Fading Memory and the Problem of Approximating Nonlinear Operators with Volterra Series

STEPHEN BOYD AND LEON O. CHUA, FELLOW, IEEE

Abstract —Using the notion of *fading memory* we prove very strong versions of two folk theorems. The first is that *any time-invariant (TI) continuous nonlinear operator* can be approximated by a *Volterra series operator*, and the second is that the approximating operator can be realized as a *finite-dimensional linear dynamical system* with a *nonlinear readout map*. While previous approximation results are valid over *finite time intervals* and for signals in *compact sets*, the approximations presented here hold *for all time* and for signals in useful (noncompact) sets. The discrete-time analog of the second theorem asserts that *any TI operator with fading memory* can be approximated (in our strong sense) by a *nonlinear moving-average operator*.

Some further discussion of the notion of fading memory is given.

I. Introduction

A *Volterra Series Operator* is one of the form

$$Nu(t) = h_0 + \sum_{n=1}^{\infty} \int \cdots \int h_n(\tau_1, \cdots, \tau_n) \cdot u(t-\tau_1) \cdots u(t-\tau_n)\, d\tau_1 \cdots d\tau_n$$

and is a generalization of the convolution description of linear time-invariant (LTI) operators to time-invariant (TI) nonlinear operators. The usefulness of Volterra series hinges on their ability to model a very wide class of nonlinear operators. Two general approaches can be taken to establish this.

First, one can demonstrate that many explicitly described systems have input/output (I/O) operators given by Volterra series. Sandberg [1] has established that a wide class of systems have I/O operators which are given by Volterra series, the requirement being, roughly speaking, that the nonlinearities are *analytic*. Thus an op-amp (with transistors modeled by the Ebers–Moll equations, which are analytic) has an I/O operator expressible, at least for small inputs, as a Volterra series.

But many common nonlinear systems are modeled with nonanalytic nonlinearities. For example the I/O operator of a control system containing an ideal *saturator*, that is, a memoryless nonlinearity with characteristic

$$\mathrm{SAT}(a) \triangleq \begin{cases} \mathrm{sign}(a) & |a| > 1 \\ a & |a| \le 1 \end{cases}$$

(which of course is not analytic) can easily be shown not to have a Volterra series representation valid for any inputs for which the saturator threshold is exceeded.[1] One could reasonably argue that even though the I/O operator of such a control system does not have an exact representation as a Volterra series operator, it could be *approximated* by one, for example by replacing the saturator with a polynomial approximation. But exactly what do we mean by approximate here, that is, over what set of signals and in what sense can the I/O operator be approximated by a Volterra series operator? This is one of the questions addressed in this paper.

The second approach to establishing the generality of Volterra series is *axiomatic* in style, and conceptually more satisfying. Here one demonstrates that under only a few physically reasonable assumptions about an operator N (such as causality, time-invariance, and some form of continuity) there is a Volterra series operator $\hat{N}$ which approximates, in some sense, N. No assumption whatever is made concerning the internal structure or realization of N.

The idea of such an approximation is not new, and in fact is discussed in the original work of Volterra [3], who cites Frechet [4]. Even in this early work one can find the basic idea (clouded by archaic mathematics): there is an analogy between ordinary polynomials and finite Volterra series, and hence some analog of the Weierstrass approximation theorem should apply to approximating general nonlinear operators with finite Volterra series.

Wiener rekindled interest in this problem at MIT in the forties and fifties [5]–[7], and since then various researchers have considered the problem [8]–[11]. A clear discussion of a typical approximation result can be found on pages 34–37 of Rugh's book [12]. The result presented there is:

Theorem: Let K be a compact subset of $L^2[0,T]$ and suppose N: $K \to C[0,T]$ is a TI causal continuous operator. Let $\epsilon > 0$. Then there is a Volterra series operator $\hat{N}$ such that for all $u \in L$ and $0 \le t \le T$

$$|Nu(t) - \hat{N}u(t)| \le \epsilon \tag{1.1}$$

(the notation will be precisely defined soon).

Roughly speaking, all of this work has the following problems:

(1) The input signals are nonzero only on a finite time interval $[0, T]$,

Manuscript received November 29, 1984; revised May 5, 1985. This work was supported in part by the Office of Naval Research under Contract N00014-76-C-0572, the National Science Foundation under Grant ECS80-20-640, and by the Fannie and John Hertz Foundation.

S. Boyd is with the Electrical Engineering Department, Stanford University, Stanford, CA 94305.

L. O. Chua is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720.

[1] A Volterra series operator which is linear for small inputs is in fact linear for all inputs; see Boyd *et al.* [2].

0098-4094/85/1150$01.00 ©1985 IEEE

(2) The approximation is always on a compact subset of the input space,

(3) The approximation only holds over a finite time interval $[0, T]$.

While demonstrating that Volterra series operators can, at least in a very weak sense, approximate a general TI causal continuous operator, these results are not really satisfying. (1), (2) and (3) are severe restrictions: we would really like an approximation which allows input signals defined on *infinite time intervals* and which approximates the operator N over an *infinite time interval.* Problems (1)–(3) preclude, for example, periodic forcing signals which start at $t = 0$. Rugh concludes his discussion with the following comments concerning (2): "...I should point out that the main drawback is in the restrictive input space K. The compactness requirement rules out many of the more natural choices for K."

