DOUBLY-FINITE VOLTERRA-SERIES APPROXIMATIONS

Irwin W. Sandberg

Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX 78712, USA
sandberg@ece.utexas.edu

ABSTRACT

We consider causal time-invariant nonlinear input-output maps that take a set of bounded functions into a set of real-valued functions, and we give criteria under which these maps can be uniformly approximated arbitrarily well using a certain structure consisting of a not-necessarily linear dynamic part followed by a nonlinear memoryless section that may contain sigmoids or radial basis functions, etc. As an application of the results, we show that system maps of the type addressed can be uniformly approximated arbitrarily well by doubly-finite Volterra-series approximants if and only if these maps have approximately-finite memory and satisfy certain continuity conditions. Corresponding results have also been obtained for (not necessary causal) multivariable input-output maps. Such multivariable maps are of interest in connection with image processing.

1. INTRODUCTION

Results concerning the representation and approximation of nonlinear maps can be of particular interest in connection with a variety of engineering problems. In [1] there began a study of the network (e.g., neural network) approximation of functionals and approximately-finite-memory maps. It was shown that large classes of approximately-finite-memory maps can be uniformly approximated (i.e., uniformly approximated arbitrarily well) by the maps of certain simple nonlinear structures using, for example, sigmoidal nonlinearities or radial basis functions.[1] This is of interest in connection with, for example, the general problem of establishing a comprehensive analytical basis for the identification of dynamic systems.[2] The approximately-finite-memory approach in [1] is different from, but is related to, the fading-memory approach in [6] where it is proved that certain scalar single-variable causal fading-memory systems with inputs and outputs defined on $\mathbb{R}$ or on $(\ldots,-1,0,1,\ldots)$ can be approximated by a finite Volterra series.

The study in [1] addresses noncausal as well as causal systems, and also systems in which inputs and outputs are functions of several variables. In recent papers [7, 8] strong corresponding results are given within the framework of an extension of the fading-memory approach. We use the term "myopic" to describe the maps we study because the term "fading-memory" is a misnomer when applied to noncausal systems, in that noncausal systems may anticipate as well as remember. Roughly speaking, an input-output map K is myopic if the value of $(Ku)(\gamma)$ is always relatively independent of the values of u at points remote from γ. [The concepts of maps that are myopic, have approximately-finite memory, fading memory [6], or decaying memory [9] are all different but are all related in that they are alternative ways of making precise, in different settings, the same general idea. There is also a history of the use of this idea in other areas and for purposes other than approximation (see, for example, [10],[11],[12],[13]).]

In [8], as well as in [6, 7] and other papers, attention centers around properties of nonlinear approximation structures of the type indicated in Figure 1 in which the box labeled N is a memoryless nonlinear system and the h_j denote linear maps. This is a structure consisting of a linear preprocessing stage followed by a memoryless nonlinear network. As mentioned earlier, such structures were first considered in an important

[1]It was later found [2] that the approximately-finite-memory condition is met by the members of a certain familiar class of stable continuous-time systems.

[2]It was also observed that any continuous real functional on a compact subset of a real normed linear space can be uniformly approximated using only a feedforward neural network with a linear-functional input layer and one hidden memoryless nonlinear (e.g., sigmoidal) layer. This has applications concerning, for instance, the theory of classification of signals, and is a kind of general extension of an idea due to Wiener concerning the approximation of input-output maps using a structure consisting of a bank of linear maps followed by a memoryless map with several inputs and a single output (see, for instance, [3, pp. 380-382]). For related results, see [4] and [5].

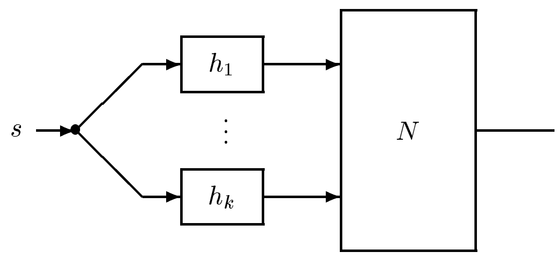

Figure 1: Approximation structure.

but very special context by Wiener. Roughly speaking, the main result in [8] is that, with N containing sigmoids, or radial basis functions, etc., a given shift-invariant input-output map K can be uniformly approximated over a certain set U of inputs if and only if K is myopic, assuming that the linear maps represented by the h_j satisfy certain conditions. A corresponding result for a different type of input set is given in [7].

