Time-Delay Polynomial Networks and Rates of Approximation

IRWIN W. SANDBERG *

Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084, USA

We consider a large family of finite memory causal time-invariant maps G from an input set S to a set of $\mathbb{R}$-valued functions, with the members of both sets of functions defined on the nonnegative integers, and we give an upper bound on the error in approximating a G using a two-stage structure consisting of a tapped delay line and a static polynomial network N. This upper bound depends on the degree of the multivariable polynomial that characterizes N. Also given is a lower bound on the worst-case error in approximating a G using polynomials of a fixed maximum degree. These upper and lower bounds differ only by a multiplicative constant. We also give a corresponding result for the approximation of not-necessarily-causal input–output maps with inputs and outputs that may depend on more than one variable. This result is of interest, for example, in connection with image processing.

Keywords: Polynomial networks; Nonlinear networks; Time-delay networks; Rates of approximation; Nonlinear input-output maps

I. INTRODUCTION

Nonlinear systems often arise in engineering applications. As is well known, the analysis of these systems can be complicated. For this reason it is frequently desirable to have available simple approximate models that can be used for synthesis, analysis, and identification. For example, the design of an equalizer to reduce the effects of nonlinear distortion in a communication channel possessing nonlinearities begins with the problem of determining a suitable model for the equalizer

---

* Fax: 512-471-5532. E-mail: sandberg@uts.cc.utexas.edu.

whose parameters can then be determined either directly or adaptively. Similarly, the problem of identifying a nonlinear system (typically the objective is to obtain a model that can later be used for extrapolation or control) can be solved satisfactorily only if the class of assumed models is both analytically manageable and broad enough to provide a sufficiently good approximation to the system to be identified. This is the motivation for what follows.

Problems involving the approximation of nonlinear input–output maps arise often in engineering and science, and there has been a large amount of recent progress in this area. For results concerning the approximation of static (i.e., memoryless) maps see, for example, [1–3]. In [4] a study was begun of the network (i.e., neural network) approximation of dynamic maps. It was shown that large classes of approximately-finite-memory maps can be uniformly approximated arbitrarily well by the maps of certain nonlinear structures using, for example, sigmoidal nonlinearities or radial basis functions or polynomials. As an application [5], it was proved that discrete-time time-delay neural networks of the type shown in Fig. 1 can be used to uniformly approximate arbitrarily well the members of a large class of causal nonlinear dynamic discrete-time input–output maps. (A similar result in a different setting is given in [6].) In the figure, s is the system input, D denotes the unit delay operator, and N stands for a static nonlinear network that contains nonlinearities drawn from any of certain families of nonlinearities. For example, it suffices that the nonlinearities be polynomial nonlinearities. The proof in [5] is nonconstructive and gives
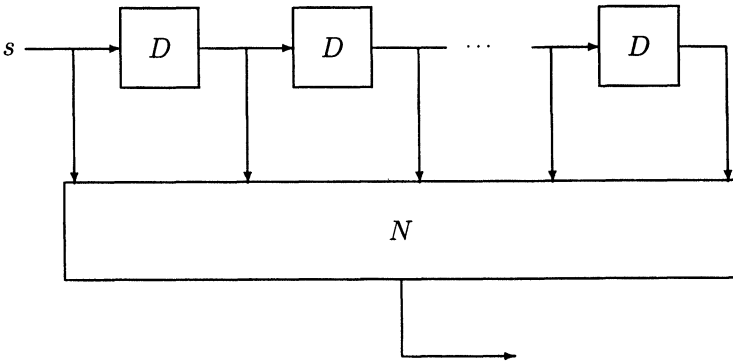

FIGURE 1   Time-delay nonlinear network.

no information concerning the needed complexity of the structure in
Fig. 1 so that a prescribed bound on the approximation error is achieved.
Here we give a result concerning the complexity problem for the case in
which N is a polynomial network. We explicitly consider only the
approximation of input–output maps that have finite memory. This is
a reasonable restriction because approximately-finite-memory maps
can be uniformly approximated arbitrarily well by finite memory maps
(see Section 2.1).

