ELSEVIER

Neurocomputing 17 (1997) 159–166

NEUROCOMPUTING

Placing spline knots in neural networks using splines as activation functions

Kateřina Hlaváčková [a,*], Michel Verleysen [b]

[a] *Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, Czech Republic*
[b] *Microelectronics Laboratory, Catholic University of Louvain, Louvain-la-Neuve, Belgium*

Received date 12 May 1995; accepted 5 May 1997

Abstract

When using feed-forward neural networks with spline activation functions, the quality of approximation depends on the knot placement of spline functions. We demonstrate a method of choosing equidistant knots in each subdivision of the space when an arbitrary initial division is given, in order to keep the approximation error under a predefined limit.

Keywords: Cubic-spline function; Knots of a spline function; Approximation error; Feed-forward neural network

1. Introduction

The approximation of functions is a classical engineering problem which can be solved by various mathematical approaches. Researchers in the field of neural networks also developed new approaches or applied the conventional ones; advantages of functional approximation by neural networks mainly consist in their simplified descriptions, usually in terms of algorithms or diagrams instead of complex mathematical equations. There are methods based on "local" activation functions among various "neural networks" algorithms used in the approximation of functions, for example radial-basis functions (RBF) networks. Methods based on splines seem to be less studied in neural network research, despite the fact that their approximation properties have been justified by numerous convergence theorems from the 1960s and 1970s.

This paper deals with the approximation of functions by cubic spline networks. The number of knots necessary for input space partition is evaluated to keep the

* Corresponding author. Fax: +420 2 8585789; e-mail: katka@uivt.cas.cz.

PII S0925-2312(97)00053-2

approximation under a predefined limit. The results are derived in one- and two-dimensional cases, where spline approximation is mostly used.

2. Neural network with cubic spline activation functions

In this paper we use the definition of cubic spline interpolation as introduced in [1].

2.1. One-dimensional case

Consider $I=[0,1]\subset\mathscr{R}$ and a function $f:I\rightarrow\mathscr{R}$. Let Δ be a division of I given by the set of knots $\Delta\equiv\{0=x_0<x_1<\cdots<x_N=1\}$.

Given Δ, let the *space of cubic splines with respect to* Δ, $S(\Delta)$, be the vector space of all twice continuously differentiable, piecewise cubic polynomials on I with respect to Δ, i.e.

$$S(\Delta)\equiv\{p(x)\in C^2(I)\}, \tag{1}$$

so that $p(x)$ is a cubic polynomial on each subinterval $[x_i,x_{i+1}]$, $0\leq i\leq N$, defined by Δ. Every cubic spline may be represented in terms of the *basis functions* $\{h_i(x), h_i^1(x)\}_{i=0}^{N+1}$: $s(x)=\sum_{i=0}^{N+1}(s(x_i)h_i(x)+s'(x_i)h_i^1(x))$, where $h_i(x)$ is the unique piecewise cubic polynomial on each subinterval defined by Δ in $C^1(I)$ such that $h_i(x_j)=\delta_{ij}$, $0\leq i,j\leq N+1$, and $h_i'(x_j)=0$, $0\leq i,j\leq N+1$, and $h_i^1(x)$ is the unique piecewise cubic polynomial on each subinterval defined by Δ in $C^1(I)$ such that $h_i^1(x_j)=0$, $0\leq i,j\leq N+1$, and $(h_i^1)'(x_j)=\delta_{ij}$, $0\leq i,j\leq N+1$.

Given $f\equiv(f_0,\ldots,f_{N+1},f_0',f_{N+1}')\in R^{N+3}$, let $\hat{f}(x)$, the $S(\Delta)$-*interpolate of* f be the unique spline $s(x)$ in $S(\Delta)$ such that $s(x_i)=f_i$, $0\leqslant i\leqslant N+1$ and $\hat{f}'(x_i)=f_i'$, $i=0$ and $N+1$. If parameters f_0' and f_{N+1}' are not known, they can be approximated by the local cubic Lagrange interpolating polynomials at both ends of the interval I, before determining the approximation $\hat{f}(x)$. Define a basis for $S(\Delta)$, consisting in the *cardinal splines* $\{C_i(x)\}_{i=0}^{N+3}$: $C_j(x_i)=\delta_{ij}, C_j'(0)=C_j'(1)=0$, for $0\leqslant i$, $j\leqslant N+1$, and by $C_{N+2}(x_i)=C_{N+3}(x_i)=0$, $0\leqslant i,j\leqslant N+1$, $C_{N+2}'(0)=C_{N+3}'(1)=1$ and $C_{N+2}'(1)=C_{N+3}'(0)=0$. In this case, the approximation $\hat{f}(x)$ can be written as

$$\hat{f}(x)=\sum_{i=0}^{N+1}f_iC_i(x)+f_0'C_{N+2}(x)+f_{N+1}'C_{N+3}(x). \tag{2}$$