In this paper attention is focused on the h_j. We consider causal time-invariant input-output maps G that take a set S of bounded vector-valued functions into a set of real-valued functions, and we give conditions on the h_j under which these G's can be uniformly approximated arbitrarily well using the structure shown in Figure 1. In our results certain separation conditions, of the kind associated with the Stone-Weierstrass theorem, play a prominent role. Here they emerge as criteria for approximation, and not just sufficient conditions under which an approximation exists. As an application of the results, we show that system maps of the type addressed can be uniformly approximated arbitrarily well by doubly-finite Volterra-series approximants *if and only if* these maps have approximately-finite memory and satisfy certain continuity conditions. By such an approximant we mean one of the form

$$\sum_{j=1}^{p} \sum_{n_j=0}^{n} \cdots \sum_{n_1=0}^{n} k_j(n-n_1,\ldots, n-n_j) s(n_1)\cdots s(n_j)$$

in which p is finite and there is a number η such that for each j $k_j(n_1,\ldots,n_j)$ vanishes if one or more of the $n_1,\ldots,n_j$ exceed η (which implies that the approximant has finite memory). Corresponding results – not presented here because of length restrictions – have also been obtained for (not necessary causal) multivariable input-output maps. Such multivariable maps are of interest in connection with image processing.

Our results are given in the next section, which begins with a section on preliminaries. As we have said, in these results certain separation conditions, of the kind associated with the Stone-Weierstrass theorem, play a prominent role but here they emerge as criteria for approximation, and not just sufficient conditions under which an approximation exists. In particular, a corollary of one of our main results in the next section is a theorem to the effect that universal approximation can be achieved using the structure of Figure 1 *if and only if* the set H from which the h_j are drawn satisfies the separation condition that for each $n:$ $(hu_1)(n) \neq (hu_2)(n)$ for some $h \in H$ whenever $u_1, u_2 \in S$ and $u_1(j) \neq u_2(j)$ for at least one $j \leq n$.. This holds even if the elements of H are not linear. For a related result in the context of "complete memories," see [14] (see also [4, Theorem 4] and [15]).

2. APPROXIMATION OF INPUT-OUTPUT MAPS

2.1. Preliminaries

The *linear-ring operations* starting with a set of real numbers consist of the linear operations and multiplication. That is, these operations consist of ordinary addition, multiplication, and multiplication by real scalars, with the understanding that operations may be performed only on numbers in the starting set and/or numbers that have been formed from the starting set. Let k be a positive integer. We say that a map $M : \mathbb{R}^k \to \mathbb{R}$ is a *linear-ring map* if Mv is generated from the components $v_1, \ldots, v_k$ of v by a finite number of linear ring operations that do not depend on v. Let $L(\mathbb{R}^k, \mathbb{R})$ stand for the set of all linear-ring maps from $\mathbb{R}^k$ to $\mathbb{R}$. We view the elements of $\mathbb{R}^k$ as row vectors.

Let $C(\mathbb{R}^k, \mathbb{R})$ denote the set of continuous maps from $\mathbb{R}^k$ to $\mathbb{R}$, and let D_k stand for any subset of $C(\mathbb{R}^k, \mathbb{R})$ that is dense in $L(\mathbb{R}^k, \mathbb{R})$ on compact sets, in the sense that given $\epsilon > 0$ and $f \in L(\mathbb{R}^k, \mathbb{R})$, as well as a compact $K \subset \mathbb{R}^k$, there is a $g \in D_k$ such that $|f(v) - g(v)| < \epsilon$ for $v \in K$. The D_k can be chosen in many different ways, and may involve, for example, radial basis functions, polynomial functions, piecewise linear functions, sigmoids, or combinations of these functions.[3]

