick $\mathbf{w}$ according to any distribution having a density (with respect to Lebesgue measure) and find with probability 1, a $\mathbf{w}$ that works. From now on, $\mathbf{w}$ denotes a fixed vector that verifies the condition in Lemma 9.

If we reorder the points of S in such a way that

$$\mathbf{w}.\mathbf{x}_0 < \mathbf{w}.\mathbf{x}_1 < \cdots < \mathbf{w}.\mathbf{x}_n,$$

we can obtain the expression of a multidimensional ai-net, simply, replacing x by $\mathbf{w}.\mathbf{x}$ in (13), that is

$$\begin{aligned} N_a(\mathbf{x}, \mathbf{w}, A) = {} & \sum_{j=0}^{n-1} (f_j - f_{j+1}) \sigma \left(\frac{(\beta - \alpha) A \mathbf{w}.\mathbf{x}}{\mathbf{w}.\mathbf{x}_{j+1} - \mathbf{w}.\mathbf{x}_j} + \frac{\alpha \mathbf{w}.\mathbf{x}_{j+1} - \beta \mathbf{w}.\mathbf{x}_j}{\mathbf{w}.\mathbf{x}_{j+1} - \mathbf{w}.\mathbf{x}_j} A \right) \\ & + f_n \sigma \left(\frac{(\beta - \alpha) A \mathbf{w}.\mathbf{x}}{\mathbf{w}.\mathbf{x}_n - \mathbf{w}.\mathbf{x}_{n-1}} + \frac{(2\alpha - \beta) \mathbf{w}.\mathbf{x}_n - \alpha \mathbf{w}.\mathbf{x}_{n-1}}{\mathbf{w}.\mathbf{x}_n - \mathbf{w}.\mathbf{x}_{n-1}} A \right). \end{aligned} \tag{28}$$

The corresponding ei-net $N_e(\mathbf{x}, \mathbf{w}, A)$ can be obtained by replacing $(f_j - f_{j+1})$ with c_j and f_n with c_n in (28). A multidimensional version of Theorem 1 can be stated using arguments similar to the ones in Section 1.

If we define

$$\begin{aligned} &\Delta_j \equiv \mathbf{w}.\mathbf{x}_j - \mathbf{w}.\mathbf{x}_{j-1} \ (j = 1, \dots, n), \\ &M \equiv \max_{j=1,\dots,n} \Delta_j, \\ &m \equiv \min_{j=1,\dots,n} \Delta_j, \\ &r \equiv m/M. \end{aligned}$$

We have, using the notation of Section 3

Theorem 4. *Let $N_a(\mathbf{x}, \mathbf{w}, A)$ be the multidimensional ai-net defined in* (28). *Then, there exists a real number A_3 such that if $A > A_3$,*

$$|f_i - N_a(\mathbf{x}_i, \mathbf{w}, A)| < \frac{2(n+1)\delta(A)}{1 - 2(n+1)\delta(A)} \left(\sum_{j=0}^{n-1} |f_j - f_{j+1}| + |f_n| \right)$$

for $i = 0, 1, \dots, n$.

Proof. If we replace in (28) $\mathbf{w}.\mathbf{x}$ by t and $\mathbf{w}.\mathbf{x}_j$ by t_j, we have the function

$$N_a(t, A) = \sum_{j=0}^{n-1} (f_j - f_{j+1})\sigma\left(\frac{(\beta-\alpha)At}{t_{j+1}-t_j} + \frac{\alpha t_{j+1} - \beta t_j}{t_{j+1}-t_j} A\right) + f_n\sigma\left(\frac{(\beta-\alpha)At}{t_n - t_{n-1}} + \frac{(2\alpha-\beta)t_n - \alpha t_{n-1}}{t_n - t_{n-1}} A\right).$$

This is the unidimensional ai-net for the problem

$$(t_i, f_i) \quad (i = 0, 1, \ldots, n).$$

It is clear that $N_a(\mathbf{x}_i, \mathbf{w}, A) = N_a(t_i, A)$ for $i = 0, 1, \ldots, n$.

If we consider the exact interpolation net $N_e(t, A)$, that by Theorem 1 exists if $A > A^*$, we have

$$|f_i - N_a(\mathbf{x}_i, \mathbf{w}, A)| = |f_i - N_a(t_i, A)| = |N_e(t_i, A) - N_a(t_i, A)|.$$

From Theorem 2, the result follows. □

In the case of logistic sigmoidal activation function we have the following result:

Corollary 4. *Let $N_a(\mathbf{x}_i, \mathbf{w}, A)$ be the multidimensional ai-net defined in* (28). *Then there exists a real number $A_4 > 0$ such that if $A > A_4$,*

$$|f_i - N_a(\mathbf{x}_i, \mathbf{w}, A)| < \frac{4e^{-\gamma A}}{1 - e^{-\omega Ar} - 4e^{-\gamma A}} \left(\sum_{j=0}^{n-1} |f_j - f_{j+1}| + |f_n|\right).$$

Fig. 4 shows the multidimensional ai-net corresponding to the function $z = x + y$ on the 49 points $\{(i, j)/0 \leqslant i \leqslant 6,\ 0 \leqslant j \leqslant 6\}$. We have considered $\mathbf{w} = (1, 10)$ and $A = 5$.