$\hat{f}(x)$ depends on all parameters $f_i, 0\leqslant i\leqslant N+1, f_0'$ and f_{N+1}' (Fig. 1). A corresponding spline feed-forward neural network has cardinal splines as activation functions.

2.2. Two-dimensional case

Consider $U\equiv[0,1]\times[0,1]\subset\mathscr{R}^2$ and a real function $f:U\rightarrow\mathscr{R}$. Let $\rho=\Delta\otimes\Delta_y$ be a rectangular grid given by $\Delta\equiv\{0=x_0<x_1<\cdots<x_N\equiv1\}$ and $\Delta_y=\{0=y_0<y_1<\cdots<y_M=1\}$ are some given divisions of I.

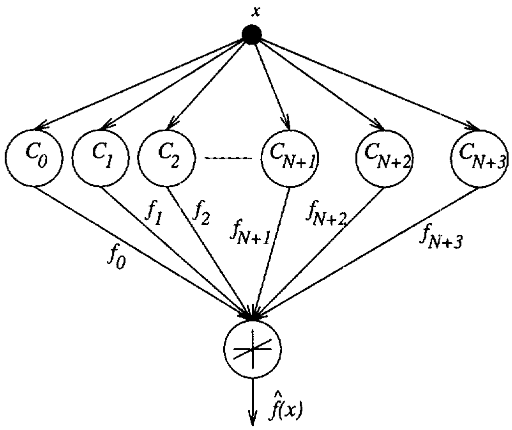

Fig. 1. Network with one-dimensional input, $N+3$ units in the hidden layer and network weights f_i, f_0 and f_{N+1}.

Let $S(\rho)$ be the $(N+4)(M+4)$-dimensional vector space of all functions of the form

$$s(x,y)=\sum_{i=0}^{N+3}\sum_{j=0}^{M+3}\beta_{ij}C_i(x)C_j(y),$$

where $C_i(x), C_j(x)$ are cardinal splines of $S(\Delta)$. $S(\rho)$ is the space of piecewise bicubic polynomials in U with respect to ρ. Now we can uniquely define the interpolation $\hat{f}(x,y)$ of f in $S(\rho)$ by (Fig. 2)

$$\begin{aligned}\hat{f}(x,y)=&\sum_{i=0}^{N+1}\sum_{j=0}^{M+1}f_{i,j}C_i(x)C_j(y)+\sum_{j=0}^{M+1}(f_{0,j}^{1,0}C_{N+2}(x)+f_{N+1,j}^{1,0}C_{N+3}(x))C_j(y)\\&+\sum_{i=0}^{N+1}(f_{i,0}^{0,1}C_{M+2}(y)+f_{i,M+1}^{0,1}C_{M+3}(y))C_i(x)+f_{0,0}^{1,1}C_{N+2}(x)C_{M+2}(y)\\&+f_{0,M+1}^{1,1}C_{N+2}(x)C_{M+3}(y)+f_{N+1,0}^{1,1}C_{N+3}(x)C_{M+2}(y)\\&+f_{N+1,M+1}^{1,1}C_{N+3}(x)C_{M+3}(y),\end{aligned}\tag{3}$$

where

$$f_{i,j}\equiv f(x_i,y_j),\quad f_{0,j}^{1,0}\equiv\frac{\partial f(0,y_j)}{\partial x},\quad f_{N+1,j}^{1,0}\equiv\frac{\partial f(1,y_j)}{\partial x},\quad f_{i,0}^{0,1}\equiv\frac{\partial f(x_i,0)}{\partial y},$$

$$f_{i,M+1}^{0,1}\equiv\frac{\partial f(x_i,1)}{\partial y},\quad f_{0,0}^{1,1}\equiv\frac{\partial^2 f(0,0)}{\partial x\partial y},\quad f_{0,M+1}^{1,1}\equiv\frac{\partial^2 f(0,1)}{\partial x\partial y},$$

$$f_{N+1,0}^{1,1}\equiv\frac{\partial^2 f(1,0)}{\partial x\partial y},\quad f_{N+1,M+1}^{1,1}\equiv\frac{\partial^2 f(1,1)}{\partial x\partial y}$$

for all $0\le i\le N+1$ and $0\le j\le M+1$.