The compactness requirement (2) and the finite time interval requirements (1) and (3) come from the use of the Stone–Weierstrass theorem, which underlies all of these approximation results, and so might seem unavoidable. Indeed we will see an example which demonstrates that without additional assumptions we *cannot* find an approximation for which (1.1) holds for all $t \in \mathbb{R}$. But we will demonstrate that all of these drawbacks can be overcome if the usual continuity assumption on N is strengthened slightly to ensure that N has *fading memory.* In particular, our approximation results (I) will hold over useful (noncompact) sets of signals, possibly nonzero for all $t \in \mathbb{R}$, and (II) will hold for all time, not just on an interval $[0, T]$.

The structure of this paper is as follows: Section II contains the preliminaries, Section III introduces the fading memory concept, Sections IV and V contain the main approximation theorems. In Section VI we give discrete-time approximation results, one of which concerns approximation by *nonlinear moving average* (*NLMA*) *operators.* In Section VII we consider a simple illustrative example, in Section VIII we give two other applications of the notion of fading memory, and in Section IX we mention how the results of this paper can be put in a cleaner (but less concrete) mathematical form.

II. Notation, Definitions, and Preliminary Discussion

2.1. Notation and Definitions

$C(\mathbb{R})$ will denote the space of bounded continuous functions $:\mathbb{R} \to \mathbb{R}$, with the usual norm $\|u\| \triangleq \sup_{t \in \mathbb{R}} |u(t)|$. $\mathbb{R}_-$ will denote $\{t | t \le 0\}$, and $C(\mathbb{R})$ will denote the space of bounded continuous functions on $\mathbb{R}_-$, with the usual norm $\|u\| \triangleq \sup_{t \le 0} |u(t)|$. A function F from $C(\mathbb{R}_-)$ into $\mathbb{R}$ is called a *functional* on $C(\mathbb{R}_-)$, and a function N from $C(\mathbb{R})$ into $C(\mathbb{R})$ is called an *operator.* We will usually drop the parentheses around the arguments of functionals and operators, writing, e.g., Fu for $F(u)$ and $Nu(t)$ for $N(u)(t)$.

U_τ will denote the τ-second *delay operator* defined by

$$(U_\tau u)(t) \triangleq u(t - \tau).$$

We say an operator N is *time-invariant* (TI) if $U_\tau N = N U_\tau$ for all $\tau \in \mathbb{R}$. N is *causal* if $u(\tau) = v(\tau)$ for $\tau \le t$ implies $Nu(t) = Nv(t)$. N is *continuous* if it is a continuous function $:C(\mathbb{R}) \to C(\mathbb{R})$.

With each TI causal operator N we associate a functional F on $C(\mathbb{R}_-)$ defined by

$$Fu \triangleq Nu_e(0) \tag{2.1.1}$$

for $u \in C(\mathbb{R}_-)$, where

$$u_e(t) \triangleq \begin{cases} u(t), & t \le 0 \\ u(0), & t > 0 \end{cases}$$

is just a continuous extension of u to $C(\mathbb{R})$ (any other would do). In words, F maps the *past input* to N (which is an element of $C(\mathbb{R}_-)$) into the *present output* of N (which is in $\mathbb{R}$). N can be recovered from its associated functional F via

$$Nu(t) = FPU_{-t}u \tag{2.1.2}$$

where $P: C(\mathbb{R}) \to C(\mathbb{R}_-)$ truncates an element $u \in C(\mathbb{R})$ into an element of $C(\mathbb{R}_-)$:

$$Pu(t) \triangleq u(t) \qquad \text{for} \quad t \le 0. \tag{2.1.3}$$

It is easy to see that N is continuous if and only if F is, so equations (2.1.1) and (2.1.2) establish a one-to-one correspondence between TI causal continuous operators N and continuous functionals F on $C(\mathbb{R}_-)$. For this reason we often see nonlinear *functionals* studied, where we are really interested in their associated TI *operators.* This has caused some confusion; some authors have mistakenly used the word *functional* to refer to what are really *operators.*

We can reexpress causality and continuity as follows: A TI operator N is causal and continuous iff for each $u \in C(\mathbb{R})$ and $\epsilon > 0$ there is a $\delta > 0$ such that for all v

$$\sup_{t \le 0} |u(t) - v(t)| < \delta \to |Nu(0) - Nv(0)| < \epsilon. \tag{2.1.4}$$

That is, a TI operator N satisfying (2.1.4) is causal and continuous, and a TI causal continuous operator satisfies (2.1.4).

2.2. Finite Volterra Series

Definition: A (finite) *Volterra Series Operator* $N: C(\mathbb{R}) \to C(\mathbb{R})$ is one of the form

$$Nu(t) = h_0 + \sum_{n=1}^{K} \int_0^\infty \cdots \int_0^\infty h_n(\tau_1, \cdots, \tau_n) \cdot u(t - \tau_1) \cdots u(t - \tau_n) \, d\tau_1 \cdots d\tau_n \tag{2.2.1}$$

where $h_n \in L^1(\mathbb{R}_+^n)$, that is,[2]

$$\int_0^\infty \cdots \int_0^\infty |h_n(\tau_1, \cdots, \tau_n)| \, d\tau_1 \cdots d\tau_n < \infty.$$

(Sometimes a finite Volterra series operator is called a

[2] Volterra series with integrable kernels might be called *stable* Volterra series; there is another interpretation of (2.2.1) which could be called *finite-time* Volterra series. For finite-time Volterra series the kernels are required to be locally integrable, but the inputs are restricted to be zero for negative time. Roughly speaking, this allows unstable systems to be considered.

polynomial operator.) That such an N is a TI causal continuous operator is easily verified; a proof can be found in Boyd *et al.* [2].

