SPIE-Aerosense96 - 10th Annual International Aerosense Symposium, Vol. 2762- Wavelet Application III:pp 365-374, 1996

Function Approximation by Polynomial Wavelets Generated from Powers of Sigmoids

J. Fernando Marar[1,2] E. C. B. Carvalho Filho[2] G. C. Vasconcelos[2]

[1]Department of Computer Science
Universidade Estadual Paulista - Unesp -
Av. Luiz Edmundo Coube, s/n,
Bauru-SP, Brazil.
[2]Department of Computer Science
Universidade Federal de Pernambuco
Av. Prof. Luiz Freire, s/n,
Cx. Postal 7851, 50732-970, Recife-PE, Brazil.
e-mail : {jfm|ecdbcf|gcv}@di.ufpe.br

Abstract

Wavelet functions have been successfully used in many problems as the activation function of feedforward neural networks [ZB92],[STK92], [PK93]. In this paper, a family of polynomial wavelets generated from powers of sigmoids is described which provides a robust way for designing neural network architectures. It is shown, through experimentation, that function members of this family can present a very good adaptation capability which make them attractive for applications of function approximation. In the experiments carried out, it is observed that only a small number of daughter wavelets is usually necessary to provide good approximation characteristics.

Keywords : Neural Networks, Function Approximation, Wavelets.

1 Introduction

In many problems in applied mathematics, dealing with certain complex functions is not a very easy task, making it necessary to approximate the original function through a combination of simpler functions. The same problem occurs with some functions whose analytical expression is not known a priori in a given environment and the only information available is a few pairs of the function values, obtained through experimental observations. It is in this context that the concepts of function approximation, signal processing, prediction and neural networks come together to create an interesting field of investigation.

The combination of neural networks and some particular types of functions known as wavelets has been used by many researchers as an efficient technique for universal function approximation

[Dau88],[ZB92],[STK92],[PK93]. Both theoretical and practical advances have been achieved, showing the benefits of using wavelets as the activation function for the processing units of neural networks (*wavenets*). Daugman, in [Dau88], has shown how neural networks can be used to learn the best coefficients for approximating images using the Gabor function (Gabor's wavelet). Szu et al [STK92] have used the function $h(t) = \cos(1.75t).\exp(-\frac{t^2}{2})$ as the mother wavelet for the representation and classification of phonemes. Zhang and Beveniste[ZB92] have investigated the power of wavenets as function approximators and, finally, Pati and Krishnaprasad [PK93] have shown the construction of a wavelet function through the combination of sigmoids, and have described how to develop its wavelet transform expansion using neural networks.

This paper, in particular, describes a effective procedure for generating polynomial forms of wavelet functions from the successive powers of sigmoid functions. The resulting functions, referred to as *polynomial wavelets*, represent a robust solution for the construction of neural networks based on wavelets and, more specificallyœ[1, for the definition of analytical polynomial functions for the derivatives of the Gaussian function. Its is shown, through experimentation, that networks of polynomial wavelets can be efficiently adapted to approximate desired functions.

2 Function Approximation and Neural Networks

The interest for the application of neural networks in function approximation was much enhanced after the results obtained by Hornik [Hor89] and Cybenko [Cyb89]. More recently, many works have been able to show the attractive characteristics of using neural networks, more specifically feedforward neural networks, as universal function approximators [Fun89], [HN89], [Hor89],[Cyb89], [LYPS93], [WG95]. In general, every continuous function can be represented by Equation 1:

$$\tilde{f}(x) = \sum_{j=1}^{m} a_i^1 \varphi(\frac{x - a_i^2}{a_i^3}) \quad (1)$$

where a_i^1,a_i^2 e a_i^3 are the coefficients, shifts and dilations of the mother function $\varphi(\bullet)$, respectively [Chu92], and m is the number of basis function employed.