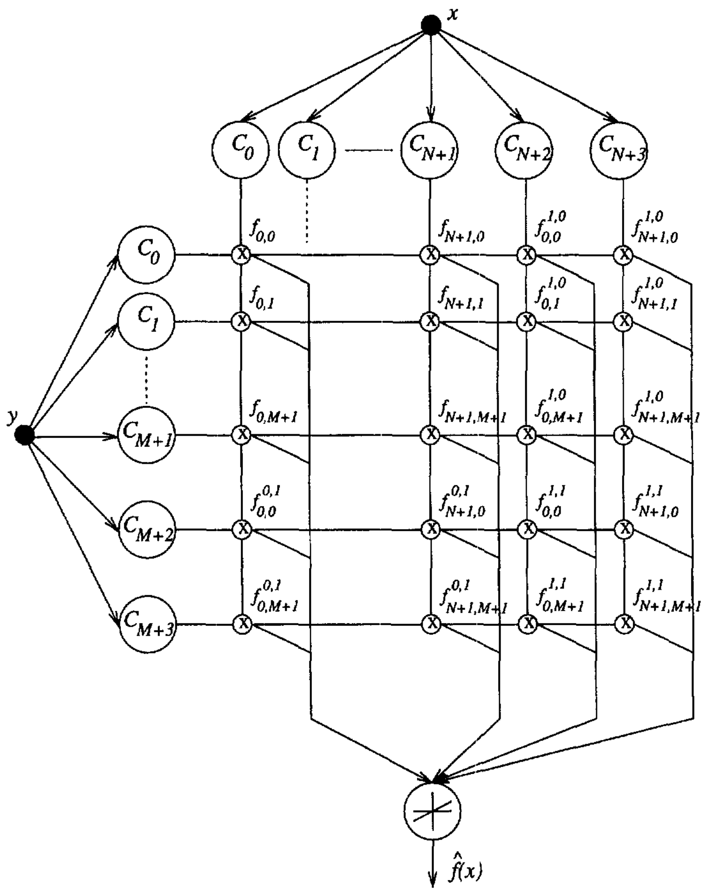

Fig. 2. Network with two-dimensional input with cardinal splines as activation functions.

3. Approximation with uniform-bounded error

In this section, we deal with the problem of finding an adequate number of knots, i.e. the number of points x_i and y_j, to ensure the approximation error under a predefined bound. We synthesize our results based on the error-bound estimations known from [1]. Both in one- and two-dimensional spaces, the approximation error is bounded by a function depending on the maximum value of the fourth derivative of function f in the considered interval. As the bounds on corresponding fourth derivative may vary from one location of the space to another, it is convenient to split the space into subdivisions. We examine only equally spaced knots into each subdivision of the space.

3.1. One-dimensional case

Let the interval $I=[0,1]$ be divided into R intervals $[\gamma_i,\gamma_{i+1}]$, $0\le i\le N$, $\gamma_0=0$, $\gamma_{N+1}=1$ given by Δ. Our concern is to find a distance h_i between equally spaced knots in each interval $[\gamma_i,\gamma_{i+1}]$, such that the approximation error between f and $\hat{f}$ is under a predefined limit ε for all points in $[\gamma_i,\gamma_{i+1}]$, i.e. $\|f-\hat{f}\|_\infty = \sup_{x\in I}\|f(x)-\hat{f}(x)\|<\varepsilon$. Consider $PC^{4,\infty}(I)$, the space of all functions on I such that

(1) f is three times continuously differentiable,
(2) there exist γ_i, $i=0\le i\le N$ with $0=\gamma_0<\gamma_1<\cdots<\gamma_N=1$ such that on each open subinterval (γ_i,γ_{i+1}), $0\le i\le N-1$, $f^{(3)}$ is continuously differentiable,
(3) $\| f^{(4)}(x)\|_\infty = \max_{0\leqslant i\leqslant N}\sup_{x\in(\gamma_i,\gamma_{i+1})}|f^{(4)}(x)|<\infty$ for $x\in I$.

