Network Approximation of Dynamical Systems

Ajit T. Dingankar* and Irwin W. Sandberg†

∗ IBM Corporation, Austin, TX 78758, U.S.A., ajit@austin.ibm.com
† The University of Texas at 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 time-varying nonlinear dynamical systems. 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.

I. Introduction

In this paper we consider the problem of approximating input-output operators G of time-varying 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 of this paper 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. Of course, (3) is of the form (2). We know of no closely related criterion in the literature. However, sufficient conditions for a specialized approximation of the general type we consider are given in [2].

II. Preliminaries

Throughout the paper m and n are arbitrary positive integers, and D denotes the domain $\mathbb{R}_+^m$ of our input functions. The Euclidean norm on D 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 continuous functions defined on D. We assume that the outputs belong to a real Banach space E with norm $||\cdot||_E$, but now it is not assumed that E has a basis.[1] An important example of an E of interest to us that does not have a basis is the set of real-valued bounded Lebesgue measurable functions defined on D.

In the development of our results a central role is played by certain weighted norm spaces.[2] To describe these spaces, let $w: D \to (0, 1]$ be a continuous map with $\lim_{|\alpha|\to\infty} w(\alpha) = 0$. Let X_w be the set of all $\mathbb{R}^n$-valued continuous functions x defined on D for which

$$||x||_w := \sup_{\alpha \in D} ||x(\alpha)w(\alpha)|| < \infty$$

[1] The case in which E does have 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.

[2] 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], [4, pp. 880–1], and in a large number of other publications (see, for instance, [5], [6], [7], [8]).

where $||\cdot||$ is the Euclidean norm on $\mathbb{R}^n$. We use X to denote the space of $\mathbb{R}^n$-valued bounded continuous functions on D, with the usual norm $\sup_{\alpha \in D} ||\cdot||$. 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 Appendix A).

For each $a \in \mathbb{R}_+$, let c_a denote $[0,a]^m$. Let λ be a function from $\mathbb{R}_+$ to $\mathbb{R}_+$, and let $r: D \to \mathbb{R}_+$ be a continuous function that has restricted growth rate at infinity and is nondecreasing, in the sense that

$$\lim_{|\alpha| \to \infty} r(\alpha) w(\alpha) = 0 \tag{4}$$

and

$$\max_{1 \le i \le m} \alpha_i \le \max_{1 \le i \le m} \beta_i \;\Rightarrow\; r(\alpha) \le r(\beta), \quad \alpha, \beta \in D \tag{5}$$

where α_i and β_i denote the i^{th} components of α and β, respectively.

Let B be a closed nonempty subset of X_w such that:

1. $||x(\alpha)|| \le r(\alpha)$ for $x \in B$ and $\alpha \in D$.
2. $||x(\alpha) - x(\beta)|| \le \lambda(a)|\alpha - \beta|$ for $a \in \mathbb{R}_+$, $x \in B$, and $\alpha, \beta \in c_a$.

Of course r can be taken to be any bounded function satisfying (5). It is not difficult to check that the set of all $x \in X_w$ that satisfy the above two conditions with r and λ any two constant functions is a closed subset of X_w.[3] Thus B can be taken to be any such set of "uniformly bounded uniformly Lipschitz" 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 $\mathbb{R}$). Let Y be any set of continuous maps from X_w to $\mathbb{R}$ 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 $u : \mathbb{R} \to \mathbb{R}$ such that given $\sigma > 0$ and any bounded interval $(\beta_1, \beta_2) \subset \mathbb{R}$ 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

$$\sum_{i=1}^{\ell} \sum_{j=1}^{k(i)} c_{ij}\, u_{ij}[y_{ij}(\cdot)] v_i \tag{6}$$

where $\ell \in \mathbb{N}$ (the positive integers), and $k(i) \in \mathbb{N}$, $v_i \in E$ for $1 \le i \le \ell$, and where $c_{ij} \in \mathbb{R}$, $u_{ij} \in U$, and $y_{ij} \in Y$ for $1 \le i \le \ell$, $1 \le j \le k(i)$. 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 [9]), 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.