Our result is an application of the theory of polynomial approxima-
tion. A key proposition in this area is the well-known Weierstrass
approximation theorem which tells us that a continuous real-valued
function f defined on $[-1, 1]$ can be uniformly approximated arbi-
trarily well using polynomials. As the reader will recall, this means
that for any such f and any $\epsilon > 0$ there is a real polynomial p for
which $|f(x) - p(x)| \leq \epsilon$ whenever $|x| \leq 1$. The theorems of D. Jackson
(see, for instance [7]) extend Weierstrass' theorem by giving informa-
tion on the quality of approximation that can be achieved with a
polynomial of a given degree. In particular, Jackson showed that given
any f as described above and any positive integer n, there is a
polynomial p of degree n or less such that

$$|f(x) - p(x)| \leq A\omega\left(\frac{1}{n}\right), \quad x \in [-1, 1]. \tag{1}$$

Here A is a positive constant, and ω is the *modulus of continuity* of f
defined by $\omega(\eta) = \sup_{|x-y| \leq \eta} |f(x) - f(y)|$. Since $\omega(\eta) \to 0$ as $\eta \to 0$,
(1) provides information concerning the n's that suffice for a given ϵ.

Jackson's result has been extended to the case of functions f of an
arbitrary finite number of real variables (see, for example [7,8]). In
Section II we make use of the extension in [8] which also gives a lower
bound on the worst-case approximation of the members of a certain
family of continuous functions specified by a single dominating modulus
of continuity. (For a related study of approximation bounds for certain
functionals, see [9].)

In Section II we consider a large family $\mathcal{G}_\tau$ of finite memory causal
time-invariant maps G from an input set S to a set of $\mathbb{R}$-valued
functions, with the members of both sets of functions defined on the
nonnegative integers, and we give an upper bound on the error in

approximating a G using a structure of the type shown in Fig. 1. This upper bound depends on the degree of the multivariable polynomial that characterizes N in Fig. 1. We also give a lower bound on the worst-case error in approximating a G using polynomials of a fixed maximum degree. These upper and lower bounds differ only by a multiplicative constant. In Section III we give a corresponding result for the approximation of not-necessarily-causal input–output maps with inputs and outputs that may depend on more than one variable. This result is of interest, for example, in connection with image processing.

II. APPROXIMATION OF INPUT–OUTPUT MAPS

2.1. Preliminaries

Let d be a positive integer, and let C stand for $[-1, 1]^d$. With $\mathcal{Z}_+ =:$ $\{0, 1, \ldots\}$ and $\mathbb{R}$ the set of real numbers, let S denote the metric space of all maps s from $\mathcal{Z}_+$ to C with the metric given by $\rho(s_a, s_b) = \sup\{\|s_a(k) - s_b(k)\|: k \in \mathcal{Z}_+\}$ in which $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$. 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

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

and

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

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

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

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

By M has *approximately-finite memory* we mean that given $\epsilon > 0$ there is an $\alpha \in \mathcal{Z}_+$ such that

$$|(Ms)(k) - (MW_{k,\alpha}s)(k)| < \epsilon, \quad k \in \mathcal{Z}_+,$$

for $s \in S$. (There is a slight difference here relative to the definition of approximately-finite memory in [4] where α is required to be positive.) Let $\tau \in \mathcal{Z}_+$. By M has *finite memory of width* τ is meant that we have

$$(Ms)(k) = (MW_{k,\tau}s)(k)$$

for all $k \in \mathcal{Z}_+$ and all $s \in S$.

