NEURAL NETWORKS AND APPROXIMATION BY SUPERPOSITION OF GAUSSIANS

Paulo Jorge S. G. Ferreira

Dep. de Electrónica e Telecomunicações / INESC
Universidade de Aveiro
3810 Aveiro Portugal

ABSTRACT

The aim of this paper is to discuss a nonlinear approximation problem relevant to the approximation of data by radial-basis-function neural networks. The approximation is based on superpositions of translated Gaussians. The method used enables us to give explicit approximations and error bounds. New connections between this problem and sampling theory are exposed, but the method used departs radically from those commonly used to obtain sampling results since (i) it applies to signals that are not band-limited, and possibly even discontinuous (ii) the sampling knots (the centers of the radial-basis functions) need not be equidistant (iii) the basic approximation building block is the Gaussian, not the usual sinc kernel. The results given offer an answer to the following problem: how complex should a neural network be in order to be able to approximate a given signal to better than a certain prescribed accuracy? The results show that $O(1/N)$ accuracy is possible with a network of N basis functions.

1. INTRODUCTION

Radial-basis-function neural networks were developed for data interpolation but have been successfully applied to other problems, including adaptive equalization and spread spectrum systems [1]. The key problem is the approximation of a given f by

$$\sum_{i=1}^{N} a_i \, g(\|t - t_i\|).$$

There are several possible choices for g, including the multi-quadratic, inverse multi-quadratic and Gaussian.

This paper turns around the approximation properties of the Gaussian function

$$g(t, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-t^2/2\sigma_i^2},$$

which can be used as a building block for approximating other continuous or even discontinuous functions to arbitrarily small tolerances. The simplest way in which Gaussian functions can be combined is illustrated by

$$\sum_{i=1}^{N} a_i \, g(t - t_i, \sigma_i). \tag{1}$$

These linear combinations of translated Gaussians have very good approximation properties. A recent and interesting work [2] addresses the approximation of finite-energy signals in this way. It argues that these representations can be used to represent signals and images efficiently, to the point of competing with standard transform techniques such as discrete cosine transform coding.

This paper discusses the extent to which a given function can be approximated by superpositions of translated Gaussian functions, or, equivalently, the approximating properties of Gaussian radial-basis-function neural networks. Section 2 recalls a general property of functions of the form

$$\sum_{i=1}^{N} a_i \, g(t - t_i, \sigma),$$

and section 3 contains the main result, a constructive approach that yields the coefficients a_i as samples of the function to be approximated, that is, $a_i = f(t_i)$. The set $\{t_i\}$ is subject only to weak restrictions. In a sense, we are in presence of a nonuniform sampling theorem for functions that are not band-limited, using a Gaussian function instead of the usual sinc kernel. The approximation result yields the values of the parameters of the approximating Gaussian functions without

Fax +351-34-370545, e-mail `pjf@inesca.pt`. This work was supported by JNICT.

any computational effort, and allows arbitrarily small errors: the error is $O(1/N)$.

The approximation properties of sigmoidal functions were discussed in [3], and a similar $O(1/N)$ result is known for sigmoidal functions [4]. It has also been shown [5] that radial-basis-function neural networks can form arbitrarily close approximations for continuous functions, provided that there are enough basis functions. Our results are quite distinct from these: we concentrate on superpositions of Gaussians, not sigmoidal functions (indeed, our methods apply for many other bell-shaped curves, not necessarily Gaussian); we impose distinct or weaker restrictions on the functions to be approximated (continuity is not required); our methods are constructive and shed some light on the connections between this problem and certain generalized sampling results.

Over a thousand papers have been written on sampling (see the references in [6–8], for example). There are many nonuniform sampling results for band-limited signals [9–16], as well as many uniform sampling results for not necessarily band-limited signals [17]. However, almost nothing is known about one aspect of the problem studied here: nonuniform sampling expansions for functions that are not band-limited, either for the sinc kernel, or for other kernels. The works [18, 19] are exceptions. More recently, [20] discusses possible connections with number theory, and [21] addresses a similar problem in the context of multiresolution analyses of L_2.

