TENSOR PRODUCT NEURAL NETWORKS AND APPROXIMATION OF DYNAMICAL SYSTEMS

Ajit T. Dingankar[1] Irwin W. Sandberg[2]

[1] IBM Corporation
Austin, TX 78758, U.S.A.
ajit@austin.ibm.com
[2] The University of Texas at Austin
Austin, TX 78712, U.S.A.
sandberg@uts.cc.utexas.edu

ABSTRACT

We consider the problem of approximating any member of a large class of input-output operators of nonlinear dynamical systems. The systems need not be shift invariant, and the system inputs need not be continuous. We introduce a family of "tensor product" dynamical neural networks, and show that a certain continuity condition is necessary and sufficient for the existence of arbitrarily good approximations using this family.

1. INTRODUCTION

In this paper we consider the problem of approximating input-output operators G of (not necessarily shift-invariant) nonlinear dynamical systems that take a subset C of a normed linear space X into another normed linear space E. Suppose that E is complete and assume initially that it has a basis $\{e_1, e_2, \ldots\}$, so that every $e \in E$ can be represented as $e = \sum_{j=1}^{\infty} g_j(e)e_j$ where the g_j are unique functionals. Then the g_j are continuous [1, p. 135], and we have

$$G(x) = \sum_{j=1}^{\infty} g_j[G(x)]e_j, \quad x \in C. \tag{1}$$

The right side of (1) is of the form

$$\sum_{j=1}^{\infty} a_j(x)v_j$$

in which the a_j are continuous functionals and the v_j are elements of the output space E. A natural question that arises is whether arbitrarily good uniform approximations of G of the form

$$\sum_{j=1}^{\ell} a_j(\cdot)v_j \tag{2}$$

can be obtained with ℓ finite. By "uniform" we mean uniform in the system inputs.

One of the principal results in [2] is a result in a certain setting that provides a criterion under which finite sums of the form

$$\sum_{i=1}^{\ell}\sum_{j=1}^{k(i)} c_{ij}u_{ij}[y_{ij}(\cdot)]v_i \tag{3}$$

achieve such an approximation. In (3) the c_{ij} are real constants, the u_{ij} are certain continuous real-valued functions of the reals, and the y_{ij} are continuous real functionals that can be taken to be linear. Of course, (3) is of the form (2). In [2] attention is restricted to the case of inputs that are continuous functions. This is an important case, but as is well known it is often desirable to assume that inputs may possess discontinuities. In this paper attention is focused on one setting in which input discontinuities are allowed. The results in this paper complement those in [2]; neither sets of results is stronger than the other. In particular, the assumptions in one paper concerning a certain weighting function w are of a different character than in the other.

2. PRELIMINARIES

Throughout the paper m and n denote arbitrary positive integers, $p \geq 1$ is a real number and D stands for the domain $\mathbb{R}_+^m$ of our input functions. The Euclidean norm on $\mathbb{R}^m$ is denoted by $|\cdot|$. We consider nonlinear dynamical systems with inputs x belonging to a subset of the set of all $\mathbb{R}^n$-valued (Lebesgue) measurable functions defined on D. We assume that the corresponding outputs belong to a real Banach space E with norm $||\cdot||_E$, but now it is not assumed that E has a basis. The case in which E possesses a basis leads to a much simpler problem that can be addressed using the characterization in [1, p.136] of compact subsets of a Banach space with a basis (see the Appendix). An important example of an E that does not have a basis is the space of real-valued bounded measurable functions defined on D, with the usual norm.

In the development of our results a central role is played by certain weighted norm spaces.[1] To describe these spaces, let $w: D \to (0,1]$ be a measurable function with $\int_D w(\alpha)d\alpha < \infty$. Let X_w be the set of all $\mathbb{R}^n$-valued measurable functions x defined on D for which

$$||x||_w := \left\{ \int_D ||x(\alpha)||^p w(\alpha)d\alpha \right\}^{1/p} < \infty,$$

where $||\cdot||$ is the Euclidean norm on $\mathbb{R}^n$. We use X to denote $L_p(D)$, the space of $\mathbb{R}^n$-valued p-integrable functions

[1] Related ideas (in a general sense) involving weighted norms on function spaces can be found in, for example, proofs of the boundedness of solutions of integral equations [3], and in a large number of other publications (see, for instance, [4], [5]).

on D, with the usual norm

$$||x|| := \left\{ \int_D ||x(\alpha)||^p d\alpha \right\}^{1/p} .$$

We view X_w as a normed linear space with norm $||\cdot||_w$. Later we will use the fact that X_w is a Banach space (see [6, Appendix B.1.2] for a proof).[2]

For each $a \in I\!R_+$ let c_a denote $[0,a]^m$, and for each $k \in I\!N$ let P_k be a partition of c_k in the sense that $P_k = \{d_1^k, \ldots, d_{n(k)}^k\}$ where:

1. Every d_i^k is compact with $\mu(d_i^k) > 0$, where μ is the Lebesgue measure on D.
2. $\mu(d_i^k \cap d_j^k) = 0,\ i \neq j$.
3. $c_k = \bigcup_{i=1}^{n(k)} d_i^k$.