Basically, the problem of function approximation can be described as follows : Let $\Omega = \{(x_i, y_i) : y_i = f(x_i)\}$ be a set of pairs corresponding to sample values generated from an unknown mapping. i.e, $f : \Re^n \rightarrow \Re^m$, where $x_i \in \Re^n$ and $y_i \in \Re^m$. These pairs are, in general, obtained through experimental observation. The set Ω is denoted the training set and the objective is to find a mapping $\tilde{f}$ that approximates $f : \Re^n \rightarrow \Re^m$. In the context of neural networks, this definition can be represented through the architecture illustrated in Figure 2[1].

In this paper, a feedforward neural network is defined by an input (x_i), where n is number of patterns, a set of coefficients $(a_j^1), j = 1..m$ that correspond to the weights of the basis functions, by the elements $(a_j^2), j = 1..m$, and $(a_j^3), j = 1..m$, that represent the coefficients of shift and dilation of the activation function $\varphi(\frac{x-a_1^2}{a_1^3})$, defined for any real value, (i.e., $\varphi : \Re \rightarrow \Re$), and by the variable (y_i) that correspond to the desired output of the network.

[1]Here f is treated as a function $f : \Re \rightarrow \Re$.

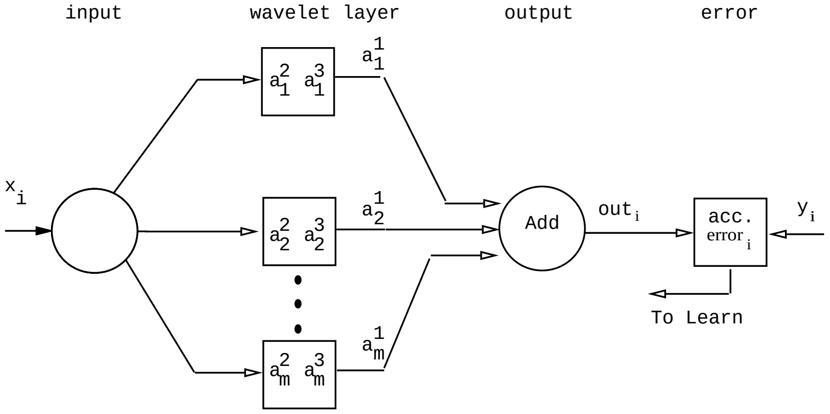

Figure 1: A neural network which employs wavelets as the activation function. The first layer corresponds to the set of adaptable wavelet basis functions. The second (output) layer combines the responses of the basis function and produces an output out_i used to compute the cost function $acc.error_i$. The error signal propagates back to the network for the continuation of the learning process if it exceeds a maximum allowed.

The network parameters (a_i^1,a_i^2 e a_i^3) are gradually adapted so as to minimize the cost function $acc.error_i$, which, in this paper, was chosen to be the function χ^2, defined by Equation 2:

$$\chi^2 = \sum_{i=1}^{n} (\frac{y_i - \sum_{k=1}^{m} a_k^1 \varphi(\frac{x_i - a_k^2}{a_k^3})}{\sigma_i})^2 \tag{2}$$

where n is the number of sample values of the function f and m is the number of basis function employed, usually referred to as daughter functions.

One of the most used activation functions φ_i in neural networks is the sigmoid function, which is given by Equation 3:

$$sig(x) = \frac{1}{1 + \exp(-\beta x)} \tag{3}$$

where β is a constant that determines the degree of smoothness of the function.

The sigmoid function presents some important properties which make them attractive to be used in many problems. One of the most important results concerning this function is attributed to Cybenko which shows that every discriminatory function, constitute a set of basis function for representing any function [Cyb89]. In this paper, however, a different kind of function is considered consisting of polynomial wavelets. These functions have to satisfy the condition of belonging to $L^2(\Re)$ and they have to be wavelets. This implies that these functions have to satisfy the conditions imposed by Equation 4 and Equation 8 presented in the next section, constraints which are not fulfilled by the sigmoid function.

Although the sigmoid function does not satisfy Equations 4 and 8, it is shown in the next section that the composition of powers of the sigmoid function can be used to generate a family of functions for the construction of neural networks that satisfy these conditions. This allows for the definition of a family of polynomial wavelets that has the important property of orthogonality [YR95], that presents the characteristic of spatial location and that has a reasonable locality in the frequency domain. Initially, however, some necessary definitions about wavelet and frames are introduced.