2. APPROXIMATION BY SUPERPOSITIONS OF A SINGLE GAUSSIAN FUNCTION

A theorem of Wiener shows that any function belonging to L_1 can be approximated to any prescribed tolerance, in the L_1 norm, by linear combinations of the translates of a single function $\psi \in L_1$

$$\sum_{i=1}^{N} a_i\,\psi(t - t_i),$$

if and only if the Fourier transform of ψ has no zeros. Wiener also showed that a similar result holds in L_2 if and only if the set of zeros of the Fourier transform of ψ has zero measure. For proofs of these results see [22–24], for example. Since the Gaussian function $g(t, \sigma)$ belongs to both L_1 and L_2, independently of σ, and its Fourier transform has no zeros, the results of Wiener imply that, for any $f \in L_1$ and $\epsilon > 0$, there is an integer N and constants $(a_i)_{1 \le i \le N}$ and $(t_i)_{1 \le i \le N}$ such that

$$\int_{-\infty}^{\infty} \left| f(t) - \sum_{i=1}^{N} a_i\, g(t - t_i, \sigma) \right| dt < \epsilon.$$

A similar result holds for any $f \in L_2$, the approximation being now in the L_2 norm,

$$\int_{-\infty}^{\infty} \left| f(t) - \sum_{i=1}^{N} a_i\, g(t - t_i, \sigma) \right|^2 dt < \epsilon,$$

that is, in the least-squares sense. These immediate corollaries of Wiener's approximation results generalize the results obtained, at much greater length, in [2].

It is remarkable that this approximation property holds independently of the value of σ. Note that the spaces L_1 and L_2 contain very rapidly varying and discontinuous functions. Any of these functions can be arbitrarily well approximated by Gaussian curves, no matter what value of σ is selected, that is, no matter how much spread out the Gaussian is.

The problem is how to select N, $(a_i)_{1 \le i \le N}$ and $(t_i)_{1 \le i \le N}$, and how to obtain error bounds. We will discuss an approach which leads to an approximation theorem that has the following characteristics: (i) it solves the evaluation problem of N, $(a_i)_{1 \le i \le N}$ and $(t_i)_{1 \le i \le N}$ with no computational effort at all (ii) it applies to signals that are not band-limited, possibly even discontinuous (iii) the sampling expansions need not be based on equidistant sets of samples or knots (iv) the basic approximation building block is the Gaussian function, not the usual sinc kernel (v) it yields an useful upper bound for the approximation error.

3. MAIN RESULT

Consider the convolution

$$f_\sigma(t) = \int_{-\infty}^{+\infty} f(\tau) g(t - \tau, \sigma)\, d\tau.$$

Generally speaking, we will require that f be such that $f_\sigma \to f$ in the pointwise sense as $\sigma \to 0$. There are several well-known conditions that ensure this, for the Gaussian kernel as well as for others (see, for example, [23]).

Theorem 1 *Let $f : \mathbb{R} \to \mathbb{R}$ be a function of bounded variation satisfying the above hypothesis, and vanishing outside* $[0, 1]$. *Denote by* $\{t_i\}$ *any N reals such that*

$$\frac{i-1}{N} < t_i < \frac{i}{N}, \tag{2}$$

for $1 \le i \le N$, and let

$$f_\sigma(t) = \int_0^1 f(\tau) g(t-\tau,\sigma)\, d\tau. \qquad (3)$$

Then,

$$\left| f_\sigma(t) - \frac{1}{N}\sum_{k=1}^{N} f(t_k) g(t-t_k,\sigma) \right| \le$$
$$\le \frac{1}{\sigma N} \frac{V_f + M_f}{\sqrt{2\pi}},$$

where M_f denotes the maximum value of f, and V_f its variation.

Proof: To simplify the notation let

$$s_N(t) = \frac{1}{N}\sum_{k=1}^{N} f(t_k) g(t-t_k,\sigma). \qquad (4)$$

Applying the mean value theorem to (3) yields

$$f_\sigma(t) = \frac{1}{N}\sum_{k=1}^{N} f(\xi_k) g(t-\xi_k,\sigma),$$

where the N reals ξ_i $(1 \le i \le N)$ satisfy (2). This leads to

$$\begin{aligned} &|f_\sigma(t) - s_N(t)| \le \\ &\le \frac{1}{N}\sum_{k=1}^{N} |f(\xi_k) g(t-\xi_k,\sigma) - f(t_k) g(t-t_k,\sigma)| \\ &\le \frac{V_F(t)}{N}, \end{aligned}$$

where $V_F(t)$ denotes the variation of the function

$$F(x) = f(x) g(x-t,\sigma)$$

in $[0,1]$, which is a function of t. But

$$V_F(t) \le V_f \|g\|_\infty + V_k(t)\|f\|_\infty.$$

Now $\|g\|_\infty$ is the maximum value of $g(x,\sigma)$, that is

$$g(0,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}.$$

The quantity $V_k(t)$ denotes the variation of $k(x) = g(x-t,\sigma)$, for $x \in [0,1]$. This is bounded by the total variation of $k(x)$,

$$V_k(t) \le \frac{2}{\sigma\sqrt{2\pi}}.$$

■

Note that the approximation error is $O(1/N)$ in the sup norm. The theorem still holds if f is discontinuous (but still of bounded variation). In this case the proofs are somewhat longer, and require the second mean-value theorem [25].