Let $r: D \to I\!R_+$ be a measurable function such that r is bounded on bounded subsets of D and

$$\int_D r(\alpha)^p w(\alpha) d\alpha < \infty, \tag{4}$$

and let λ be a function from $I\!N$ to $I\!R_+$.

Let B be any closed nonempty subset of X_w satisfying the following condition: Each $b \in B$ is equivalent[3] to a function $f: D \to I\!R^n$ such that

1. $||f(\alpha)|| \leq r(\alpha)$ for $\alpha \in D$.
2. For every $k \in I\!N$ and $d_i^k \in P_k$,
$$||f(\alpha) - f(\beta)|| \leq \lambda(k)|\alpha - \beta|, \quad \alpha, \beta \in \mathrm{int}(d_i^k),$$
where $\mathrm{int}(d_i^k)$ denotes the interior of d_i^k.

Of course r can be taken to be any bounded function. It can be shown that the set of all $x \in X_w$ that satisfy the above condition with r any constant function is a closed subset of X_w. Thus B can be taken to be any such set of piecewise continuous functions.

We denote by X_w^* the set of bounded linear functionals on X_w (i.e., the set of bounded linear maps from X_w to the reals $I\!R$). Let Y be any set of continuous maps from X_w to $I\!R$ that is dense in X_w^* on B, in the sense that for each $\phi \in X_w^*$ and any $\rho > 0$ there is a $y \in Y$ such that $|\phi(x) - y(x)| < \rho$, $x \in B$.

Also, let U be any set of continuous maps $u : I\!R \to I\!R$ such that given $\sigma > 0$ and any bounded interval $(\beta_1, \beta_2) \subset I\!R$ there exists a finite number of elements $u_1, \ldots, u_q$ of U for which $|\exp(\beta) - \sum_j u_j(\beta)| < \sigma$ for $\beta \in (\beta_1, \beta_2)$.[4]

Let $\mathcal{T}$ denote the set of maps from B to E of the form

$$\sum_{i=1}^{\ell} \sum_{j=1}^{k(i)} c_{ij} u_{ij}[y_{ij}(\cdot)] v_i \tag{5}$$

where $\ell \in I\!N$, and $k(i) \in I\!N$, $v_i \in E$ for $1 \leq i \leq \ell$, and where $c_{ij} \in I\!R$, $u_{ij} \in U$, and $y_{ij} \in Y$ for $1 \leq i \leq \ell,\ 1 \leq j \leq k(i)$. Since a sum of the form $\sum_{i=1}^{\ell} \phi_i(\cdot) v_i$ is an element of the so-called tensor product (e.g., see [8]), a general element M of $\mathcal{T}$ can be realized by what may naturally be called the *tensor product neural network* shown in Figure 1.

3. CHARACTERIZATION OF CONTINUITY OF INPUT-OUTPUT MAPS

Let $\mathcal{G}$ denote the linear space of all maps $G: B \to E$ such that $\sup_{x \in B} ||Gx||_E < \infty$. The set $\mathcal{G}$ is a large set in that (as will become clear) it contains all continuous maps from B to E. It is easy to give examples of $G \in \mathcal{G}$ that are not continuous [2].

3.1. The approximation theorem

Theorem 1 *Let $G \in \mathcal{G}$ be given. Then the following two conditions are equivalent.*

1. *G is continuous on B with respect to the norm $||\cdot||_w$.*
2. *For any $\epsilon > 0$, there exists an $M \in \mathcal{T}$ such that $||Gx - Mx||_E < \epsilon,\ x \in B$.*

Proof: (1) $\Rightarrow$ (2). We first prove that B is relatively compact in X_w. We do this by showing that for any positive ϵ, B has a finite ϵ-net (recall that X_w is complete). Let $\epsilon > 0$ be given. Using (4) choose $k \in I\!N$ such that

$$\int_{D \setminus c_k} r(\alpha)^p w(\alpha) d\alpha < (\epsilon/2)^p. \tag{6}$$

Define $T_k: X_w \to X_w$ by

$$(T_k s)(\alpha) = \begin{cases} s(\alpha) & \text{if } \alpha \in c_k, \\ \theta & \text{otherwise,} \end{cases}$$

where θ denotes the zero element of $I\!R^n$. We will prove that $T_k(B)$ is relatively compact in X_w. It suffices to prove that $T_k(B)|_{c_k}$ is relatively compact in $L_p(c_k)$ since $\eta: L_p(c_k) \to X_w$ given by

$$(\eta f)(\alpha) = \begin{cases} f(\alpha) & \text{if } \alpha \in c_k, \\ \theta & \text{otherwise,} \end{cases}$$