3 Wavelets, Frames and Wavenets

Consider the basic definitions of wavelets for functions belonging to $L^2(\Re)$. The idea is to find a set of functions that suffice the conditions of Equation 1 for function approximation. For this purpose, it is necessary to take into account some of the fundamental properties found in the Hilbert space, which is a space where the inner product is applicable.

Every function, $f(x)$, is said to be square integrable ($L^2(\Re)$) in the Hilbert space, iff it satisfies Equation 4 :

$$\int_{\Re} |f(x)|^2 \, dx < \infty \tag{4}$$

If $f(x)$ and $g(x) \in L^2(\Re)$, then the inner product $< f(x), g(x) >$ is given by Equation 5:

$$\int_{\Re} f(x)\overline{g(x)}dx \tag{5}$$

where $\overline{g(x)}$ corresponds to the complex conjugate of the function $g(x)$, and the norm of $L^2(\Re)$ é s defined by $\|g(x)\|^2 =< g(x), g(x) >$.

Noticing that these concepts are extremely important for the definition of frames in the Hilbert space, the set of all possible functions $\varphi : \Re \rightarrow \Re$ with the following property 6 have to be considered :

$$\Psi = \{r^{\frac{n}{2}}\varphi(r^n x - t)|t, r \in \Re \ , \ n \in Z\} \tag{6}$$

where t and r are shifts and dilations, and $r^{\frac{n}{2}}$ is a constant used to set $\left\| r^{\frac{n}{2}}\varphi(r^n x - t) \right\|^2$ normalized. The set Ψ is said to be a frame, if it satisfies the property that there exists two constants $A > 0$ and $B < \infty$ such that for all the functions $g(x) \in L^2(\Re)$, the relation 7 applies :

$$A \|g(x)\|^2 \leq \sum_{\varphi \in \Psi} |< \varphi, g(x) >|^2 \leq B \|g(x)\|^2 \tag{7}$$

Following directly from relation 7 comes the fact that the family Ψ is dense in $L^2(\Re)$ and Equation 1 is seen as a wavelet expansion of $f(x)$. In order to have a frame from the mother function, $\varphi(\bullet)$, the final condition of admissibility of Equation 8 has to be satisfied:

$$\int \frac{|\widehat{\varphi}(w)|^2}{|w|} dw < \infty \tag{8}$$

In 8, it is sufficient to have $\int \varphi(x)dx = 0$ [Dau92].

3.1 Polynomial Wavelets Generated from Powers of Sigmoids

One important case of wavelet function, shown in Equation 9, that satisfies the condition of admissibility for wavelets as described above, was proposed by Pati e Krishnaprasad [PK93]. This function has been used in different situations with its analytical equation assuming varied forms. One the most popular examples is found in David Marr's work, which presents an application for this function in the detection of contours [Mar82].

$$\varphi(x) = sig(x+p) + sig(x-p) - 2.sig(x) \tag{9}$$

where $p \in Z$ and $\varphi(x)$ is valid only for p ≥ 2.

Consider now Equation 9. If both sides of this equation are divided by p^2 and the limit of the function $\varphi(x)$ is taken when p tends to zero, according to Equation 10 :

$$\lim_{p \to 0} \frac{sig(x+p) + sig(x-p) - 2.sig(x)}{p^2} \tag{10}$$

The original function becomes exactly the second derivative of the sigmoid function. In order to clarify this argument, consider initially the first derivative of the sigmoid function, given by :

$$\frac{dsig(x)}{dx} = \frac{d(\frac{1}{1+\exp(-x)})}{dx} = \frac{\exp(-x)}{(1+\exp(-x))^2}$$

expanding the squared term in the equation above, it is obtained :

$$\frac{dsig(x)}{dx} = \frac{\exp(-x)}{(1+\exp(-x)).(1+\exp(-x))} = sig(x)\left(\frac{\exp(-x)}{1+\exp(-x)}\right) \tag{11}$$

if we add and subtract 1 from Equation 11, we get :