Throughout the paper, λ denotes an $\mathbb{R}$-valued map on $[0, \infty)$ such that λ satisfies $\lambda(0) = 0$ as well as $\lambda(\eta) > 0$ for $\eta > 0$ and is bounded, nondecreasing, continuous, and subadditive. (λ is *subadditive* if $\lambda(\eta_1 + \eta_2) \le \lambda(\eta_1) + \lambda(\eta_2)$ for $\eta_1, \eta_2 \ge 0$.) We use $\mathcal{G}_\tau$ to denote the family of all causal time-invariant maps G from S to the set of $\mathbb{R}$-valued functions defined on $\mathcal{Z}_+$ such that G has finite memory of width τ, the functional $G(\cdot)(\tau) : S \to \mathbb{R}$ is continuous, and $\omega(\eta)$ defined by

$$\omega(\eta) = \sup_{\rho(s_a, s_b) \le \eta} |G(s_a)(\tau) - G(s_b)(\tau)|$$

satisfies $\omega(\eta) \le \lambda(\eta)$ for $\eta \ge 0$. (The conditions on $\mathcal{G}_\tau$ are redundant: The finite memory assumption implies causality, and also $\omega(\eta) \le \lambda(\eta)$ for $\eta \ge 0$ implies the continuity of $G(\cdot)(\tau)$.) For the important case in which $G(\cdot)(\tau)$ is Lipschitz continuous with Lipschitz constant ℓ for all $G \in \mathcal{G}_\tau$, we have $\omega(\eta) \le \lambda(\eta)$ for $\eta \ge 0$ with $\lambda(\eta) = \ell\eta$ for $\eta \le 2\sqrt{d}$ and $\lambda(\eta) = 2\sqrt{d}\,\ell$ for $\eta > 2\sqrt{d}$.

For any positive integers n and k, let P_n^k denote the family of maps P from $\mathbb{R}^k$ (viewed as row vectors) to $\mathbb{R}$ such that Px is a real polynomial of degree at most n in $x_1, \ldots, x_k$.

Finally, let k_τ stand for $d(\tau + 1)$, and for each $k \in \mathcal{Z}_+$ let L_k^τ denote the map from S to $[-1, 1]^{k_\tau}$ given by

$$L_k^\tau s = [s_1(k - \tau), \ldots, s_1(k), \ldots, s_d(k - \tau), \ldots, s_d(k)]$$

with $s_j(\zeta) = 0$ for $\zeta < 0$ and each j. In particular, for the special case in which $d = 1$, we have simply

$$L_k^\tau s = [s(k - \tau), \ldots, s(k)]$$

(which corresponds to the vector of outputs, given in reverse order, of the delay line with unit delays in Fig. 1).

2.2. Approximation Bounds

Our main result is the following.

THEOREM 1

(i) *For each $G \in \mathcal{G}_\tau$, and each positive integer n, there is a $P \in P_n^{k_\tau}$ such that*

$$|(Gs)(k) - P(L_k^\tau s)| < C\lambda\left(\frac{k_\tau^{3/2}}{n}\right)$$

for all s and k. Here C is a positive constant that does not depend on G, τ, d or n.

(ii) *There is a positive constant c, independent of τ, d and n, with the following property: for any positive integer n there is a $G \in \mathcal{G}_\tau$ for which*

$$\inf_{s \in S} |(Gs)(\tau) - P(L_\tau^\tau s)| \geq c\lambda\left(\frac{k_\tau^{3/2}}{n}\right) \qquad (2)$$

for all $P \in P_n^{k_\tau}$.

The proof is given in the appendix.

2.3. Comments

Since P in Theorem 1 is a polynomial map, the approximants $P(L_k^\tau s)$ of the theorem are discrete-time Volterra series.

The fact that the upper and lower bounds in Theorem 1 differ by only a multiplicative constant shows that C in Theorem 1 cannot be replaced with a function of n that approaches zero as $n \to \infty$. Similarly, c in the lower bound cannot be replaced with a function of n that is unbounded. In this sense the upper and lower bounds cannot be improved.