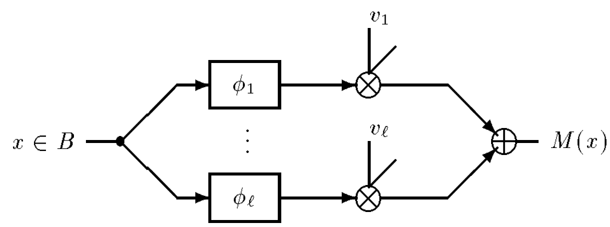
is continuous. Let S_i, $i = 1, \ldots, n(k)$, denote the set of continuous extensions of the elements of $T_k(B)|_{\mathrm{int}(d_i^k)}$ to d_i^k. By the generalized Ascoli-Arzelà theorem [9, p. 107, Problem 5], each S_i is relatively compact in $\mathcal{C}(d_i^k)$ (the space of continuous $I\!R^n$-valued functions defined on d_i^k). Hence each S_i is also relatively compact in $L_p(d_i^k)$. Since $S_i = T_k(B)|_{d_i^k}$ in the sense of the norm in $L_p(d_i^k)$, repeated application of the following proposition shows that $T_k(B)|_{c_k}$ is relatively compact in $L_p(c_k)$.

Proposition 1 *Let d_1, d_2 be compact subsets of D such that $\mu(d_1 \cap d_2) = 0$, and let A be a subset of $L_p(d_1 \cup d_2)$. If $A|_{d_i}$ is relatively compact in $L_p(d_i)$, $i = 1, 2$, then A is relatively compact in $L_p(d_1 \cup d_2)$.*

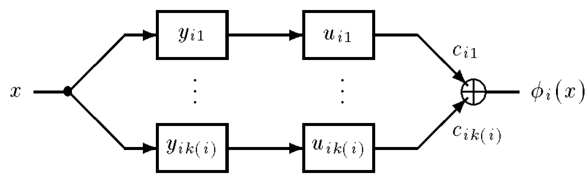
[2] A similar result is given in [7].

[3] The elements of X_w are equivalence classes, with equivalence meaning almost everywhere equality with respect to Lebesgue measure.

[4] Of course we can take U to be the set whose only element is $\exp(\cdot)$, or the set $\{u : u(\beta) = (\beta)^n/n!,\ n \in \{0, 1, \ldots\}\}$.



(a) Network representation of the tensor product $M(x) = \sum_{i=1}^{\ell} \phi_i(x) v_i$.



(b) Network representation of the i^{th} input functional $\phi_i(x) = \sum_{j=1}^{k(i)} c_{ij} u_{ij}[y_{ij}(x)]$.

Figure 1.

Proof of the Proposition: Choose $\delta/\sqrt[p]{2}$-nets $\{s_i\}$, $\{t_j\}$ for $A|_{d_1}$ and $A|_{d_2}$ in $L_p(d_1)$ and $L_p(d_2)$ respectively. Let $\{r_{ij}\} \subset L_p(d_1 \cup d_2)$ be given by

$$r_{ij}(\alpha) = \begin{cases} s_i(\alpha) & \text{if } \alpha \in \text{int}(d_1), \\ t_j(\alpha) & \text{if } \alpha \in \text{int}(d_2), \\ \theta & \text{otherwise.} \end{cases}$$

where $\text{int}(d_1)$ denotes the interior of d_1, and similarly for $\text{int}(d_2)$. Using the identity

$$\int_{d_1 \cup d_2} ||a(\alpha)||^p d\alpha = ||a|_{d_1}||^p + ||a|_{d_2}||^p,$$

it is easy to see that $\{r_{ij}\}$ is an δ-net for A in $L_p(d_1 \cup d_2)$.

Returning to the proof of the relative compactness of B, there exists in X_w a finite $\epsilon/2$-net $\{s_1, \ldots, s_j\}$ for $T_k(B)$ since it is relatively compact in X_w. Given $s \in B$, choose $i \in \{1, \ldots, j\}$ such that

$$||T_k s - s_i||_w < \epsilon/2. \qquad (7)$$

Note that

$$\begin{aligned} ||s - T_k s||_w{}^p &= \int_{D \setminus C_k} ||s(\alpha)||^p w(\alpha) d\alpha \\ &\leq \int_{D \setminus C_k} r(\alpha)^p w(\alpha) d\alpha \qquad (8) \end{aligned}$$

By the triangle inequality and (8), (6), and (7) we have $||s - s_i||_w < \epsilon$, showing the existence of a finite ϵ-net for B. Thus B, which is assumed to be closed, is relatively compact, and thus compact.[5]

[5] This shows that as mentioned earlier, $\mathcal{G}$ contains all continuous maps from B to E.