$$\frac{dsig(x)}{dx} = sig(x)\left(\frac{\exp(-x)+1-1}{1+\exp(-x)}\right) = sig(x)\left(\frac{1+\exp(-x)}{1+\exp(-x)}.\frac{-1}{1+\exp(-x)}\right) \tag{12}$$

now, simplifying the terms of Equation 12 provides:

$$\frac{dsig(x)}{dx} = sig(x)(1 - sig(x)) \tag{13}$$

which represents the first derivative of Equation 3. Since the objective is to find the second derivative of the sigmoid function, Equation 13 can be further developed to obtain :

$$\frac{d^2 sig(x)}{dx^2} = sig(x).(1 - sig(x))^2 + (-sig(x)^2.(1 - sig(x))) \tag{14}$$

re-arranging the terms and re-writing Equation 14, the following polynomial form is achieved :

$$\varphi_2(x) = 2.sig(x)^3 - 3.sig(x)^2 + sig(x) \tag{15}$$

where $\varphi_2(x)$ is denominated the first polynomial wavelet function generated by the second derivative of the sigmoid function. Normalizing this function produces Equation 16:

$$\int_{-\infty}^{\infty} |\varphi_2(x)|^2 \, dx = \frac{1}{30} \qquad (16)$$

which determines the normalized expression for the function $\varphi_2(x)$, given by Equation 17:

$$\varphi_{2norm}(x) = \sqrt{30}.\varphi_2(x). \qquad (17)$$

By proceeding in this way of calculating derivatives of sigmoid functions, a family of polynomial wavelets φ_x can be obtained, which can be used as the activation function for the processing units of neural networks.

Equations 18, 19 and 20, for example, represent the analytical expressions for the elements $\varphi_3(x)$,$\varphi_4(x)$, $\varphi_5(x)$, respectively, in the family, and Figure 2 illustrates the nine first polynomial wavelets in the set.

$$\begin{aligned}
\varphi_3 &= -6.sig(x)^4 + 12.sig(x)^3 - 7.sig(x)^2 + sig(x) && (18)\\
\varphi_4 &= 24.sig(x)^5 - 60.sig(x)^4 + 50.sig(x)^3 - 15.sig(x)^2 + sig(x) && (19)\\
\varphi_5 &= -120.sig(x)^6 - 360.sig(x)^5 - 390.sig(x)^4 + 180.sig(x)^3 - 31.sig(x)^2 + sig(x) && (20)
\end{aligned}$$