III. The Fading Memory Concept

Roughly speaking, an operator is *continuous* if input signals which are *close* (meaning, the *peak deviation* of the signals *over all past time* is small) have present outputs which are close. We will see that a slight strengthening of continuity is much more useful. Intuitively, an operator has *fading memory* if two input signals which are close in the *recent past*, but not necessarily close in the *remote past* yield present outputs which are close. For dynamical systems, fading memory is related to the notion of a *unique steady state* (see Section 8.2).

The concept of fading memory has a history at least as long as Volterra series themselves. Indeed we find it in Volterra [3, p. 188]:

> A first extremely natural postulate is to suppose that the influence of the (input) a long time before the given moment gradually fades out.

and in Wiener [5, p. 89]:

> We are assuming (the output) of the network does not depend on the infinite past. If the response of this apparatus depends on the remote past, then the Brownian motion is not a good approximation because we shall always have to consider the remote past. So we are considering networks in which the output is asymptotically independent of the remote past input...

and in various other work over the years [13], [6]. In [14] Root mentions operators with finite memory. The fading memory assumption, then, is by no means a new stronger restriction on the operators to be approximated. It is simply an old assumption whose full power has not been used.

How should we define fading memory? The problem is that in (2.14) we want $Nu(0)$ to depend less and less on the input when elapsed time $-t$ is large. To do this we simply introduce a weight in (2.1.4).

Definition: N has *Fading Memory* (FM) on a subset K of $C(\mathbb{R})$ if there is a decreasing function w: $\mathbb{R}_+ \to (0,1]$, $\lim_{t\to\infty} w(t) = 0$, such that for each $u \in K$ and $\epsilon > 0$ there is a $\delta > 0$ such that for all $v \in k$

$$\sup_{t \le 0} |u(t) - v(t)| w(-t) < \delta \to |Nu(0) - Nv(0)| < \epsilon. \tag{3.1}$$

(This should be compared to (2.1.4).)

w will be called the weighting function; we will say that N has a w-fading memory, for example if $w(t) = e^{-\lambda t}$ then we might say N has a λ-exponentially fading memory on K. Note that since $w(t) \le 1$, an operator with FM is continuous, so FM is indeed stronger than continuity.[3]

[3] Our requirements on the weighting function w are more stringent than necessary. All we really need is $w \geqslant 0$ and $\lim_{t\to\infty} w(t) = 0$; our additional assumptions simplify some of the proofs in the sequel.

The FM property can be clearly expressed in terms of the functional F associated with N as follows: On $C(\mathbb{R}_-)$ define the *weighted norm*

$$\|u\|_w \triangleq \|u(t)w(-t)\| = \sup_{t \le 0} |u(t)w(-t)|. \tag{3.2}$$

Then N has FM on K if and only if F is *continuous* with respect to the weighted norm $\|\cdot\|_w$ on $PK \triangleq \{Pu | u \in K\}$.

Remark 1: As in (2.1.4) above, if a TI N has fading memory, then N is causal.

Remark 2: It is interesting to note that this is very close to Volterra's "definition" of fading memory given on p. 188 of [3] (which unfortunately is not clear enough to be a real definition).

Remark 3: For LTI operators, having a fading memory is equivalent to having a convolution representation; see Section 8.1.

Remark 4: It can be shown that all finite Volterra series operators have fading memory on all of $C(\mathbb{R})$

Perhaps the best way to appreciate the notion of fading memory is to consider an example of a continuous operator which does not have fading memory.

Example (Peak-Hold Operator): Define N_{pk}: $C(\mathbb{R}) \to C(\mathbb{R})$ by

$$N_{pk}u(t) \triangleq \sup_{\tau < t} u(\tau)$$

that is, N_{pk} is a *peak-hold* operator. N_{pk} is continuous, since for all u, $v \in C(\mathbb{R})$

$$\|N_{pk}u - N_{pk}v\| \le \|u - v\|.$$

Nevertheless N_{pk} does not have a fading memory.[4]

Let us consider the problem of approximating N_{pk} by a Volterra series operator $\hat{N}$. Consider the signal

$$u_0(t) \triangleq \begin{cases} 1 - |t|, & |t| \le 1 \\ 0, & |t| > 1. \end{cases}$$

Then

$$Nu_0(t) = \begin{cases} 0, & t \le -1 \\ 1, & t \ge 0. \end{cases}$$

Now for *any* Volterra series operator $\hat{N}$ we have

$$\hat{N}u_0(t) = h_0 \qquad \text{for} \quad t < -1$$

and

$$\lim_{t\to\infty} \hat{N}u_0(t) = h_0.$$

(This is a consequence of the steady-state theorem [2].) Hence for *any* Volterra series operator $\hat{N}$

$$\|N_{pk}u_0 - \hat{N}u_0\| \ge \max\{|h_o|, |1 - h_0|\} \ge \tfrac{1}{2}.$$

Thus we may conclude *no Volterra series operator* can approximate N_0 within 0.1 over all time, even for the single input u_0. In fact the same argument holds for *any* operator $\hat{N}$ with fading memory, if we substitute $\hat{N}0$ (which must be a constant) for h_0. In particular, N_{pk} itself does not have fading memory.