$P_n^{k_\tau}$ is a subspace of dimension $N_{k_\tau} \stackrel{\text{def}}{=} \binom{n+k_\tau}{k_\tau}$ of the space $\mathcal{C}(\mathbb{R}^{k_\tau})$ of continuous functions from $\mathbb{R}^{k_\tau}$ (viewed as row vectors) to $\mathbb{R}$. From the

theory of n-widths introduced by Kolmogorov (see [8]) and the result in [8] cited in the appendix in connection with (7), it follows that the lower bound (2) holds also for $P_n^{k_\tau}$ replaced with *any* subspace of $C(\mathbb{R}^k)$ of dimension N_{k_τ}. Thus, in so far as the approximation of the entire set $\mathcal{G}_\tau$ is concerned, and aside from numerical constants in the upper and lower bounds, the use of polynomials is as effective as the use of any other subspace of $C(\mathbb{R}^{k_\tau})$ of the same dimension. (In other words, the proposition noted in [8] concerning the effectiveness of polynomials holds in our setting too.)

It may be possible to give useful estimates of the constants C and c by going through arguments in [8] and supplying omitted details. We have not addressed that problem.

In Section I we (in effect) indicated that an approach to the approximation of maps that have approximately-finite-memory, but not finite memory, is to first obtain a suitably accurate finite-memory approximation. This determines τ. There are important cases in which τ can be determined analytically [10]. This is interesting because, for these cases and together with the upper bound in Theorem 1, it leads to a solution of the problem of specifying a bound on the needed order of a Volterra-series approximant so that a prescribed quality of uniform approximation is achieved.

In the following section we give a result corresponding to Theorem 1 for the approximation of not-necessarily-causal input–output maps with inputs and outputs that may depend on more than one variable. As mentioned earlier, this result is of interest, for example, in connection with image processing.

III. APPROXIMATION OF MULTIDIMENSIONAL INPUT–OUTPUT MAPS

3.1. Preliminaries

Let $\mathcal{Z}_+$, $\mathbb{R}$, d, C, and $\|\cdot\|$ be as described in Section 2.1. Let $\mathcal{Z}$ denote the set of all integers, and let m be a positive integer.

Let U denote the metric space of all maps u from $\mathcal{Z}^m$ to C, with the metric given by $\rho(u_a, u_b) = \sup\{\|u_a(\gamma) - u_b(\gamma)\|: \gamma \in \mathcal{Z}^m\}$. Here the set U is our set of inputs.

For each $\beta \in \mathcal{Z}^m$ and each $\alpha \in \mathcal{Z}_+$, let maps $W_\beta^\alpha : U \to U$ and $T^\beta : U \to U$ be defined by

$$(W_\beta^\alpha u)(\gamma) = \begin{cases} u(\gamma), & \max_i |\gamma_i - \beta_i| \leq \alpha, \\ 0, & \text{otherwise} \end{cases}$$

(where γ_i and β_i denote the ith components of γ and β, respectively) and

$$(T^\beta u)(\gamma) = u(\gamma - \beta).$$

Let M denote a map from U to the set of $\mathbb{R}$-valued functions defined on $\mathcal{Z}^m$. We say that M is *shift-invariant* if for each $\beta \in \mathcal{Z}^m$ we have $(MT^\beta u)(\gamma) = (Mu)(\gamma - \beta)$ for all γ and u.

We say that M is *myopic* if for each $\epsilon > 0$ there is an $\alpha \in \mathcal{Z}_+$ such that

$$|(Mu)(\gamma) - (MW_\gamma^\alpha u)(\gamma)| < \epsilon, \quad \gamma \in \mathcal{Z}^m$$

for $u \in U$. Let $r \in \mathcal{Z}_+$. By M has *finite myopicity of width* $2r$ we mean that we have

$$(Mu)(\gamma) = (MW_\gamma^r u)(\gamma)$$

for all $\gamma \in \mathcal{Z}^m$ and all $u \in U$ (if $m = 1$ and such an M is causal in the usual sense we say that it has *finite memory of width* r).