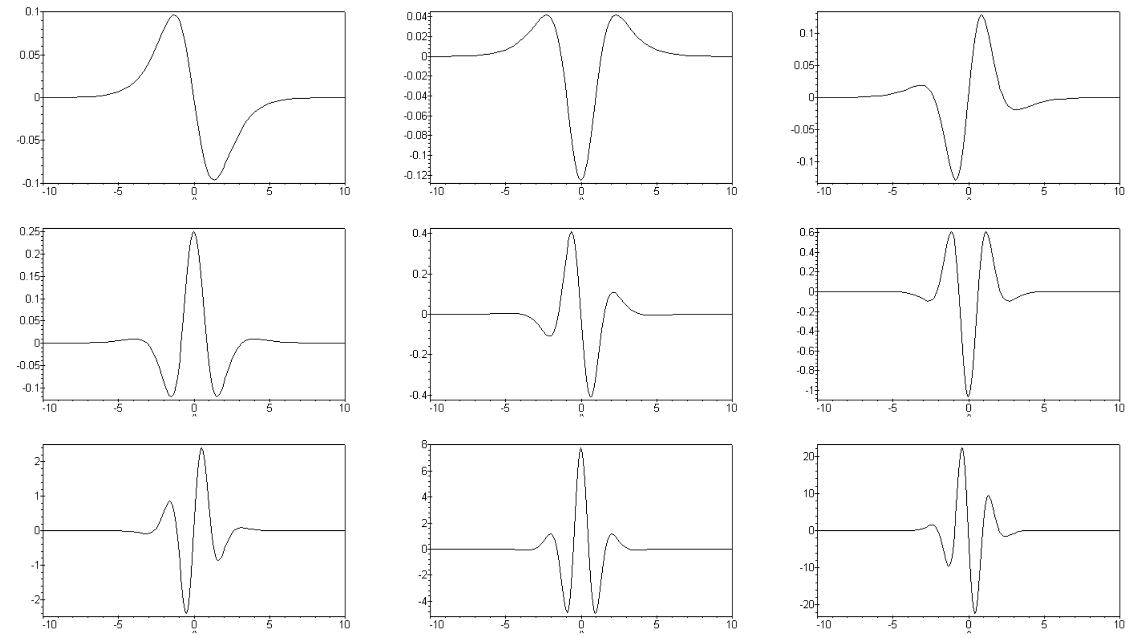

Figure 2: Family of polynomial wavelets obtained from powers of sigmoids. The diagrams correspond, respectively, from left to right, top to bottom, to the elements φ_2,φ_3,φ_4, φ_5,φ_6 , φ_7 φ_8,φ_9 and φ_{10}. in the family

One interesting observation about the polynomials described above is that they constitute a different way of representing successive derivatives of the Gaussian function in the form of polynomials. This procedure of finding successive derivatives of the Gaussian function has been useful when utilized in conjunction with the polynomials of Hermite [YR95] to study some processes in the human spatial vision.

Consider the derivatives of the Gaussian function in its standard form, given by the Equation :

$$G_n(x) = \frac{d^n \exp(-x^2)}{dx^n}, n = 0, 1, 2, ...$$

For any analytical expression of the Gaussian function, the successive derivatives of this function produces the same waveforms shown in Figure 2, apart from the characteristics of phase and amplitude [2]. However, it is not always the case that the analytical expressions generated initially from Gaussians are as elegant as those observed in the polynomials forms obtained when the sigmoid function is used as the seed.

The most important aspect to take into account is that the family of polynomials wavelets, $\{\varphi_i\}$, presented here is composed of functions belonging to $L^2(\Re)$ which satisfy the condition of admissibility imposed by Equation 8. As a result, the set of functions obtained from shifts and dilations of each of these functions creates a dense family in the space of functions square integrable.

4 Experimental Results with the Polynomial $\varphi_2(x)$

In this section, it is shown how the particular function $\varphi_2(x)$ taken from the family $\{\varphi_i\}$ can be adapted to approximate a given function. The process of network learning takes place using the cost function defined by Equation 2. In principle, any learning algorithm based on gradient descent or the conjugate gradient can be used to adapt the network's parameters. However, in the experiments carried out, the method of Levenberg-Marquardt [Mar63],[PFTV92] was chosen because it retains some important characteristics of gradient descent and also makes use of the speed of the method of the Hessian matrix inverse.

The training data used in the experiments was obtained from the function defined in Equation 21 whose diagram is shown in Figure 3.

$$f(x) = sig(x).(cos(x) - sin(x)) \tag{21}$$

The network simulation was carried out using 9 wavelets and 90 training samples from the function 21, uniformly distributed in the range $[-6, 6]$. The error criterion for stopping the training process was chosen to be 0.01. The network weights were initialized with random values. The results of the function approximation obtained with the method of Levemberg-Marquardt with the mother function $\varphi_2(x)$ is depicted in Figure 4. In this figure, it is shown 4 different points in the training phase. The first diagram corresponds to the function approximation after 28 training epochs and the other diagrams represent 65, 164 and 198, respectively. The process was stopped after only 198 iterations, denoting a fast convergence of the method according to the error criterion.

Table 1 illustrates the final values for the weight, dilations and shifts of the 9 daughter functions $\varphi_2(x)$. These functions combined at the output layer of the network represent a very close approximation to the function $sig(x).(cos(x) - sin(x))$ within the acceptable margin of error. The final

[2] parameters which can be compensated by a constant.