[4] There are also continuous LTI operators which don't have fading memory, but they are quite pathological; see Section A3.

This example suggests that approximation results which rely only on the continuity of the operator, and no fading memory assumption, will be very weak. In particular, the approximations need not hold for all time, even on compact sets of signals (in this example, the signal set has only one element, u_0, and so is compact). And yet a very strong approximation is possible for operators with fading memory.

IV. Approximation by Volterra Series

Theorem 1 (Approximation by Volterra Series): Let $\epsilon > 0$ and

$$K \triangleq \{u \in C(\mathbb{R}) \mid \|u\| \le M_1,\ \|U_\tau u - u\| \le M_2 \tau \text{ for } \tau \ge 0\}. \tag{4.1}$$

Suppose that N is *any* TI operator with fading memory on K. Then there is a finite Volterra series operator $\hat{N}$ such that for all $u \in K$

$$\|Nu - \hat{N}u\| \le \epsilon. \tag{4.2}$$

Remark 1: The assumption on N is extremely weak. As mentioned earlier, it does not in any way concern the internal structure or realization of N. For example, N could arise from a nonlinear PDE, but even this is not necessary.

Remark 2: We can reexpress K as

$$K = \{u \in C(\mathbb{R}) \mid |u(t)| \le M_1,\ |u(s) - u(t)| \le M_2(s-t) \text{ for } t \le s\}.$$

Thus K can be described as those signals bounded by M_1 and having Lipschitz constant M_2, that is, *slew-limited* by M_2.[5]

Remark 3: The signals in K are not "time-limited" (i.e., zero outside of some interval such as $[0, T]$), and the approximation, $|Nu(t) - \hat{N}u(t)| \le \epsilon$ holds *for all* $t \in \mathbb{R}$, not just in some interval $[0, T]$ (cf. the theorem in Section I, (1.1)).

Remark 4: K is *not* a compact subset of $C(\mathbb{R})$!

Before starting the proof of Theorem 1, we state the Stone–Weierstrass theorem in a convenient form (see, e.g., Dieudonne [15]):

Suppose E is a compact metric space and $\boldsymbol{G}$ a set of continuous functionals on E which separate points, that is, for any distinct u, $v \in E$ there is a $G \in \boldsymbol{G}$ such that $Gu \ne Gv$. Let F by any continuous functional on E and $\epsilon > 0$. Then there is a polynomial $p: \mathbb{R}^M \to \mathbb{R}$ and $G_1, \cdots, G_M \in \boldsymbol{G}$ such that for all $u \in E$

$$|Fu - p(G_1 u, \cdots, G_M u)| < \epsilon.$$

Proof of Theorem 1: Suppose K is given by (4.1) and N has fading memory on K, with weighting function w. Let F be the functional associated with N, given by (2.1.1), and define $K_- \triangleq PK$, that is

$$K_- = \{Pu \mid u \in K\}$$

(P is the projection (2.1.3).)

[5] In fact K can be any bounded equicontinuous set in $C(\mathbb{R})$; the K defined in (4.1), while far from the most general, has a nice engineering description.

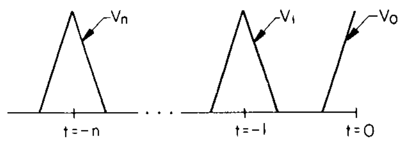

Fig. 1. A sequence of signals in K_- which contains no ($C(\mathbb{R}_-)$, i.e., uniformly) convergent subsequence. But in the weighted norm, $v_n \to 0$.

Lemma 1: Consider the weighted norm $\|\cdot\|_w$ on $C(\mathbb{R}_-)$ defined above in (3.2). K_- is compact with the weighted norm $\|\cdot\|_w$.

The proof uses the Arzela–Ascoli theorem and a diagonal argument and is in the Appendix, Section A1. Since Lemma 1 is the key to obtaining approximations valid for all time and on noncompact sets, some discussion is in order. Note that K_- is not compact with the standard norm $\|\cdot\|$. To see this, let

$$u_0(t) \triangleq \max\{0, M_1 - M_2|t|\}$$

and consider the sequence $v_n \triangleq PU_{-n}u_0$ in K_- (see Fig. 1). With the standard norm, this sequence has no convergent subsequence, and hence K_- is not compact in $C(\mathbb{R}_-)$. Yet intuitively, *to a device with fading memory* the sequence v_n should appear to be *converging to zero*, and this is indeed true: $\|v_n\|_w \to 0$ as $n \to \infty$. The idea of lemma 1 is that the fading memory makes K_- "appear" compact to our functional F.

Continuing our proof, we define a set of functionals $\boldsymbol{G}$ on K_- which are continuous with respect to the weighted norm $\|\cdot\|_w$.

$$\boldsymbol{G} \triangleq \left\{ G \,\middle|\, Gu = \int_0^\infty g(\tau) u(-\tau)\, d\tau,\ \int_0^\infty |g(\tau)| w(\tau)^{-1} d\tau < \infty \right\}. \tag{4.3}$$

Note that since $0 < w(t) \le 1$, the condition $g/w \in L^1(\mathbb{R}_+)$ implies $g \in L^1(\mathbb{R}^+)$. The fact that any $G \in \boldsymbol{G}$ is continuous with respect to the weighted norm $\|\cdot\|_w$ follows from

$$\begin{aligned}|Gu - Gv| &\le \left| \int_0^\infty \left(|g(t)| w(t)^{-1}\right)\left(|u(-t) - v(-t)| w(t)\right) dt \right. \\ &\le \left| \sup_{t \ge 0} |u(-t) - v(-t)| w(t) \int_0^\infty |g(t)| w(t)^{-1} dt \right. \\ &= \|u - v\|_w \int_0^\infty |g(t)| w(t)^{-1} dt.\end{aligned}$$

Lemma 2: The functionals $\boldsymbol{G}$ separate points in K_-.