With λ as described in Section 2.1, let $\mathcal{K}_r$ denote the family of all shift-invariant maps K from U to the set of $\mathbb{R}$-valued functions defined on $\mathcal{Z}^m$ such that K has finite myopicity of width $2r$, the functional $K(\cdot)(0) : U \to \mathbb{R}$ is continuous, and $\omega(\eta)$ now defined by

$$\omega(\eta) = \sup_{\rho(s_a, s_b) \leq \eta} |K(s_a)(0) - K(s_b)(0)|$$

satisfies $\omega(\eta) \leq \lambda(\eta)$ for $\eta \geq 0$. (The conditions on $\mathcal{K}_r$ are redundant in that $\omega(\eta) \leq \lambda(\eta)$ for $\eta \geq 0$ implies the continuity of $K(\cdot)(0)$.) Paralleling the example in Section 2.1, for the important case in which $K(\cdot)(0)$ is Lipschitz continuous with Lipschitz constant ℓ for all $K \in \mathcal{K}_r$, we have $\omega(\eta) \leq \lambda(\eta)$ for $\eta \geq 0$ with $\lambda(\eta) = \ell\eta$ for $\eta \leq 2\sqrt{d}$ and $\lambda(\eta) = 2\sqrt{d}\ell$ for $\eta > 2\sqrt{d}$.

As in Section 2.1, for any positive integers n and k, let P_n^k denote the family of maps P from $\mathbb{R}^k$ (viewed as row vectors) to $\mathbb{R}$ such that Px is a real polynomial of degree at most n in $x_1, \ldots, x_k$.

Finally, let $\zeta_1, \ldots, \zeta_{(2r+1)^m}$ be an enumeration of the points of $\{-r, \ldots, r\}^m$, and here let k_r stand for $d(2r+1)^m$, and here for each γ let L_γ^r denote the map from U to $[-1,1]^{k_r}$ given by

$$
\begin{aligned}
L_\gamma^r u = [&u_1(\gamma + \zeta_1), \ldots, u_1(\gamma + \zeta_{(2r+1)^m}), \ldots, u_d(\gamma + \zeta_1), \\
&\ldots, u_d(\gamma + \zeta_{(2r+1)^m})].
\end{aligned}
$$

In particular, for the important special case in which $d = m = 1$ and $\zeta_i = -(r+1) + i$ for each i, we have simply

$$
L_\gamma^r u = [u(\gamma - r), \ldots, u(\gamma + r)].
$$

3.2. Approximation Bounds: Multidimensional Case

Our result is the following.

THEOREM 2

(i) *For each $K \in \mathcal{K}_r$, and each positive integer n, there is a $P \in P_n^{k_r}$ such that*

$$
|(Ku)(\gamma) - P(L_\gamma^r u)| < C\lambda \left(\frac{k_r^{3/2}}{n} \right)
$$

for all u and γ. Here C is a positive constant that does not depend on K, r, m, d or n.

(ii) *There is a positive constant c, independent of r, d and n, with the following property: for any positive integer n there is a $K \in \mathcal{K}_r$ for which*

$$
\inf_{u \in U} |(Ku)(0) - P(L_0^r u)| \geq c\lambda \left(\frac{k_r^{3/2}}{n} \right) \tag{3}
$$

for all $P \in P_n^{k_r}$.

The proof parallels the proof of Theorem 1, and is given in the appendix. Of course comments similar to those in Section 2.3 are pertinent in the setting of Theorem 2.