Fig. 5 shows the multidimensional ai-net corresponding to the above function, but the 49 points have been equidistantly placed on the intersection of the straight line $y = 10x$ and the square $[0, 6] \times [0, 6]$. In this case the approximation to the function $z = x + y$ is much better than in the above case.

Multidimensional ai-nets are constant in the hyperplanes orthogonal to $\mathbf{w}$, therefore, in general, we cannot obtain a good approximation by a only ai-net. For example, in the case of functions from R^2 to R, the graph of the corresponding bidimensional ai-nets are ruled surfaces.

6. Conclusions

We give a new and quantitative proof of the fact that a single layer neural network with n hidden neurons can learn n distinct samples with zero error.

Based on this result we introduce a family of neural networks (*ai-nets*). These networks can approximately interpolate n samples in any dimension with arbitrary precision and without training. We give a rigorous bound of the interpolation error.

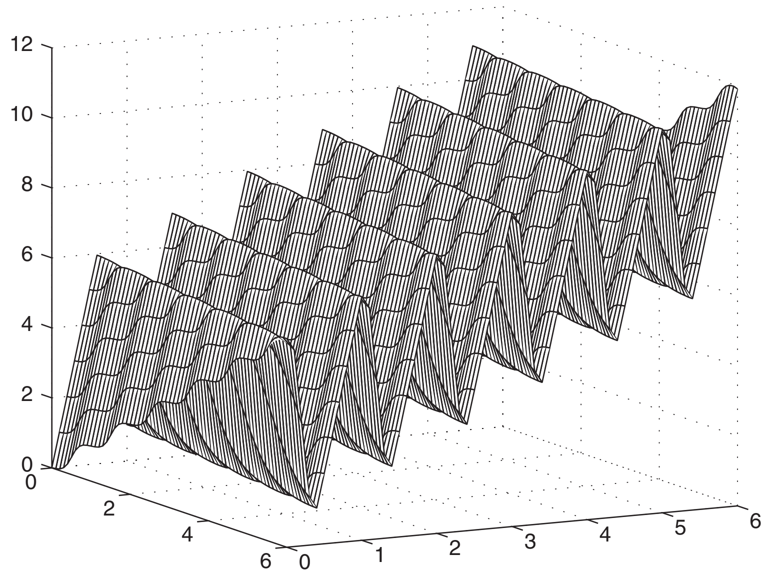

Fig. 4. Interpolation of $z = x + y$. Uniform distribution of interpolation points.

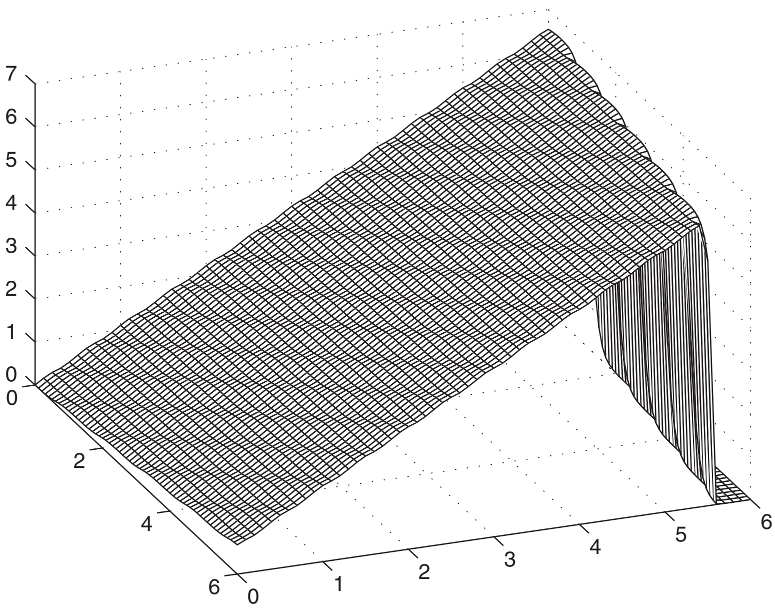

Fig. 5. Interpolation of $z = x + y$. Interpolation points on the straight line $y = 10x$.