Lemma 1 (Schultz [1]). *If $f\in PC^{4,\infty}(I)$, then*

$$\|f-\hat{f}\|_\infty \le \frac{5}{384}h^4\|f^{(4)}\|_\infty, \tag{4}$$

where $h\equiv \max_{0\le i\le N}(\gamma_{i+1}-\gamma_i)$.

Applying this lemma on each subdivision $[\gamma_i,\gamma_{i+1}]$, we get the following theorem.

Theorem 2. *Let $\Delta=\{0=\gamma_0<\cdots<\gamma_N=1\}$ be a given division of I and let $L_i=\sup_{x\in(\gamma_i,\gamma_{i+1})}|f^{(4)}(x)|$ for all $i=0,\ldots,N$, $B=\frac{5}{384}$. Then for every $\varepsilon>0$ there exists a neural network with spline activation functions given on the subdivisions of $[\gamma_i,\gamma_{i+1}]$, $i=0,\ldots,N$ consisting of M_i+1 equidistant knots $\gamma_{i_0},\ldots,\gamma_{i_{M_i}}$ on each $[\gamma_i,\gamma_{i+1}]$ realizing a function $\hat{f}$ so that for every $x\in I$*

$$|f(x)-\hat{f}(x)|\le\varepsilon. \tag{5}$$

This M_i can be found in the way

$$M_i=\left\lceil\frac{\gamma_{i+1}-\gamma_i}{h_i}\right\rceil, \tag{6}$$

where $h_i=\sqrt[4]{\varepsilon/(BL_i)}$. If $L_i\ne L_j$ then $M_i\ne M_j$. ($\lceil a\rceil$ means the smallest integer greater than a).

Proof. Let the right-hand side of the equation in Lemma 1 be at most ε. Applying this condition to each interval $[\gamma_i,\gamma_{i+1}]$, we have

$$Bh_i^4\|f^{(4)}(x)\|_\infty\le\varepsilon, \tag{7}$$

or for $x\in[\gamma_i,\gamma_{i+1}]$ and each $0\le i\le N$,

$$h_i\leqslant\sqrt[4]{\frac{\varepsilon}{BL_i}},\quad \text{where } L_i=\|f^{(4)}(x)\|_\infty \tag{8}$$

The best choice for h_i is obviously the largest possible value, in order to decrease the number of necessary knots. However, to keep limits γ_i and γ_{i+1} of the interval as knots and to have an integer value for $(\gamma_{i+1} - \gamma_i)/h_i$, the number of interior knots M_i (without γ_i and γ_{i+1}) is

$$M_i = \left\lceil \frac{\gamma_{i+1} - \gamma_i}{\sqrt[4]{\varepsilon/BL_i}} \right\rceil \quad \text{and consequently } h_i = \frac{\gamma_{i+1} - \gamma_i}{M_i + 1}. \tag{9}$$

3.2. Two-dimensional case

Similarly, we proceed in the two-dimensional case. We divide the space U by a rectangular grid ρ where $\rho = \Delta \otimes \Delta_y$ is a rectangular grid given by $\Delta \equiv \{0 = \gamma_0 < \cdots < \gamma_N = 1\}$ and $\Delta_y \equiv \{0 = \mu_0 < \cdots < \mu_M = 1\}$, which are given divisions of I. In each rectangle $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$, we find distances h_{ij} between equally spaced knots for both coordinates so that the approximation error is under the limit ε for all points in U in the rectangle. Denote $PC^{4,\infty}(U)$, the space of all functions on I such that