Recall that we are assuming that G is continuous in the norm $||\cdot||_w$. We will use the following lemma proved in [2].

Lemma 1 ([2]) *Let H be a compact metric space, and let $\mathcal{P}$ be the set of all maps of the form $\sum_{i=1}^{\ell} \phi_i(\cdot) v_i$ where the ϕ_i are continuous functionals on H, $v_i \in E$, and $\ell \in I\!N$. Then a map $G\colon H \to E$ is continuous if and only if for any $\epsilon > 0$, there exists an $M \in \mathcal{P}$ such that $||Gx - Mx||_E < \epsilon$, $x \in H$.*

Continuing with the proof of the theorem, let $\epsilon > 0$ be given. By Lemma 1, there exist ℓ together with continuous functionals a_i and elements v_i of E for $i = 1, \ldots, \ell$ such that

$$||Gx - \textstyle\sum_{i=1}^{\ell} a_i(x) v_i||_E < \epsilon/2, \quad x \in B.$$

The case in which $||v_i||_E = 0$ for all i is trivial. So assume otherwise. By Theorem 1 of [10] there exist $k(i) \in I\!N$ as well as $c_{ij} \in I\!R$, $u_{ij} \in U$, and $y_{ij} \in Y$, $1 \leq j \leq k(i)$ with the property that

$$\left| a_i(x) - \sum_{j=1}^{k(i)} c_{ij} u_{ij}[y_{ij}(x)] \right| < \epsilon/(2\ell\gamma), \quad x \in B,\ 1 \leq i \leq \ell,$$

where $\gamma = \max_{1 \leq i \leq \ell} ||v_i||_E$. By combining the above estimates using the triangle inequality, we see that **(1)** $\Rightarrow$ **(2)**.

(2) $\Rightarrow$ **(1)**. This follows from the compactness of B and Lemma 1.

4. CONCLUSION

We have introduced a family $\mathcal{T}$ of maps of "tensor product" dynamical neural network structures and, given the input-output map G of any member of a certain large class of (not necessarily shift-invariant) nonlinear dynamical systems, we have shown that continuity of G with respect to the norm $||\cdot||_w$ is necessary and sufficient for $\mathcal{T}$ to contain an arbitrarily good approximation to G. Inputs and outputs are not restricted to be defined on finite intervals, nor need they be functions of only one variable. In this paper we have not considered the important problem of actually determining the elements of the approximating structures.

Related results can be found in [6] for cases in which D is replaced with $\mathbf{Z}_+^m$, $I\!R^m$ or $\mathbf{Z}^m$.

APPENDIX

The case in which E has a basis

Here we suppose that E is a Banach space with a basis $\{e_1, e_2, \ldots\}$. Let $G: C \to E$ be a continuous input-output map of a dynamical system, where C is now a compact subset of a normed linear space X. Using the characterization in [1, p. 136] of compact subsets of E, given any positive ϵ there exists an $\ell \in I\!N$ such that

$$||G(x) - \sum_{j=1}^{\ell} g_j[G(x)]e_j||_E < \epsilon, \quad x \in C.$$

This yields an output approximation of the form (2), with the approximation uniform in the system inputs x.

REFERENCES

[1] L. A. Liusternik and V. J. Sobolev, *Elements of Functional Analysis.* New York: Frederick Ungar Publishing Co., 1961.

[2] A. T. Dingankar and I. W. Sandberg, "Network Approximation of Dynamical Systems," in *Proceedings of the International Symposium on Nonlinear Theory and its Applications (NOLTA'95),* Vol. 2, (Las Vegas, Nevada), pp. 757–762, December 10–14 1995.

[3] I. W. Sandberg, "On the boundedness of solutions of nonlinear integral equations," *Bell System Technical Journal*, vol. 44, pp. 439–453, March 1965.

[4] S. Boyd, *Volterra Series: Engineering Fundamentals.* PhD thesis, The University of California at Berkeley, Berkeley, California, 1985.

[5] B. D. Coleman and V. J. Mizel, "On the general theory of fading memory," *Arch. Rational Mech. Anal.*, vol. 29, pp. 18–31, 1968.

[6] A. T. Dingankar, *On Applications of Approximation Theory to Identification, Control and Classification.* PhD thesis, The University of Texas at Austin, Austin, Texas, 1995.

[7] I. W. Sandberg and L. Xu, to appear.

[8] E. W. Cheney, *Multivariate Approximation Theory: Selected Topics*, vol. 51 of *The Regional Conference Series in Applied Mathematics.* Philadelphia, Pennsylvania: SIAM, 1986.

[9] A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis.* New York: Dover, 1975.

[10] I. W. Sandberg, "General structures for classification," *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 41, pp. 372–376, May 1994.