As an application, we prove, in the unidimensional case and with logistic activation function, that ai-nets can be used as uniform approximation operators for continuous functions. These operators are somewhat more general than those presented in [5,8,19]. In this way, we give a connection between interpolation and approximation in the case of neural networks.

From Lemma 8 and the definitions in Section 5, we can prove that a multidimensional ai-net can uniformly approach, without training, a linear function defined on a simplex of R^d. This opens the way to the study of localized approximation [6,7,18] based on ai-nets. This topic will be addressed in future work.

Acknowledgements

The authors wish to thank the reviewers for their constructive suggestions and comments.

References

[1] M. Anthony, Function learning from interpolation, London School of Economics, -Mathematics Preprint Series-73, 1994.
[2] M. Anthony, P.L. Bartlett, Neural Network Learning: Theoretical Foundations, Cambridge University Press, Cambridge, 1999.
[3] J. Barhen, R. Cogswell, V. Protopopescu, Single iteration training algorithm for multilayer feedforward neural networks, Neural Process. Lett. 11 (2000) 113–129.
[4] P. Cardaliaguet, G. Euvrard, Approximation of a function and its derivative with a neural network, Neural Networks 5 (1992) 207–220.
[5] T. Chen, H. Chen, R.-W. Liu, A constructive proof and an Extension of Cybenko's approximation theorem, in: Computing Science and Statistics, Proceedings of the 22nd Symposium on the Interface, Springer, Berlin, 1991, pp. 163–168.
[6] C.K. Chui, X. Li, H.N. Mhaskar, Neural networks for localized approximation, Mathematics of Computation 63 (1994) 607–623.
[7] C.K. Chui, X. Li, H.N. Mhaskar, Limitations of the approximation capabilities of neural networks with one hidden layer, Adv. Comput. Math. 5 (1996) 233–243.
[8] Ch. Debao, Degree of approximation by superpositions of a sigmoidal function, Approx. Theory & its Appl. 9 (1993) 17–28.
[9] W. Duch, N. Jankowski, Survey of neural transfer functions, Neural Computing Surveys 2 (1999) 163–212.
[10] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (1989) 359–366.
[11] G.B. Huang, H.A. Babri, Feedforward neural networks with arbitrary bounded nonlinear activation functions, IEEE Trans. Neural Networks 9–1 (1998) 224–229.
[12] Y. Ito, Nonlinearity creates linear independence, Adv. Comput. Math. 5 (1996) 189–203.
[13] Y. Ito, Independence of unscaled basis functions and finite mappings by neural networks, Math. Sci. 26 (2001) 117–126.
[14] Y. Ito, K. Saito, Superposition of linearly independent functions and finite mappings by neural networks, Math. Sci. 21 (1996) 27–33.
[15] R. Kress, Numerical Analysis, Springer, New York, 1998.
[16] P. Lascaux, T. Theodor, Analyse Numerique Matricielle Apliquée a l'Art de l'Ingenieur, Masson, Paris, 1986.
[17] X. Li, Interpolation by ridge polynomials and its application in neural networks, J. Comput. Appl. Math. 144 (2002) 197–209.
[18] H.N. Mhaskar, Neural networks for localized approximation of real functions, in: Neural Networks for Signal Processing, Proceedings of the 1993 IEEE-SP Workshop, 6–9 September 1993, IEEE, 1993, pp. 190–196.
[19] H.N. Mhaskar, Ch.A. Michelli, Approximation by superposition of sigmoidal and radial basis functions, Adv. Appl. Math. 13 (1992) 350–373.
[20] A. Pinkus, Approximation theory of the MLP model in neural networks, Acta Numerica (1999) 143–195.
[21] Th.J. Rivlin, An Introduction to the Approximation of Functions, Dover, Mineola, New York, 1981.
[22] M.A. Sartori, P.J. Antsaklis, A simple method to derive bounds on the size and to train multilayer neural networks, IEEE Trans. Neural Networks 2–4 (1991) 467–471.

[23] Y. Shristava, S. Dasgupta, Neural networks for exact matching of functions on a discrete domain, in: Proceedings of the 29th IEEE Conference on Decision and Control, Honolulu, 1990, pp. 1719–1724.
[24] E.D. Sontag, Feedforward nets for interpolation and classification, J. Comp. Syst. Sci. 45 (1992) 20–48.
[25] E.M. Stein, G. Weiss, Introduction to Fourier Analysis on Euclidean Spaces, Princeton University Press, Princeton, 1975.
[26] S. Tamura, M. Tateishi, Capabilities of a four-layered feedforward neural network, IEEE Trans. Neural Networks 8-2 (1997) 251–255.