Proof: Let u, $v \in K_-$, $u \ne v$. Define

$$g_0(t) \triangleq (u(-t) - v(-t)) w(t) e^{-t}.$$

Then

$$\int_0^\infty |g_0(t)| w(t)^{-1} dt \le \|u\| + \|v\| < \infty$$

so let G_0 be the functional in $\boldsymbol{G}$ associated with g_0 as in (4.3). Then

$$G_0 u - G_0 v = \int_0^\infty (u(-t) - v(-t))^2 w(t) e^{-t} dt > 0$$

since u and v are continuous and $u \neq v$. This proves Lemma 2.

Now by Lemmas 1 and 2 and the Stone-Weierstrass theorem, we conclude that there is a polynomial $p: \mathbb{R}^M \to \mathbb{R}$ and $G_1, \cdots, G_M \in \boldsymbol{G}$ such that for all $u \in K_-$

$$|Fu - p(G_1 u, \cdots, G_m u)| < \epsilon. \tag{4.4}$$

Explicitly writing out p:

$$\begin{aligned} p(G_1 u, \cdots, G_M u) &= \alpha_0 + \sum_{n=1}^{K} \sum_{i_1, \cdots, i_n \leq M} \alpha_{i_1 \cdots i_n} G_{i_1} u \cdots G_{i_n} u \\ &= h_0 + \sum_{n=1}^{K} \int \cdots \int h_n(\tau_1, \cdots, \tau_n) \\ &\quad \cdot u(-\tau_1) \cdots u(-\tau_n)\, d\tau_1 \cdots d\tau_n \end{aligned}$$

where $h_0 \triangleq \alpha_0$ and

$$h_n(\tau_1, \cdots, \tau_n) \triangleq \sum_{n=1}^{K} \sum_{i_1, \cdots, i_n \leq M} \alpha_{i_1 \cdots i_n} g_{i_1}(\tau_1) \cdots g_{i_n}(\tau_n)$$

and the g_l are the kernels of the functionals G_l as in (4.3).

We mentioned above that the g_l's are in $\boldsymbol{L}^1(\mathbb{R}_+)$, so $h_n \in \boldsymbol{L}^1(\mathbb{R}_+^n)$, and thus they are the kernels of a finite Volterra series operator which we call $\hat{N}$. We finally show that $\hat{N}$ is the desired finite Volterra series approximator of N. Let $u \in K$ and $t \in \mathbb{R}$. Then $PU_{-t}u \in K_-$, hence by (4.4)

$$\begin{aligned} |FPU_{-t}u - p(G_1 PU_{-t}u, \cdots, G_M PU_{-t}u)| \\ = |Nu(t) - \hat{N}u(t)| < \epsilon. \end{aligned} \tag{4.5}$$

Since (4.5) is true for all $t \in \mathbb{R}$, we conclude for all $u \in K$

$$\|Nu - \hat{N}u\| \leq \epsilon$$

which proves Theorem 1.

V. Approximation by Finite-Dimensional Dynamical Systems

5.1. Linear-Dynamic Polynomial Readout Approximators

The block diagram of $\hat{N}$ is shown in Fig. 2. Note that it consists of a single-input multi-output *linear time-invariant operator* followed by a multi-input single-output *memoryless nonlinearity*. One question arises immediately: can the LTI block be realized as a finite dimensional linear dynamical system? We will now show that it can.

In the proof of the approximation theorem we used only two properties of the set $\boldsymbol{G}$ of functionals: first, that each $G \in \boldsymbol{G}$ has a w-fading memory, and second, that $\boldsymbol{G}$ separates points in K_-.

Let us examine the first property. For a functional G on $\boldsymbol{C}(\mathbb{R}_-)$ given by

$$Gu = \int_0^\infty g(\tau) u(-\tau)\, d\tau \tag{5.1.1}$$

(where $g \in \boldsymbol{L}^1(\mathbb{R}_+)$) the necessary and sufficient condition that it have w-fading memory, that is, be continuous with respect to the w-weighted norm, is

$$\int_0^\infty |g(\tau)| w(\tau)^{-1} d\tau < \infty. \tag{5.1.2}$$

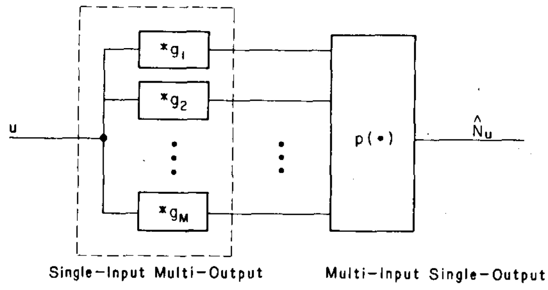

Fig. 2. Structure of the Volterra series approximator.