III. 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.

A. 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) ⇒ (2). We first prove that B is relatively compact in the norm $||\cdot||_w$. We will use the following result.[5]

Lemma 1 ([11]) *Let S be a subset of a complete metric space A with metric ρ, and let $T_1, T_2, \ldots$ be maps of A into itself such that*

(i) $T_k(S)$ is a relatively compact subset of A for each k, and

(ii) $\rho(s, T_k s) \to 0$ as $k \to \infty$ uniformly for $s \in S$.

Then S is a relatively compact subset of A.

Continuing with the proof of the theorem, note that c_k is closed and convex for every positive integer k. Hence [12, p. 98] for any $x \in D$ and any k, there exists a unique $\hat{x} \in c_k$ such that $|x - \hat{x}| = \inf_{y \in c_k} |x - y|$. For $k = 1, 2, \ldots$ define $T_k : X_w \to X_w$ by

$$(T_k s)(x) = s(\hat{x}). \tag{7}$$

Referring to the lemma, set $A = X_w$, set $S = B$, and take ρ to be the usual metric induced by the norm $||\cdot||_w$.

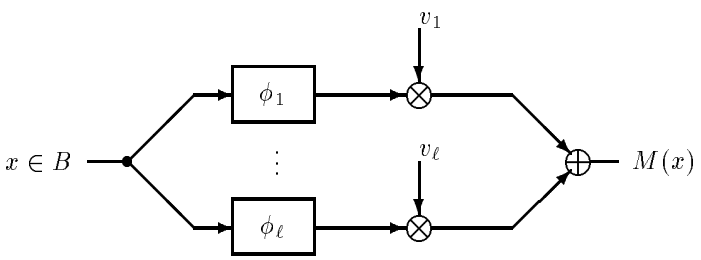
Consider Hypothesis (i). Let k be an arbitrary positive integer. Note that X is a subset of X_w and that the map $X \to X_w$ given by $x \mapsto x$ is continuous. Hence, to show that $T_k(B)$ is relatively compact it suffices to prove that it is relatively compact in X.

Let $\mathcal{C}(c_k)$ denote the space of continuous $\mathbb{R}^n$-valued functions on c_k, with the usual norm. Using the two conditions in the definition of B together with an extension [13, p. 76] of the classical Ascoli-Arzelà theorem, it follows directly that the restriction $T_k(B)|_{c_k}$ of

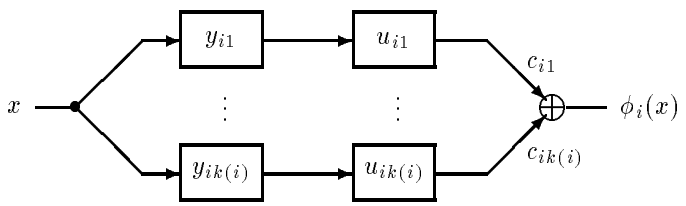
[3] To show this it is helpful to notice that convergence with respect to the norm in X_w implies uniform convergence on any c_a.

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

[5] In [10] a similar application is made of Lemma 1.



(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:

$T_k(B)$ to c_k is relatively compact in $\mathcal{C}(c_k)$. Consider the map Q from $\mathcal{C}(c_k)$ to X given by $Qf = g$, where $g(\alpha) = f(\hat{\alpha})$, $\alpha \in D$. Since Q is continuous, $T_k(B)$ is relatively compact in X.