As σ grows, f_σ becomes an increasingly better approximation to f ($f_\sigma \to f$ in the sense of pointwise convergence). By selecting σ and N, it is always possible to obtain arbitrarily good approximations to f, since

$$\|f - s_N\| \le \|f - f_\sigma\| + \|f_\sigma - s_N\|$$

in any norm. The method used is similar to the method used in [18,19] to derive nonuniform sampling theorems for the sinc and other kernels.

The result given offers an answer to the following problem: how complex should a radial-basis-function neural network be in order to be able to approximate a given signal to better than a certain prescribed accuracy? The results show that $O(1/N)$ accuracy is possible with a network of N basis (Gaussian) functions.

4. REFERENCES

[1] B. Mulgrew. Applying radial basis functions. *IEEE Sig. Proc. Mag.*, 13(2):50–65, March 1996.

[2] J. Ben-Arie and K. R. Rao. Nonorthogonal signal representation by gaussians and Gabor functions. *IEEE Trans. Circuits Syst. II*, 42(6):402–413, June 1995.

[3] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems*, 2:303–314, 1989.

[4] A. R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inform. Theory*, 39(3):930–945, May 1993.

[5] J. Park and I. W. Sandberg. Universal approximation using radial-basis-function networks. *Neural Comp.*, 3:245–257, 1991.

[6] R. J. Marks II, editor. *Advanced Topics in Shannon Sampling and Interpolation Theory.* Springer, Berlin, 1993.

[7] A. I. Zayed. *Advances in Shannon's Sampling Theory.* CRC Press, Boca Raton, 1993.

[8] J. R. Higgins. *Sampling Theory in Fourier and Signal Analysis: Foundations.* Oxford University Press, Oxford, 1996.

[9] J. L. Brown, Jr. On the error in reconstructing a non-bandlimited function by means of the bandpass sampling theorem. *J. Math. Anal. Appl.*, 18:75–84, 1967.

[10] P. L. Butzer and W. Splettstößer. A sampling theorem for duration-limited functions with error estimates. *Infor. Control*, 34:55–65, 1977.

[11] P. L. Butzer and W. Splettstößer. On quantization, truncation and jitter errors in the sampling theorem and its generalizations. *Sig. Proc.*, 2:101–112, 1980.

[12] R. L. Stens. Approximation to duration-limited functions by sampling sums. *Sig. Proc.*, 2:173–176, 1980.

[13] S. Cambanis and M. K. Habib. Finite sampling approximations for non-band-limited signals. *IEEE Trans. Inform. Theory*, 28(1):67–73, January 1982.

[14] R. L. Stens. A unified approach to sampling theorems for derivatives and Hilbert transforms. *Sig. Proc.*, 5:139–151, 1983.

[15] P. L. Butzer, S. Ries, and R. L. Stens. Approximation of continuous and discontinuous functions by generalized sampling series. *J. Approx. Theory*, 50(1):25–39, 1987.

[16] W. Engels, E. L. Stark, and L. Vogt. Optimal kernels for a general sampling theorem. *J. Approx. Theory*, 50(1):69–83, 1987.

[17] P. L. Butzer and R. L. Stens. Sampling theory for not necessarily band-limited functions: a historical overview. *SIAM Rev.*, 34(1):40–53, March 1992.

[18] P. J. S. G. Ferreira. Nonuniform sampling of nonbandlimited signals. *IEEE Sig. Proc. Letters*, 2(5):89–91, May 1995.

[19] P. J. S. G. Ferreira. Approximating non-band-limited functions by nonuniform sampling series. In *SampTA'95, 1995 Workshop on Sampling Theory and Applications*, pages 276–281, Jurmala, Latvia, September 1995.

[20] P. J. S. G. Ferreira. On the approximation of non-bandlimited signals by nonuniform sampling series. In *Proceedings of EUSIPCO-96, VIII European Signal Processing Conference*, pages 1567–1570, Trieste, Italy, September 1996.

[21] P. J. S. G. Ferreira. Approximation by nonuniform sampling series and multiresolution analysis. In *Proceedings of the 7th IEEE Digital Signal Processing Workshop, DSPW-96*, pages 137–140, Loen, Norway, September 1996.

[22] N. Wiener. *The Fourier Integral and Certain of Its Applications*. Cambridge University Press, Cambridge, 1933. Reprinted by Dover Publications, 1958.

[23] K. Chandrasekharan. *Classical Fourier Transforms*. Springer, Berlin, 1989.

[24] N. I. Achieser. *Theory of Approximation*. Frederick Ungar Publishing Co., New York, 1956. Reprinted by Dover Publications, 1992.

[25] E. C. Titchmarsh. *The Theory of Functions*. Oxford University Press, Oxford, second edition, 1968.