Let d be a positive integer, and let $\mathcal{C}$ stand for any bounded closed subset of $\mathbb{R}^d$ that contains the origin of $\mathbb{R}^d$. For example, $\mathcal{C}$ can be chosen to be $\{v \in \mathbb{R}^d : \|v\| \leq \gamma\}$ where $\|\cdot\|$ is any norm on $\mathbb{R}^d$ and γ is a positive constant. With $\mathcal{Z}_+ =: \{0, 1, \ldots\}$, let S denote

[3] The term D_k is used also in, for example, [7] where the meaning is different. Here the conditions on the D_k are even less restrictive.

the family of all maps s from $\mathcal{Z}_+$ to $\mathcal{C}$. The set S is our set of inputs.

For each α and β in $\mathcal{Z}_+$, let maps $W_{\beta,\alpha} : S \to S$ and $T_\beta : S \to S$ be defined by

$$(W_{\beta,\alpha}s)(n) = \begin{cases} s(n), & \beta - \alpha \le n \le \beta \\ 0, & \text{otherwise} \end{cases}$$

and

$$(T_\beta s)(n) = \begin{cases} 0, & n < \beta \\ s(n-\beta), & n \ge \beta \end{cases} .$$

We say that a map M from S into the set of real-valued functions on $\mathcal{Z}_+$ is *time-invariant* if for each $\beta \in \mathcal{Z}_+$ we have

$$(MT_\beta s)(n) = \begin{cases} 0, & n < \beta \\ (Ms)(n-\beta), & n \ge \beta \end{cases}$$

for all s. M is *causal* if $(Mu)(n) = (Mv)(n)$ whenever $n \in \mathcal{Z}_+$ and u and v satisfy $u(\alpha) = v(\alpha)$ for $\alpha \le n$.

Throughout the paper, G denotes a causal time-invariant map from S to the set of real-valued functions defined on $\mathcal{Z}_+$. We assume that G has *approximately-finite memory* in the sense that given $\epsilon > 0$ there is an $\alpha \in \mathcal{Z}_+$ such that

$$|(Gs)(n) - (GW_{n,\alpha}s)(n)| < \epsilon, \quad n \in \mathcal{Z}_+$$

for $s \in S$.[4]

For each $n \in \mathcal{Z}_+$, let c_n stand for $\{0, 1, \ldots n\}$, and let S_n denote the restriction of S to c_n. We view each S_n as a metric space with metric ρ_n defined by $\rho_n(x,y) = \max\{\, \|x(j) - y(j)\| : j \in c_n \,\}$.

Let H be a family of time-invariant causal maps h from S to the set of $I\!R$-valued functions defined on $\mathcal{Z}_+$, and for each h and each n in $\mathcal{Z}_+$ define the functional $q(h,n,\cdot)$ on S_n by $q(h,n,u) = (hs)(n)$, where s is any element of S whose restriction to c_n is u. We assume that $q(h,n,\cdot)$ is continuous for each $h \in H$. We also assume that H is closed under the memory-limiting operation, in the sense that g defined on S by $(gs)(n) = (hW_{n,\alpha}s)(n)$ belongs to H whenever $h \in H$ and $\alpha \in \mathcal{Z}_+$.[5]

As an example, we can take H to be the set H_0 of all maps h for which

$$(hs)(n) = \phi\Big[\sum_{j=0}^{n} s(j)a(n-j)\Big], \quad n \in \mathcal{Z}_+ \qquad (1)$$

where ϕ, which depends on h, is a continuous map from $I\!R$ into $I\!R$ with $\phi(0) = 0$, and a, which also depends on h, is real $d \times 1$ matrix valued with $a(j)$ the zero $d \times 1$ matrix for j sufficiently large. As another example, note that H can be taken to be any subset of H_0 that is closed under the memory-limiting operation.