Now consider Hypothesis (ii). Note that here

$$\begin{aligned}\rho(s, T_k s) &= \sup_{\alpha \in D \setminus c_k} ||[s(\alpha) - s(\hat{\alpha})]w(\alpha)|| \\ &\leq \sup_{\alpha \in D \setminus c_k} r(\alpha)w(\alpha) + \\ &\qquad \sup_{\alpha \in D \setminus c_k} \left[\max_{\beta \in c_k} r(\beta)\right] w(\alpha).\end{aligned}$$

By (4) the first term tends to 0 as $k \to \infty$. By (5) we have $\max_{\beta \in c_k} r(\beta) \leq r(\alpha)$ for $\alpha \in D \setminus c_k$ and any k. Hence $\rho(s, T_k s) \to 0$ uniformly for $s \in B$, as required. Thus B, which is assumed to be closed, is relatively compact, and thus compact.[6]

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

Lemma 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 $\ell < \infty$. Then a map $G: H \to E$ is continuous if and only if for any $\epsilon > 0$, there exists an $M \in \mathcal{P}$ such that $||Gx - Mx||_E < \epsilon$, $x \in H$.*

Proof of the Lemma: (**Necessity**) Assume that G is continuous. Let the space of all continuous functionals on H be denoted by $\mathcal{C}(H)$, and let the set of continuous maps from H to E be denoted by $\mathcal{C}(H, E)$. The so-called *tensor product* $\mathcal{C}(H) \otimes E$ of $\mathcal{C}(H)$ and E is the linear manifold of $\mathcal{C}(H, E)$ consisting of finite sums of the form $\sum_{i=1}^{\ell} a_i(\cdot)v_i$, where $a_i \in \mathcal{C}(H)$ and $v_i \in E$ for each i.

The following proposition establishes necessity.

Proposition: $\mathcal{C}(H) \otimes E$ is dense in $\mathcal{C}(H, E)$.

Proof of the Proposition: Let $\mathcal{M} = \mathcal{C}(H) \otimes E$, and for each $x \in H$ let $\mathcal{M}(x) = \{f(x): f \in \mathcal{M}\}$. Since $\mathcal{C}(H)$ contains the constant function taking the value unity, $\mathcal{M}(x) = E$ for every x. The proposition thus follows from [14, Corollary 1].

(**Sufficiency**) Suppose that for any $\epsilon > 0$, there exists an $M \in \mathcal{P}$ such that $||Gx - Mx||_E < \epsilon$, $x \in H$. Thus we can find a sequence $\{G_p\}$ of elements of $\mathcal{P}$ that is uniformly convergent to G. By the continuity of the G_p and a well-known result [15, p.121] concerning uniform convergence, it follows that G is continuous.

Continuing with the proof of the theorem, let $\epsilon > 0$ be given. By Lemma 2, there exist ℓ and 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 each i is trivial. So assume otherwise. By Theorem 1 of [16] (see Appendix B), there exist $k(i) \in \mathbb{N}$, $c_{ij} \in \mathbb{R}$, $u_{ij} \in U$, and $y_{ij} \in Y$, $1 \leq j \leq k(i)$ such that

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

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 2.

Comment. With regard to Condition 1 of Theorem 1, it is easy to give examples of G's in $\mathcal{G}$ that are not continuous. For instance, suppose that B, viewed as a metric space with its metric derived in the usual way from the norm $||\cdot||_w$, has a limit point x_0. Suppose also that G is defined on B so that Gx is an element of E with nonzero norm only for $x = x_0$. Then G is obviously not continuous.

IV. 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 time-varying 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 [17], [18] for cases in which D is replaced with $\mathbb{Z}_+^m$, or $\mathbb{R}^m$ or $\mathbb{Z}^m$, and for classes of continuous-time input functions that need not be continuous.

Appendix A

Proof of the Completeness of X_w

To see that X_w is complete,[7] let $\{x_p\}$ be a Cauchy sequence in X_w, and notice that for each positive ϵ there is a positive N such that

[6] See [5] for a related result for $m = 1$. Also, this shows that as mentioned earlier, $\mathcal{G}$ contains all continuous maps from B to E.