(1) $f(x,y)$ is three times continuously differentiable, i.e. $\partial^l \partial^k f(x,y)/\partial x^l \partial y^k$ exist and are continuous for all $0 \leq l + k \leq 3$,
(2) there exist γ_i, $0 \leq i \leq R$ and μ_j $0 \leqslant j \leqslant S$ with $0 = \gamma_0 < \gamma_1 < \cdots < \gamma_R = 1$ and $0 = \mu_0 < \cdots < \mu_s = 1$ such that on each open subrectangle $(\gamma_i, \gamma_{i+1}) \times (\mu_j, \mu_{j+1})$, $0 \leq i \leq R - 1$, $0 \leq j \leq S - 1$, $\partial^l \partial^k f(x,y)/\partial x^l \partial y^k$ are continuously differentiable for $0 \leq l + k \leq 3$, and
(3) for all $0 \leq l + k \leq 3$,

$$\left\| \frac{\partial^l \partial^k f(x,y)}{\partial x^l \partial y^k} \right\|_\infty = \max_{\substack{0 \leq i \leq R \\ 0 \leq j \leq S}} \sup_{(x,y) \in (\gamma_i, \gamma_{i+1}) \times (\mu_j, \mu_{j+1})} \left| \frac{\partial^l \partial^k f(x,y)}{\partial x^l \partial y^k} \right| < \infty$$

for $(x,y) \in U$.

Lemma 3 (Schultz [1]). *If $f \in PC^{4,\infty}(U)$, then*

$$\|f - \hat{f}\|_\infty \leq \hat{\rho}^4 \left(\frac{5}{384} \left\| \frac{\partial^4 f}{\partial x^4} \right\|_\infty + \frac{4}{9} \left\| \frac{\partial^4 f}{\partial x^2 \partial y^2} \right\|_\infty + \frac{5}{384} \left\| \frac{\partial^4 f}{\partial y^4} \right\|_\infty \right), \tag{10}$$

where $\hat{\rho} = \max\{h,k\}$ and h,k are the distances between knots in the x- and y-coordinates, respectively.

Denote $B = \frac{5}{384}$ and $D = \frac{4}{9}$. Applying the lemma on each subrectangle of the space $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$, we get the following theorem:

Theorem 4. *Let the space $U = [0,1] \times [0,1]$ (the domain) of function f be divided by a rectangular grid ρ, where $\rho = \Delta \otimes \Delta_y$, $\Delta \equiv \{0 = \gamma_0 < \gamma_1 < \cdots < \gamma_N \equiv 1\}$ and $\Delta_y \equiv \{0 = \mu_0 < \mu_1 < \cdots < \mu_M = 1\}$ are given divisions of I. Let*

$$L_{ij} = \sup \left(B \left\| \frac{\partial^4 f}{\partial x^4} \right\|_\infty + D \left\| \frac{\partial^4 f}{\partial x^2 \partial y^2} \right\|_\infty + B \left\| \frac{\partial^4 f}{\partial y^4} \right\|_\infty \right) \tag{11}$$

in the rectangle $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$. *Then for every* $\varepsilon > 0$ *there exists a neural network with spline activation functions given on the subdivisions of equidistant knots on* $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$ $i = 0, \dots, R, j = 0, \dots, S$ *consisting of* $M_{ij} = (R_i + 1) \times (S_j + 1)$ *equidistant knots realizing a function* $\hat{f}$ *so that for every* $(x, y) \in U$

$$|f(x,y) - \hat{f}(x,y)| \le \varepsilon, \tag{12}$$

where $R_i = |\gamma_{i+1} - \gamma_i|/h_i$, $h_i = \gamma_{i+1} - \gamma_i$, $S_j = |\mu_{j+1} - \mu_j /k_j$, $k_j = \mu_{j+1} - \mu_j$ *and where* $\hat{\rho}_{i,j} = \sqrt[4]{\varepsilon/L_{i,j}}$ *and* $h_{ij} = h_i/(R_{ij} + 1)$, $k_{ij} = k_j/(S_{ij} + 1)$, $h_i, k_j \leqslant \hat{\rho}_{i,j}$. *The best solution (in terms of knots* $M_{i,j}$*) is* $h_i = k_j = \rho_{i,j}$.