For each $n \in \mathcal{Z}_+$, let F_n denote the functional defined on S_n by $F_n u = (Gs)(n)$, where s is any element of S whose restriction to c_n is u. We shall use A.1 to denote the following condition:

> For each $n \in \mathcal{Z}_+$ and each $(u_1, u_2) \in E_n$ there is an $h \in H$ such that $q(h,n,u_1) \ne q(h,n,u_2)$,

in which

$$E_n = \{(u_1,u_2) \in S_n \times S_n : F_n u_1 \ne F_n u_2\}.$$

2.2. Approximation and Discrete-Time Systems

One of our main results is the following.

Theorem 1: The following two statements are equivalent.

(i) For each $\epsilon > 0$, there are an $\alpha \in \mathcal{Z}_+$, a positive integer k, elements $h_1, \ldots, h_k$ of H, and an $N \in D_k$ such that

$$|(Gs)(n) - N[(MW_{n,\alpha}s)(n)]| < \epsilon, \quad n \in \mathcal{Z}_+$$

for all $s \in S$, where

$$(Ms)(n) = [(h_1 s)(n), \ldots, (h_k s)(n)].$$

(ii) Each F_n is continuous and A.1 is met.

The proof makes use of the following lemma, but the remaining details are omitted in this version of the paper.

Lemma 1: Let A be a compact topological space, let f belong to the set C of all continuous real-valued functions on A, and let B be a subset of C. Suppose that there is an element a of A such that $f(a) = 0$ and $b(a) = 0$ for $b \in B$. Then statements 1) and 2) below are equivalent.

1) For each $\epsilon > 0$ there are a positive integer k, elements $b_1, \ldots, b_k$ of B, and an $N \in D_k$ such that $|f(x) - N[B(x)]| < \epsilon$ for $x \in A$, where $B(x) = [(b_1)(x), \ldots, (b_k)(x)]$.

2) For $(x_1, x_2) \in \{(x_1,x_2) \in A \times A : f(x_1) \ne f(x_2)\}$ there is a $b \in B$ such that $b(x_1) \ne b(x_2)$.

[4] There is a slight difference here relative to the definition of approximately-finite memory in [1] where α is required to be positive.

[5] In this connection, it is not difficult to check that g defined above is causal and time invariant for $h \in H$ and $\alpha \in \mathcal{Z}_+$.

Our next result is a corollary of Theorem 1. It focuses attention on the conditions required of the preprocessing stage represented by the h_j in Figure 1 so that universal approximation is achieved. We find that a separation condition is the key condition.

Theorem 2: Let $\mathcal{G}$ be the set of all time-invariant causal approximately-finite-memory maps G from S to the set of real-valued functions on $\mathcal{Z}_+$ such that each associated functional F_n is continuous. Then (i) of Theorem 1 holds for each $G \in \mathcal{G}$ if and only if for every n the $q(h,n,\cdot)$ separate the points of S_n. [i.e., if and only if $q(h,n,u_1) \neq q(h,n,u_2)$ for some $h \in H$ whenever $u_1, u_2 \in S_n$ and $u_1 \neq u_2$].

The proof is omitted in this version of the paper.

2.3 Comments

For a given G the condition of Theorem 1 that A.1 be met can be much less restrictive than the separation condition of Theorem 2. For instance, suppose that G is the zero map [i.e., suppose that $(Gs)(n) = 0$ for all s and n]. Then each E_n is empty and A.1 imposes no restrictions on H, as one would expect.

Using the closure of H under the memory-limiting operation, a check of the omitted proof of Theorem 1 shows that Statement (i) is equivalent to: For each $\epsilon > 0$, there are a positive integer k, elements $h_1, \ldots, h_k$ of H, and an $N \in D_k$ such that

$$|(Gs)(n) - N[(Ms)(n)]| < \epsilon, \quad n \in \mathcal{Z}_+$$

for all $s \in S$, where $(Ms)(n) = [(h_1 s)(n), \ldots, (h_k s)(n)]$.