References

[1] G. Cybenko, Approximation by superposition of a single function, *Mathematics of Control, Signals and Systems*, **2**, 303–314, 1989.
[2] H.N. Mhaskar and C.A. Micchelli, Approximation by superposition of sigmoidal and radial basis functions, *Advances in Applied Mathematics*, **3**, 350–373, 1992.
[3] J. Park and I.W. Sandberg, Approximation and radial-basis function networks, *Neural Computation*, **5**(2), 305–316, 1993.
[4] I.W. Sandberg, Structure theorems for nonlinear systems, *Multidimensional Systems and Signal Processing*, **2**(3), 267–286, 1991. (See also the Errata in **3**(1), 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.
[5] I.W. Sandberg, Approximation theorems for discrete-time systems, *IEEE Transactions on Circuits and Systems*, **38**(5), 564–566, 1991.
[6] I.W. Sandberg and L. Xu, Uniform approximation of discrete-space multidimensional myopic maps, *Journal of Circuits, Systems, and Signal Processing*, **16**(3), 387–403, 1997.
[7] A.F. Timan, *Theory of Approximation of Functions of a Real Variable*, Pergamon Press, Oxford, 1963.
[8] D.J. Newman and H.S. Shapiro, Jackson's theorem in higher dimensions, in *On Approximation Theory*, eds. P.L. Butzer and J. Korevaar, Birkhauser, Basel, 1964.
[9] H.N. Mhaskar and N. Hahm, Neural networks for functional approximation and system identification, *Neural Computation*, **9**, 143–159, 1997.
[10] M. Ciraula and I.W. Sandberg, to appear.

APPENDIX

Proof of Theorem 1

Consider (i). Let n and $G \in \mathcal{G}_\tau$ be given. We are to show that there is a P as described such that

$$|(GW_{k,\tau}s)(k) - P(L_k^\tau s)| < C\lambda \left(\frac{k_\tau^{3/2}}{n} \right) \tag{4}$$

for each s and k. We do this as follows.

Let Q denote the functional defined on $[-1, 1]^{k_\tau}$ by $Qv = (Gs)(\tau)$ where s is the element of S that satisfies $L_\tau^\tau s = v$ and $s(k) = 0$ for

$k \notin \{0, \ldots, \tau\}$. By the continuity of $G(\cdot)(\tau)$, Q is continuous with respect to the Euclidean norm $\| \cdot \|_{k_\tau}$ in $\mathbb{R}^{k_\tau}$. And, with ω_Q the modulus of continuity of Q,

$$\omega_Q(\eta) \overset{\text{def}}{=} \sup_{\|v_a - v_b\|_{k_\tau} \leq \eta} |Qv_a - Qv_b| \leq \sup_{\rho(s_a, s_b) \leq \eta} |(Gs_a)(\tau) - (Gs_b)(\tau)| \leq \lambda(\eta)$$

for $\eta \geq 0$.

By a version of Jackson's theorem [8, Theorem 4] and the inequality $\omega_Q(\eta) \leq \lambda(\eta)$, there is an element P of $P_n^{k_\tau}$ such that

$$|Qv - Pv| < C\lambda\left(\frac{k_\tau^{3/2}}{n}\right), \quad v \in [-1, 1]^{k_\tau}$$

where C is a constant that does not depend on G, τ, d or n. Thus,

$$|(Gw)(\tau) - P(L_\tau^\tau w)| < C\lambda\left(\frac{k_\tau^{3/2}}{n}\right) \tag{5}$$

for all $w \in S$.

Now let s and k be given. Suppose first that $k \geq \tau$. By the time-invariance of G,

$$(GW_{k,\tau}s)(k) = (Gw)(\tau)$$

with $w(\beta) = s(\beta + k - \tau)$ for $\beta = 0, 1, \ldots, \tau$, showing that

$$|(GW_{k,\tau}s)(k) - P(L_k^\tau s)| < C\lambda\left(\frac{k_\tau^{3/2}}{n}\right). \tag{6}$$

For $k < \tau$,

$$(GW_{k,\tau}s)(k) - P(L_k^\tau s) = (Gs)(k) - P(L_k^\tau s) = (Gw)(\tau) - P(L_\tau^\tau w)$$

where now $w(\beta) = 0$ for $\beta < (\tau - k)$ and $w(\beta) = s(\beta + k - \tau)$ for $\beta = (\tau - k), \ldots, \tau$, showing that (6) holds even if $k < \tau$. This proves part (i).