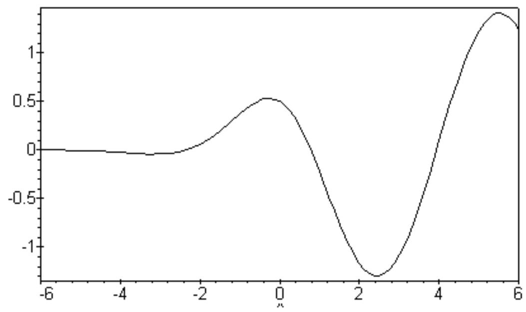

Figure 3: function $sig(x).(cos(x) - sin(x))$.

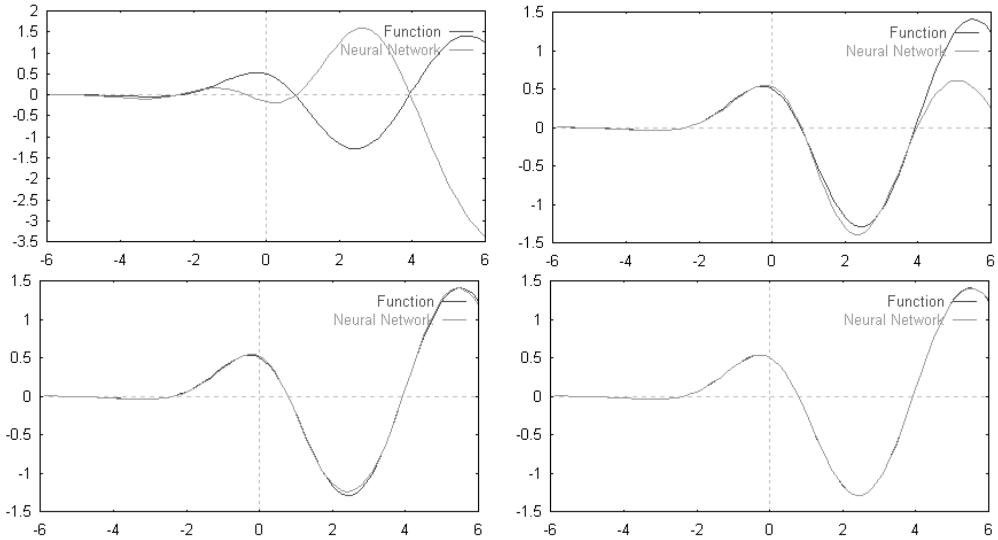

Figure 4: Adaptation process of the daughter wavelets. The first diagram (left corner, on the top) corresponds to 28 training epochs. The next diagrams (from left to right, top to bottom) represent 65, 164 and 198 epochs, respectively.

configurations for the daughter wavelets can also be observed through the diagrams shown in Figure 5.

5 Conclusions

In this paper, a family of polynomial wavelets was presented constructed from powers of sigmoids. It has been shown, in the context of function approximation, that this set of functions can provide a very good approximation capability, with a fast convergence of the training process. One of the advantages of using wavelets in the design of neural networks is the reduced number of basis function usually required to give a good approximation to a function [STK92],[PK93],[MFV96]. In particular, the procedure for defining the family of functions described in this paper possesses the attractive characteristic of providing a simple way of defining wavelet functions. This family of functions also

weights(a^1)	dilations(a^2)	shifts(a^3)
2.11992	-2.84857	7.60694
0.77675	-1.92266	10.28019
-14.51028	4.06153	1.43123
2.58381	0.98175	11.66697
-4.15489	-0.52987	1.22267
3.92829	2.40124	3.01965
-28.41081	-149.37083	-36.42442
7.39287	1.08307	1.23855
16.06220	7.45983	2.30509

Table 1: Final values for the weights, dilations and shifts for the daughter functions.