The conditions of Statement (ii) of Theorem 1 can be expressed in other ways. For example, using the causality of G, it is not difficult to check that the condition that each F_n is continuous is equivalent to the condition that each functional $G(\cdot)(n)$ is continuous with respect to the metric on S given by $\rho(x,y) = \sup\{\, \|x(j) - y(j)\| \,:\, j \in \mathcal{Z}_+\}$. As another example, the statement that each each F_n is continuous contains redundancy in the sense that, by the time-invariance of G, if n_1 and n_2 are elements of $\mathcal{Z}_+$ such that $n_2 > n_1$ then the continuity of F_{n_2} implies the continuity of F_{n_1}. Similarly, and concerning A.1 (and this time by the time-invariance of G and the elements of H), if the following holds for $n = n_2$ and $n_2 > n_1$ then it holds for $n = n_1$.

> For each $(u_1, u_2) \in E_n$ there is an $h \in H$ such that $q(h,n,u_1) \neq q(h,n,u_2)$,

in which

$$E_n = \{(u_1, u_2) \in S_n \times S_n : F_n u_1 \neq F_n u_2\}.$$

In applications it is often possible to choose H so that its elements are linear and A.1 is met. However, for a given degree of approximation (i.e., for a given ϵ) a much lower overall degree of complexity of the approximation structure can sometimes result if H is allowed to contain relatively simple elements that are not linear. As an example of how H can be chosen, let H_{00} stand for the subset of H_0 (of Section 2.1) containing only the members for which the functions ϕ are strictly monotone increasing. Given n as well as u_1 and u_2 in S_n such that $u_1 \neq u_2$, choose $h \in H_{00}$, so that it is given by (1) with $a(n-j) = (u_1 - u_2)(j)^{\mathrm{tr}}$ for $j = 1, \ldots, n$. Then

$$\sum_{j=0}^{n} (u_1 - u_2)(j) a(n-j) \neq 0$$

and so, by the strict monotonicity of the ϕ's, $q(n,h,u_1) \neq q(n,h,u_2)$. This shows that A.1 is always met for for $H = H_{00}$.

Now assume that $H = H_{00}$ with the ϕ's linear, and assume also that the D_k are linear-ring maps. Let Q denote the approximating map of Theorem 1 given by $(Qs)(n) = N[(MW_{n,\alpha}s)(n)]$, and let $h_1', \ldots, h_k'$ be elements of H such that $(h_j' s)(n) = (h_j W_{n,\alpha} s)(n)$ Consider the $d = 1$ case. By writing products of sums as iterated sums, it is not difficult to see that $(Qs)(n)$ has the form

$$\sum_{j=1}^{p} \sum_{n_j=0}^{n} \cdots \sum_{n_1=0}^{n} k_j(n-n_1, \ldots,$$

$$n - n_j) s(n_1) \cdots s(n_j) \tag{2}$$

in which p is a positive integer, and each $k_j(n-n_1, \ldots, n-n_j)$ is a finite linear combination of products of the form $a'_{i_j(1)}(n-n_1) a'_{i_j(2)}(n-n_2) \cdots a'_{i_j(j)}(n-n_j)$, where the indices $i_j(1), \ldots, i_j(j)$ are drawn from $\{1, \ldots, k\}$ and each a_i' is the kernel (i.e., discrete impulse response) associated with h_i'. Note that each k_j vanishes whenever one or more of its arguments exceed α.

Expression (2) is a doubly-finite Volterra series approximant for G in the sense that p is finite and each k_j vanishes if at least one of its arguments is sufficiently large in the sense indicated (which of course has the interpretation that the approximants have finite memory).[6] Let $\mathcal{M}$ stand for the set of all causal time-invariant maps from S to the real-valued functions defined on $\mathcal{Z}_+$, and let $\mathcal{V}$ denote the family of all members of $\mathcal{M}$ that have a representation of the form (2) for some p and some $k_1, \ldots, k_p$ with each $k_j(n_1, \ldots, n_j) = 0$ when one or more of the $n_1, \ldots, n_j$

[6] For early studies and other references concerning Volterra-series representations, see [16] and [3].

exceeds some number η. We have just seen that sufficient conditions for a member of $\mathcal{M}$ to be uniformly approximated arbitrarily well by an element of $\mathcal{V}$ are that the member possesses approximately-finite memory and that its corresponding F_n functionals are continuous. In the full-length version of the paper we show that these conditions are in fact necessary.