Now consider (ii). It follows directly from [8, Part (iii) of Theorem 1] that there is a positive constant c independent of τ, d and n such that given n there exists a continuous real functional Q on $[-1, 1]^{k_\tau}$ with

modulus of continuity μ satisfying $\mu(\eta) \leq \lambda(\eta)$ $(\eta \geq 0)$ for which

$$|Qv - Pv| \geq c\lambda\left(\frac{k_\tau^{3/2}}{n}\right), \quad \text{some } v \in [-1, 1]^{k_\tau} \tag{7}$$

for each $P \in P_n^{k_\tau}$. We assume without loss of generality that $Q(0) = 0$.
  Define the map G on S by

$$(Gs)(k) = \begin{cases} Q\Lambda_\tau A^{(k-\tau)}s, & k > \tau, \\ Q\Lambda_\tau T_{(\tau-k)}s, & k \leq \tau, \end{cases}$$

where Λ_τ is the map of S into $[-1, 1]^{k_\tau}$ given by $\Lambda_\tau u = L_\tau^\tau u$, and $A^{(k-\tau)}s$
with $k > \tau$ stands for the element of S defined by $(A^{(k-\tau)}s)(j) = s[(k-\tau)+j]$ for $j \in \mathcal{Z}_+$.
  This G has finite memory of width τ and is causal because
$Q\Lambda_\tau A^{(k-\tau)}W_{k,\tau}s = Q\Lambda_\tau A^{(k-\tau)}s$ for $t > \tau$ and $Q\Lambda_\tau T_{(\tau-k)}W_{k,\tau}s = Q\Lambda_\tau T_{(\tau-k)}s$ where $k \leq \tau$. To see that G is time-invariant, let β belong
to $\mathcal{Z}_+$ and observe that $(GT_\beta s)(k+\beta) = (Gs)(k)$ for $k \in \mathcal{Z}_+$ and that
$(GT_\beta s)(k) = 0$ for $k < \beta$. (Here we have used the condition that
$Q(0) = 0$.) Using $(Gs)(\tau) = Q\Lambda_\tau s$ as well as $\mu(\eta) \leq \lambda(\eta)$ $(\eta \geq 0)$ and the
continuity of Q, we see that $G(\cdot)(\tau)$ is continuous, and that

$$\sup_{\rho(s_a, s_b) \leq \eta} |G(s_a)(\tau) - G(s_b)(\tau)| = \mu(\eta)$$

for $\eta \geq 0$, showing that $G \in \mathcal{G}_\tau$. And, using (7) as well as $(Gs)(\tau) = Q\Lambda_\tau s$
and $\Lambda_\tau s = L_\tau^\tau s$, we have

$$\inf_{s \in S} |(Gs)(\tau) - P(L_\tau^\tau s)| \geq c\lambda\left(\frac{k_\tau^{3/2}}{n}\right)$$

for all $P \in P_n^{k_\tau}$. This completes the proof.

Proof of Theorem 2

Consider (i) of Theorem 2. Let n and $K \in \mathcal{K}_r$ be given. We are to show
that there is a P as indicated such that

$$|(KW_\gamma^r u)(\gamma) - P(L_\gamma^r u)| < C\lambda\left(\frac{k_r^{3/2}}{n}\right) \tag{8}$$

for each u and γ. We do this as follows.

Let Q denote the functional defined on $[-1, 1]^{k_r}$ by $Qv = (Ku)(0)$ where u is the element of U that satisfies $L_0^r u = v$ and $u(\gamma) = 0$ for $\gamma \notin \{-r, \ldots, r\}^m$. By the continuity of $K(\cdot)(0)$, Q is continuous with respect to the Euclidean norm $\|\cdot\|_{k_r}$ in $\mathbb{R}^{k_r}$. And, with ω_Q the modulus of continuity of Q,

$$\omega_Q(\eta) \overset{\text{def}}{=} \sup_{\|v_a - v_b\|_{k_r} \leq \eta} |Qv_a - Qv_b| \leq \sup_{\rho(u_a, u_b) \leq \eta} |(Ku_a)(0) - (Ku_b)(0)| \leq \lambda(\eta)$$

for $\eta \geq 0$.