Proof. Applying the condition of Lemma 3 to each subrectangle of the space $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$, we have

$$\overline{\rho}_{ij}^4 \left(\frac{5}{384} \left\| \frac{\partial^4 f}{\partial x^4} \right\|_\infty + \frac{4}{9} \left\| \frac{\partial^4 f}{\partial x^2 \partial y^2} \right\|_\infty + \frac{5}{384} \left\| \frac{\partial^4 f}{\partial y^4} \right\|_\infty \right) \le \varepsilon, \tag{13}$$

$\hat{\rho}_{ij} \leqslant \sqrt[4]{\varepsilon/L_{ij}}$, where $\hat{\rho}_{ij} = \max\{h_{ij}, k_{ij}\}$ and

$$L_{ij} = \frac{5}{384} \left\| \frac{\partial^4 f}{\partial x^4} \right\|_\infty + \frac{4}{9} \left\| \frac{\partial^4 f}{\partial x^2 \partial y^2} \right\|_\infty + \frac{5}{384} \left\| \frac{\partial^4 f}{\partial y^4} \right\|_\infty \tag{14}$$

on the rectangle $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$. Under the condition (20), the best choice for h_{ij} and k_{ij} is obviously the largest possible values, in order to decrease the number of necessary knots. However, to keep knots on the boundaries of the rectangle and to have integer values for $(\gamma_{i+1} - \gamma_i)/h_{ij}$ and $(\mu_{j+1} - \mu_j)/k_{ij}$, we will choose the number of interior knots in the x-direction R_{ij} (not including γ_i and γ_{i+1}) and in the y-direction S_{ij} (not including μ_j and μ_{j+1}) to be equal to

$$R_{ij} = \left\lceil \frac{\gamma_{i+1} - \gamma_i}{\sqrt[4]{\varepsilon/L_i j}} \right\rceil, \quad S_{ij} = \left\lceil \frac{\mu_{j+1} - \mu_j}{\sqrt[4]{\varepsilon/L_i j}} \right\rceil, \quad M_{ij} = (R_{ij} + 1) \times (S_{ij} + 1) \tag{15}$$

and, consequently,

$$h_{ij} = \frac{\gamma_{i+1} - \gamma_i}{R_{ij} + 1} \quad \text{and} \quad k_{ij} = \frac{\mu_{j+1} - \mu_j}{S_{ij} + 1}. \tag{16}$$

We see that the lattice of knots is regular inside each subrectangle of the space but need not be regular over the whole space U.

4. Discussion

Knot placement in neural network approximation by superposition of local functions (e.g. RBF or splines) is a non-trivial problem, as well in one- as in multi-dimensional spaces. The number of knots to be placed in a defined region of the space must also be determined. This number can be given or estimated a priori, assuming some knowledge on the function to approximate.

By using results on the approximation error in one- and two-dimensional spaces, we presented a method to determine a lower bound for the number of knots in each region of the partitioned space, assuming that a bound on the fourth derivative of the function to approximate is known.

Further studies could include the extension of the results to higher-dimensional spaces, and the search for the bound on the fourth derivative of the function to approximate, given a set of data, but without any other information about the function.

Acknowledgements

The authors thank the reviewers for their suggestions. The paper was partially supported by GACR under Grant 201/93/0427 and 201/96/0917. M. Verleysen is a Research Associate of the Belgian F.N.R.S. (National Fund for Scientific Research).

References

[1] M.H. Schultz, Spline Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1973.
[2] L.L. Schumaker, Spline Functions: Basic Theory, Wiley, New York, 1981.

Kateřina Hlaváčková is a research scientist in the Department of Neural Networks, Institute of Computer Science, Academy of Sciences of the Czech Republic. She deals with mathematical theory of neural networks, in particular the approximation of functions by feedforward networks. She received her Master degree in mathematics in Charles University, Faculty of Mathematics and Physics in 1988 and her PhD degree in computer science in the Institute of Computer Science, Academy of Sciences of the Czech Republic in 1993. She has been an author and co-author of more than 20 research papers on approximation of functions and Kohonen's type self-organization in international journals and conferences. She is a member of organizing and programme committees for several international conferences (MFCS, ICANNGA).

Michel Verleysen is a research associate of the Belgian National Fund for Scientific Research, and invited lecturer at the Catholic University in Louvain (Louvain-la-Neuve, Belgium). He received his engineering and PhD degrees in microelectronics from the Catholic University in Louvain, in 1987 and 1992, respectively. He was an invited professor at the E.P.F. Lausanne in 1992. He has been an author and co-author of more than 50 publications in the field, and wrote a popular book in the "Que Sais-je?" series (in French). Each year he organizes the European Symposium on Artificial Neural Networks, and is the editor-in-chief of the journal Neural Processing Letters. He is a member of the IEEE and the International Neural Network Society.