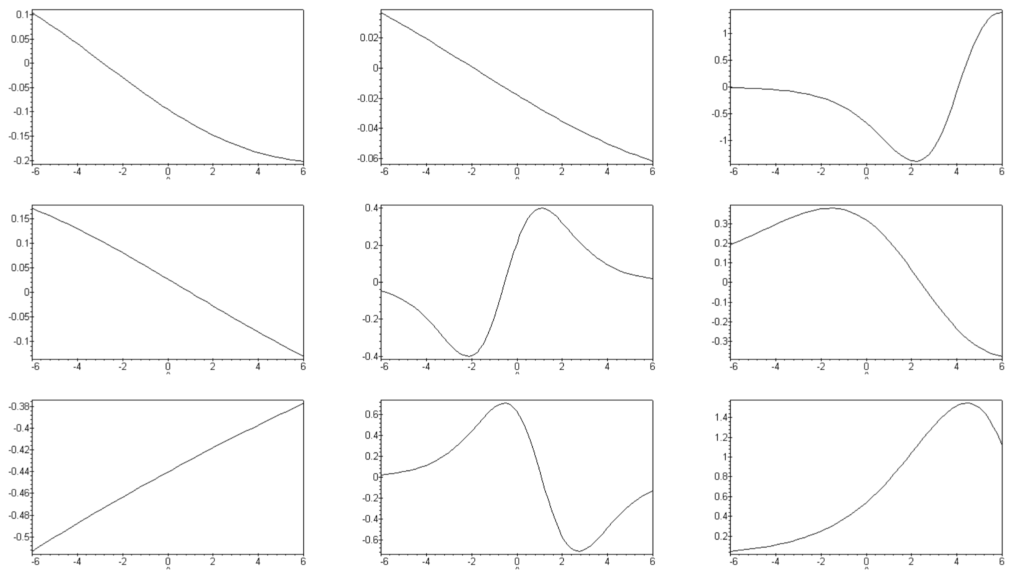

Figure 5: Final shape of the 9 daughter wavelets used to approximate the function $f(x) = sig(x).(cos(x) - sin(x))$.

represents a potential scheme for the development of efficient neural network structures in pattern classification problems [STK92].

5.1 Acknowledgements

The authors would like to acknowledge the Brazilian Research Council (CNPq), the Brazilian Federal Agency for Postgraduate Studies (CAPES) and UNESP-Bauru for partially supporting this work.

References

[Chu92] C.S. Chui. *An Introduction to Wavelets.* Academic Press,Inc,Boston, 1992.

[Cyb89] G. Cybenko. Approximation by superposition of a sigmoidal function. *Mathematics of Control, signals and Systems*, 2:303–314, 1989.

[Dau88] J. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoust. Speech,Signal Proc.*, 8, July 1988.

[Dau92] I. Daubechies. *Ten Lectures on Wavelets.* SIAM,Philadelphia, 1992.

[Fun89] K. Funahashi. On the approximate realization of continuos mappings by neural networks. *Neural Networks*, 2:183–192, April 1989.

[HN89] R. Hecht-Nielsen. Theory of the backpropagation neural network. *Proc. IJCNN (Washington) D.C.*, I-593:18–22, June 1989.

[Hor89] K. Hornik. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.

[LYPS93] M. Leshno, L.V Ya, A. Pinkus, and S. Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6:861–867, 1993.

[Mar63] D.W. Marquardt. *Journal Soc. Ind. Appl. Math*, 11:431–441, 1963.

[Mar82] D. Marr. *Vision.* Freeman- New York, 1982.

[MFV96] J.Fernando Marar, E.C.B.Carvalho Filho, and G.C. Vasconcelos. Function approximation by classical methods and neural networks. *Tech. Report - in preparation*, 1996.

[PFTV92] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T Vetterling. *Numerical Recipes in C.* Cambridge University Press, Cambridge, 1992.

[PK93] Y.C. Pati and P.S. Krishnaprasad. Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformations. *IEEE Transactions on Neural Networks*, 4:73–85, January 1993.

[STK92] Harold Szu, Brian Telfer, and Shubha Kadambe. Neural network adaptive wavelets for signal representation and classification. *Optical Engineering*, 31:1907–1916, September 1992.

[WG95] J. Wray and G.G.R. Green. Neural network, approximation theory, and finite precision computation. *Neural Network*, 8:31–37, January 1995.

[YR95] J. Yang and A. Reeves. Bottom-up visual image processing probed with weighted hermite polynomials. *Neural Network*, 8:669–691, May 1995.

[ZB92] Qinghua Zhang and Albert Benveniste. Wavelet networks. *IEEE Transactions on Neural Networks*, 3:889–898, November 1992.