[7] Similar results are given in [18] and [10].

$$||wx_p - wx_q||_0 < \epsilon, \quad p, q > N.$$

where $||\cdot||_0$ is the norm on X of Section II. By the completeness of X, there is a $y \in X$ such that $\{wx_p\}$ converges to y in X. Thus, for an arbitrary positive ϵ there exists a positive N' such that

$$||y - wx_p||_0 < \epsilon, \quad p > N'.$$

Since the weighting function w never vanishes, the function $x = y/w$ can be seen to be the limit in X_w of $\{x_p\}$.

Appendix B

In Section III.reference is made to Theorem 1 of [16], a part of which is used in the proof of our theorem. Since the proof of the result in [16] contains some typos etc., for the sake of completeness we state below as a theorem the result we use and include a proof.

Let K be a nonempty compact subset of a real normed linear space Z, and let Z^* be the set of bounded linear functionals on Z (i.e., the set of bounded linear maps from Z to the reals $\mathbb{R}$). Let W be any set of continuous maps from Z to $\mathbb{R}$ that is dense in Z^* on K, in the sense that for each $\phi \in Z^*$ and any $\epsilon > 0$ there is a $y \in W$ such that $|\phi(x) - y(x)| < \epsilon$, $x \in K$. Let U be as described in Section II.

Theorem. Let f be a continuous map of K into $\mathbb{R}$. Then given $\epsilon > 0$ there are a positive integer k, real numbers $c_1, \ldots, c_k$, elements $u_1, \ldots, u_k$ of U, and elements $y_1, \ldots, y_k$ of W such that $|f(x) - \sum_j c_j\, u_j[y_j(x)]| < \epsilon$ for $x \in K$.

Proof

Let f be given, and notice that the set V of all functions $v : K \to \mathbb{R}$ of the form $v(x) = \sum_j a_j \exp[\phi_j(x)]$, in which the sum is finite and the a_j and the ϕ_j belong to $\mathbb{R}$ and Z^*, respectively, is an algebra under the natural definition of addition and multiplication. By a consequence [19, p.198] of the Hahn-Banach theorem, given distinct x_a and x_b in K there is a ϕ in Z^* such that $\exp[\phi(x_a)] \neq \exp[\phi(x_b)]$, showing that V separates the points of K. It is clear that $v(x) \neq 0$ for some $v \in V$ for each x. Thus, by a version of the Stone-Weierstrass Theorem [20, p.162], given $\epsilon > 0$ there are a positive integer p, real numbers $d_1, \ldots, d_p$, and elements $w_1, \ldots, w_p$ of Z^* such that

$$|f(x) - \sum_j d_j \exp[w_j(x)]| < \epsilon/3$$

for $x \in K$.[8]

[8] Here we view K as a metric space with the metric derived in the usual way from the norm in Z.

We may assume that $\sum_j |d_j| \neq 0$. Choose $\gamma > 0$ such that $\gamma \sum_j |d_j| < \epsilon/3$. Let $[a', b']$ be an interval in $\mathbb{R}$ that contains all of the sets $w_j(K)$, and let real a and b be such that $a < a', b > b'$. Select $\eta > 0$ such that $|\exp(\beta_1) - \exp(\beta_2)| < \gamma$ for $\beta_1, \beta_2 \in [a, b]$ with $|\beta_1 - \beta_2| < \eta$. With $\rho = \min(\eta, a' - a, b - b')$, choose $y_j \in W$ such that $|w_j(x) - y_j(x)| < \rho$, $x \in K$ for all j. This gives $|\exp[w_j(x)] - \exp[y_j(x)]| < \gamma$, $x \in K$ for each j (because we have $y_j(K) \in [a, b]$ and $|w_j(x) - y_j(x)| < \eta$ for each j and x), and thus

$$\begin{aligned}
&|f(x) - \sum_j d_j \exp[y_j(x)]| \\
&\quad \leq |f(x) - \sum_j d_j \exp[w_j(x)]| + \\
&\qquad |\sum_j d_j \exp[w_j(x)] - \sum_j d_j \exp[y_j(x)]| \\
&\quad \leq \epsilon/3 + \sum_j |d_j| \cdot |\exp[w_j(x)] - \exp[y_j(x)]| \\
&\quad \leq (2\epsilon)/3, \ x \in K.
\end{aligned}$$