The methods we have used can be employed to obtain corresponding results for other types of input-output maps. A particularly important case is the one in which inputs and outputs are real-valued functions of a finite number of integer-valued variables. This case is of interest in connection with, for example, image processing. We consider this case in the full-length version of this paper where a large class of myopic maps are the focus of attention. As an application, we give a criteria in this multivariable setting for the existence of arbitrarily good doubly-finite Volterra series approximations.

REFERENCES

[1] I. W. Sandberg, "Structure theorems for nonlinear systems," *Multidimensional Systems and Signal Processing*, vol. 2, no. 3, pp. 267–286, 1991. (See also the Errata in vol. 3, no. 1, p. 101, 1992.) A conference version of the paper appears in *Integral Methods in Science and Engineering-90* (Proceedings of the International Conference on Integral Methods in Science and Engineering, Arlington, Texas, May 15-18, 1990, ed. A. H. Haji-Sheikh), New York: Hemisphere Publishing, pp. 92–110, 1991.

[2] ____________, "Uniform approximation and the circle criterion," *IEEE Transactions on Automatic Control*, vol. 38, no. 10, pp. 1450–1458, October 1993.

[3] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, John Wiley, New York, 1980.

[4] I. W. Sandberg and L. Xu, "Network approximation of input-output maps and functionals," *Journal of Circuits, Systems, and Signal Processing*, vol. 15, no. 6, pp. 711–725, 1996.

[5] T. Chen and H. Chen, "Approximations of continuous functionals by neural networks with application to dynamical systems," *IEEE Transactions on Neural Networks*, vol. 4, no. 6, pp. 910–918, November 1993.

[6] S. Boyd and L. O. Chua, "Fading memory and the problem of approximating nonlinear operators with volterra series," *IEEE Transactions on Circuits and Systems*, vol. CAS-32, no. 11, pp. 1150–1161, November 1985.

[7] I. W. Sandberg and L. Xu, "Uniform approximation of multidimensional myopic maps," *IEEE Transactions on Circuits and Systems -I*, May 1997 (to appear).

[8] I. W. Sandberg and L. Xu, "Approximation of myopic systems whose inputs need not be continuous," *Multidimensional Systems and Signal Processing*, 1997 (to appear).

[9] I. W. Sandberg, "The mathematical foundations of associated expansions for mildly nonlinear systems," *IEEE Transactions on Circuits and Systems*, vol. CAS-30, no. 7, pp. 441–455, July 1983.

[10] B. D. Coleman and V. J. Mizel, "Norms and semi-groups in the theory of fading memory," *Arch. Rational Mech. Anal.*, Vol. 23, pp. 87–123, 1966.

[11] B. D. Coleman and V. J. Mizel, "On the general theory of fading memory," *Arch. Rational Mech. Anal.*, vol. 6, pp. 180–231, 1968.

[12] I. W. Sandberg, "Approximately-finite memory and input-output maps," *IEEE Transactions on Circuits and Systems*, vol. 39, no. 7, pp. 549-556, July 1992.

[13] J. S. Shamma and R. Zhao, "Fading-memory feedback systems and robust stability," *Automatica*, vol. 29, no.1, pp. 191-200, 1993.

[14] B. W. Stiles, I. W. Sandberg, and J. Ghosh, "Complete Memory Structures for Approximating Nonlinear Discrete-Time Mappings," to appear in *IEEE Transactions on Neural Networks*.

[15] I. W. Sandberg, "Separation conditions and criteria for uniform approximation of input-output maps," submitted.

[16] W. J. Rugh, *Nonlinear System Theory: The Volterra/Wiener Approach*, Johns Hopkins University Press, Baltimore, 1981.