By a version of Jackson's theorem [8, Theorem 4] and the inequality $\omega_Q(\eta) \leq \lambda(\eta)$ there is an element P of $P_n^{k_r}$ such that

$$|Qv - Pv| < C\omega\left(\frac{k_r^{3/2}}{n}\right), \quad v \in [-1, 1]^{k_r}$$

where C is independent of K, r, m, d and n. Now let u and γ be given, and set $v = L_0^r T^{-\gamma} u$. Since $Qv = (KW_0^r T^{-\gamma} u)(0)$,

$$|(KW_0^r T^{-\gamma} u)(0) - P(L_0^r T^{-\gamma} u)| < C\omega\left(\frac{k_r^{3/2}}{n}\right). \tag{9}$$

It is not difficult to check that the map M from U to the set of $\mathbb{R}$-valued functions defined on $\mathcal{Z}^m$ given by $(Mw)(\xi) = (KW_\xi^r w)(\xi)$ is shift-invariant. (By $W_\xi^r T^\beta w = T^\beta W_{(\xi-\beta)}^r w$ and the shift-invariance of K we have $(KW_\xi^r T^\beta w)(\xi) = (KW_{(\xi-\beta)}^r w)(\xi - \beta)$.) Thus, by (9) and $L_0^r T^{-\gamma} u = L_\gamma^r u$ we have (8). This proves (i).

Now consider (ii). It follows directly from [8, Part (iii) of Theorem 1] that there is a positive constant c independent of r, d and n such that given n there exists a continuous real functional Q on $[-1, 1]^{k_r}$ with modulus of continuity μ satisfying $\mu(\eta) \leq \lambda(\eta)$ $(\eta \geq 0)$ for which

$$|Qv - Pv| \geq c\lambda\left(\frac{k_r^{3/2}}{n}\right), \quad \text{some } v \in [-1, 1]^{k_r} \tag{10}$$

for each $P \in P_n^{k_r}$.

Define the map K on U by

$$(Ku)(\gamma) = Q\Lambda_r T^{-\gamma} u,$$

where Λ_r is the map of U into $[-1, 1]^{k_r}$ given by $\Lambda_r u = L_0^r u$.

This K has finite myopicity of width $2r$ because $\Lambda_r T^{-\gamma} W_\gamma^r u = \Lambda_r T^{-\gamma} u$. To see that K is shift-invariant, let β belong to $\mathcal{Z}^m$ and observe that $(KT^\beta u)(\gamma) = Q\Lambda_r T^{-\gamma} T^\beta u = Q\Lambda_r T^{-(\gamma-\beta)} u = (Ku)(\gamma - \beta)$. Using $(Ku)(0) = Q\Lambda_r u$ as well as $\mu(\eta) \leq \lambda(\eta)$ $(\eta \geq 0)$ and the continuity of Q, we see that $K(\cdot)(0)$ is continuous, and that

$$\sup_{\rho(u_a,u_b)\leq\eta} |K(u_a)(0) - K(u_b)(0)| = \mu(\eta)$$

for $\eta \geq 0$, showing that $K \in \mathcal{K}_r$. And, using (10) as well as $(Ku)(0) = Q\Lambda_r u$ and $\Lambda_r u = L_0^r u$, we have

$$\inf_{u\in U} |(Ku)(0) - P(L_0^r u)| \geq c\lambda\left(\frac{k_r^{3/2}}{n}\right)$$

for all $P \in P_n^{k_r}$. This completes the proof of Theorem 2.