Now let $[c, d] \subset \mathbb{R}$ be such that $y_j(K) \subset [c, d]$ for each j. Pick $u_1, \ldots, u_\ell \in U$ so that $|\exp(\beta) - \sum_i u_i(\beta)| \leq \gamma_1$, $\beta \in [c, d]$ where $\gamma_1 \sum_j |d_j| < \epsilon/3$. Then

$$\begin{aligned}
&|f(x) - \sum_j \sum_i d_j\, u_i[y_j(x)]| \\
&\quad \leq |f(x) - \sum_j d_j \exp[y_j(x)]| + \\
&\qquad |\sum_j d_j \exp[y_j(x)] - \sum_j \sum_i d_j\, u_i[y_j(x)]| \\
&\quad \leq (2\epsilon)/3 + \sum_j |d_j \exp[y_j(x)] - d_j \sum_i u_i[y_j(x)]| \\
&\quad \leq (2\epsilon)/3 + \sum_j |d_j| \cdot |\exp[y_j(x)] - \sum_i u_i[y_j(x)]| \\
&\quad \leq (2\epsilon)/3 + \gamma_1 \sum_j |d_j| < \epsilon.
\end{aligned}$$

Since $\sum_j \sum_i d_j\, u_i[y_j(x)]$ can be written in the form $\sum_j c_j\, u_j[y_j(x)]$, with the c_j, u_j,and y_j in $\mathbb{R}, U$, and W, respectively, we have proved the theorem.

References

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

[2] T. Chen and H. Chen, "Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks," *IEEE Transactions on Neural Networks*, vol. 6, pp. 904–910, July 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] I. W. Sandberg, "Some results on the theory of physical systems governed by nonlinear functional equations," *Bell System Technical Journal*, vol. 44, pp. 871–898, May–June 1965.

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

[6] B. D. Coleman and W. Noll, "An approximation theorem for functionals, with applications in continuum mechanics," *Arch. Rational Mech. Anal.*, vol. 6, pp. 355–370, 1960.

[7] B. D. Coleman and V. J. Mizel, "Norms and semigroups in the theory of fading memory," *Arch. Rational Mech. Anal.*, vol. 23, pp. 87–123, 1966.

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

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

[10] I. W. Sandberg and L. Xu, "(in preparation)," 1996.

[11] I. W. Sandberg, "Weighted norms and network approximation of functionals," in *Proceedings of the special sessions on The Mathematical Theory of Electrical Networks presented at the MTNS'96 Conference, College of Engineering Technical Report no. 717*, State University of New York at Stony Brook, (to appear) 1996.

[12] A. A. Goldstein, *Constructive Real Analysis*. Harper's Series in Modern Mathematics, New York: Harper & Row, 1967.

[13] A. Mukherjea and K. Pothoven, *Real and Functional Analysis*. New York: Plenum Press, 1978.

[14] R. C. Buck, "Approximation properties of vector valued functions," *Pacific Journal of Mathematics*, vol. 53, no. 1, pp. 85–94, 1974.

[15] W. A. Sutherland, *Introduction to Metric and Topological Spaces*. Oxford: Clarendon Press, 1975.

[16] 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.

[17] A. T. Dingankar and I. W. Sandberg, "Tensor Product Neural Networks and Approximation of Dynamical Systems," in *Proceedings of the International Symposium on Circuits and Systems*, Atlanta, Georgia, (Atlanta, Georgia), May 12–15 (to appear) 1996.

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

[19] G. Bachman and L. Narici, *Functional Analysis*. New York: Academic Press, Inc., 1966.

[20] W. Rudin, *Principles of Mathematical Analysis*. International Series in Pure and Applied Mathematics, New York: McGraw-Hill, Inc., Third ed., 1976.