Now we make the observation that if a TI operator N has a w-fading memory, then it has a $\tilde{w}$-fading memory for any weighting function $\tilde{w}$ which dominates w (i.e., $\tilde{w}(t) \geq w(t)$). By using the weight

$$\tilde{w}(t) \triangleq \max\left\{ w(t), (1+t)^{-1} \right\}$$

(and relabeling it w) we may simply assume that the weight satisfies $w(t)^{-1} \leq 1 + t$. Under this assumption if follows that *every G which comes from a finite dimensional (exponentially stable) linear dynamical system has a w-fading memory*, since the integrand on left-hand side of (5.1.2) is exponentially decaying, that is,

$$\int_0^\infty |g(\tau)| w(\tau)^{-1} d\tau \leq \int_0^\infty M e^{-\lambda t}(1+t)\, dt < \infty$$

if $|g(t)| \leq M e^{-\lambda t}$. In the next subsection we will show that the G's which come from finite-dimensional linear dynamical systems separate points in $\boldsymbol{C}(\mathbb{R}_-)$. From this discussion we conclude:

Theorem 2 (Approximation by Finite-Dimensional Dynamical Systems): Let $\epsilon > 0$ and K be given by (4.1). Suppose that N is *any* TI operator with fading memory on K. Then there is a finite Volterra series operator $\hat{N}$ such that for all $u \in K$

$$\|Nu - \hat{N}u\| \leq \epsilon$$

where $\hat{N}$ is the I/O operator of the dynamical system

$$\dot{x} = Ax + bu \qquad y = p(x) \tag{5.1.3}$$

where A is the exponentially stable $M \times M$ matrix and $p: \mathbb{R}^M \to \mathbb{R}$ is a polynomial.

We have shown that under one extremely weak condition on a TI operator, namely that it have fading memory, it can be approximated in the strong sense of (4.2) by the I/O operator of a finite-dimensional *linear* dynamical system with a nonlinear (indeed, polynomial) readout map, as shown in Fig. 3. In principle, then, a dynamical system of the form (5.1.3) can always be used as a *macro-model*[16] of a complicated or large-scale nonlinear system, as long as the system has a fading memory. Whether an acceptable approximation is possible with M reasonably small is, of course, a harder question.

5.2. Wiener's Laguerre System

The idea that a system of the form (5.1.3), shown in Fig. 3, could be used to approximate a very wide class of TI operators is not new. Wiener considered the case where the LTI block in Fig. 3 consists of a set of Laguerre filters, that

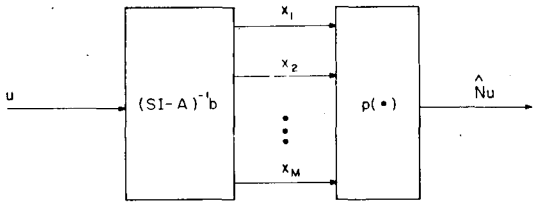

Fig. 3. Approximator consisting of linear dynamical system with polynomial readout map.

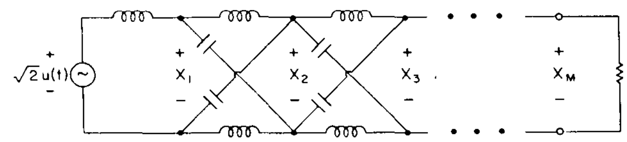

Fig. 4. Wiener and Lee's Laguerre lattice filter. All components have value 1.

is,

$$(sI-A)^{-1}b=\sqrt{2}\left[\frac{1}{1+s},\frac{1-s}{(1+s)^2},\cdots,\frac{(1-s)^{M-1}}{(1+s)^M}\right]^T \tag{5.1.4}$$

which Lee realized with the lattice filter shown in Fig. 4 (see Wiener [5, p. 92]).[6]

To see that Wiener's Laguerre system can approximate any TI causal operator with fading memory in the strong sense of Theorems 1 or 2 (a result evidently unknown to Wiener and his coworkers), we need to establish that the Laguerre functionals $\{L_1, L_2, \cdots\}$ given by

$$L_k u \triangleq \int_0^\infty l_k(t)u(-t)\,dt$$

where $\hat{l}_k(s)=\sqrt{2}(1-s)^{k+1}(1+s)^{-k}$, separate points in $C(\mathbb{R}_-)$. If not, there are $u_1, u_2 \in C(\mathbb{R}_-)$ such that $L_k u_1 = L_k u_2$ for all k. Let $u=u_1-u_2$, so that $L_k u=0$ for all k. We will show that $u=0$, which will prove that the Laguerre functionals separate points in $C(\mathbb{R}_-)$. Note that $l_k(t)e^{t/2} \in L^2(\mathbb{R}_+)$ and $u(-t)e^{-t/2} \in L^2(\mathbb{R}_+)$ and

$$L_k u = \int_0^\infty \left(l_k(t)e^{t/2}\right)\left(u(-t)e^{-t/2}\right)dt = 0$$

for all k. But the span of the functions $l_k(t)e^{t/2}$ is dense in $L^2(\mathbb{R}_+)$, so we conclude $u(-t)e^{-t/2}=0$ and hence $u=0$. This proves that the Laguerre functionals separate points in $C(\mathbb{R}_-)$; since they are a subset of the functionals which come from finite-dimensional linear dynamical systems, *a fortiori* these functionals separate points, a fact used in the previous subsection. Of course there are many other sequences of functionals which separate points in $C(\mathbb{R}_-)$.

5.3. A Note on Approximation by Bilinear Systems

The dynamical system approximator (5.1.3) can be realized as a bilinear system, that is, one of the form

$$\dot{z} = Ez + Fzu + Gu \tag{5.3.1}$$

$$y = Hz \tag{5.3.2}$$

[6] The only real difference between (5.1.4) and (5.1.3) is that in (5.1.4) we require the minimal polynomial of A to be $(s+1)^M$, since a change of coordinates can change the numerator polynomials. See, e.g., Section 7.2.

where $z \in \mathbb{R}^r$ (usually r is much larger than M) and $z(0)=0$. In fact this is a special case of an exercise in Rugh's book [12, p. 130]; here is a simple way to see it. Suppose the polynomial p in (5.1.3) is of degree n.

Let z be a vector consisting of all

$$r=\sum_{k=0}^{n}\binom{M+k-1}{k}$$

monomials of degree $\leq n$ formed from $x_1, \cdots, x_m$. Clearly we can write $y=p(x)$ in the form (5.3.2), where H contains the coefficients of p.

We will now verify that z satisfies an equation of the form (5.3.1). Consider the lth component of z, say $z_l = x_1^{i_1} \cdots x_M^{i_M}$, where $i_1 +, \cdots, + i_M \leq n$. Then

$$\dot{z}_l = \sum_{m=1}^{M} i_m \dot{x}_m x_1^{i_1} \cdots x_m^{m-1} \cdots x_M^{i_M} \tag{5.3.3}$$

$$= \sum_{m,k=1}^{M} i_m a_{mk} x_k x_1^{i_1} \cdots x_m^{i_m-1} \cdots x_m^{i_m} \tag{5.3.4}$$

$$+ \sum_{m=1}^{M} i_m x_1^{i_1} \cdots x_m^{i_m-1} \cdots x_M^{i_M} b_m u \tag{5.3.5}$$

using (5.1.3). Since each monomial in (5.3.4) and (5.3.5) has degree (in x) $\leq n$, we can reexpress this as

$$\dot{z}_l = \sum_{p=1}^{r} E_{lp} z_p + \sum_{p=1}^{r} F_{lp} z_p + G_l u$$

which is of the form (5.3.1).

In (5.3.1) the readout map is *linear*, but the vector field contains the product term Fzu (cf. (5.1.3)).

Approximation by bilinear systems has received much attention, but in a context different from that considered here. Usually (but not always) the systems to be approximated are dynamical systems with analytic vector fields. The approximation is generally not in an I/O sense, but rather in the sense of a perturbational expansion of x in u, meaning the input-to-state maps agree to order r in u. See, for example, Fliess [17], Sussman [18], or Brockett [19].

The discrete-time analog of bilinear systems are *state-affine systems*, Which have been used to model complicated processes, e.g., in [20].

VI. Discrete-Time Theorems

6.1. Approximation by Discrete-Time Volterra Series

In this section we present analogous results for discrete-time systems. $\mathbb{Z}$ will denote the integers, $\mathbb{Z}_+(\mathbb{Z}_-)$ the nonnegative (nonpositive) integers. Our signal space $C(\mathbb{R})$ is replaced by l^∞, the space of bounded sequences (i.e., functions $: \mathbb{Z} \to \mathbb{R}$) with norm

$$\|u\| \triangleq \sup_k |u(k)|.$$

The definitions of time-invariance, causality, and fading memory for discrete-time systems require only notational changes. For example a TI operator $N: l^\infty \to l^\infty$ has fading memory on a subset K of l^∞ if there is a decreasing sequence $w: \mathbb{Z}_+ \to (0,1]$, $\lim_{k\to\infty} w(k)=0$, such that for each $u \in K$ and $\epsilon > 0$ there is a $\delta > 0$ such that for all

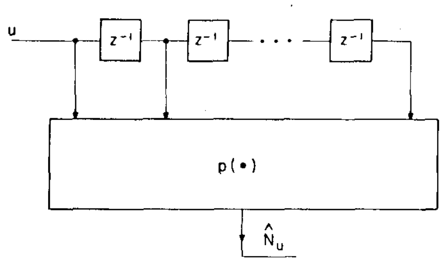

Fig. 5. Block diagram of the nonlinear moving-average (NLMA) operator.

$v \in K$

$$\sup_{k \le 0} |u(k)-v(k)|w(-k) < \delta \rightarrow |Nu(0)-Nv(0)| < \epsilon$$

(cf. (3.1)).

A (finite) discrete-time Volterra series operator N: $l^\infty \rightarrow l^\infty$ is one of the form

$$Nu(k) = h_0 + \sum_{n=1}^{K} \sum_{i_1,\cdots,i_n \ge 0} h_n(i_1,\cdots,i_n) \cdot u(k-i_1)\cdots u(k-i_n)$$

where $h_n \in l^1(\mathbb{Z}_+^n)$, that is,

$$\sum_{i_1,\cdots,i_n \ge 0} |h_n(i_1,\cdots,i_n)| < \infty$$

(cf. (2.2.1)).

Theorem 3 (Discrete-Time Approximation Theorem): Let $\epsilon > 0$ and

$$K \triangleq \{u \in l^\infty | \|u\| \le M_1\}$$

Suppose that N is *any* TI operator : $l^\infty \rightarrow l^\infty$ with fading memory on K. Then there is a finite Volterra series operator $\hat{N}$ such that for all $u \in K$

$$\|Nu - \hat{N}u\| \le \epsilon.$$

Remark: In the discrete-time theorem there is no "slew-limit" requirement on the signals in K; K here is just the ball of radius M_1 in l^∞.

In the next subsection we will see a stronger form of Theorem 3, so we omit the proof.

6.2. *Approximation by Nonlinear Moving-Average Operators*

As in Section V, the Volterra series approximator $\hat{N}$ can be realized as a finite-dimensional LTI dynamical system with a polynomial readout map. But for discrete-time systems we can choose the LTI dynamical system to have a particularly simple form: its transfer function can be simply

$$H_{xu}(z) = [1, z^{-1}, \cdots, z^{-M+1}]^T.$$

(This should be compared to the Laguerre system described in Section 5.2.) The approximator has the block diagram shown in Fig. 5; $\hat{N}$ is simply a *nonlinear moving-average operator*. To summarize:

Theorem 4 (NLMA Approximation Theorem): Let $\epsilon > 0$, K be any ball in l^∞, and suppose N is any TI operator: $l^\infty \rightarrow l^\infty$ with fading memory on K.

Then there is a polynomial p: $\mathbb{R}^M \rightarrow \mathbb{R}$ such that for all $u \in K$

$$\|Nu - \hat{N}u\| \le \epsilon$$

where $\hat{N}$ is the NLMA operator given by

$$\hat{N}u(k) \triangleq p(u(k), u(k-1), \cdots, u(k-M+1)).$$

The proof is in Section A2. Note that this theorem implies Theorem 3, since every NLMA operator with polynomial nonlinearity is also a finite Volterra series operator.

VII. A Simple Example

In this section we consider a simple example, one which illustrates some of the previous ideas and results. We consider the simple RMS detector N shown in Fig. 6(a), and show how a Volterra series approximation and a Laguerre system approximation can be found. More precisely, N is given by

$$Nu(t) \triangleq \left\{0.1\int_0^\infty e^{-0.1(t-\tau)}\left(\int_0^\infty e^{-(\tau-s)}u(s)\,ds\right)^2 d\tau\right\}^{1/2}.$$

We chose this example for several reasons. First, N has no Volterra series representation. To see this, suppose N were a Volterra series operator with kernels h_n. Let $u(t) = \alpha$, a constant. For any Volterra series operator N, $N\alpha$ is also a constant, in fact an analytic function of α (see Boyd *et al.* [2]). But in this case $N\alpha = |\alpha|$, which is not even differentiable at $\alpha = 0$, let alone analytic. So our RMS detector N is not given (exactly) by a Volterra series. Yet it can be shown to have a fading memory on any set K of the form (4.1), and hence our approximation theorems hold for this N.

Another reason for choosing this example is that it is typical of the operators for which the Laguerre system approximation requires very many terms, that is, N is hard to approximate with a Laguerre system. Roughly speaking, this is because N has its nonlinearity near the *input*, and we seek to approximate N with a system with nonlinearity at the *output*.

7.1. *Finding a Volterra Series Approximation*

To find a Volterra series approximation of N on the set K given by (4.1), we find a polynomial $q(x)$ such that $|q(x) - \sqrt{|x|}| < \epsilon$ for $|x| \le M_1^2$.[7]

The mean-square operator N_1 shown in Fig. 6(b) is a Volterra series operator, its only nonzero kernel

$$h_2(\tau_1,\tau_2) = 19^{-1}(e^{1.9\min\{\tau_1,\tau_2\}} - 1)e^{-(\tau_1+\tau_2)}. \qquad (7.1.1)$$

It follows that the operator $\hat{N}_{\mathrm{vlt}}$ shown in Fig. 6(c) is a Volterra series operator, whose kernels could be computed, if desired, from (7.1.1) and the composition formula [2]. For $u \in K$ we have $0 \le N_1 u \le M_1^2$ and hence,

$$\|Nu - \hat{N}_{\mathrm{vlt}}u\| \le \epsilon, \qquad \text{for} \quad u \in K.$$

[7] For example, let q_M be the even polynomial of degree $2M$ which agrees with $\sqrt{|x|}$ at the points $0, M_1^2/M, \ldots, M_1^2$. Then for M large enough, q_M will work.

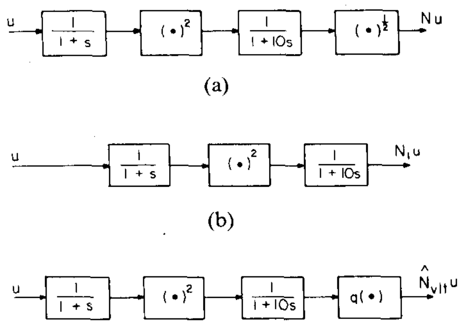

Fig. 6. (a) RMS operator N. (b) Mean square operator N_1, given by a Volterra series. (c) Approximating Volterra series operator $\hat{N}_{\mathrm{vlt}}$.

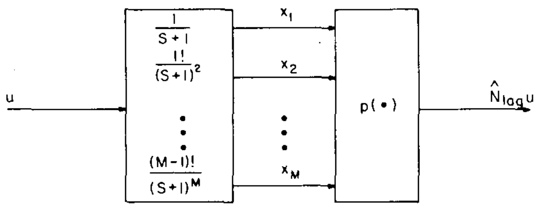

Fig. 7. Laguerre system approximator $\hat{N}_{\mathrm{lag}}$.

7.2. Finding a Laguerre System Approximation

We will now show how a Laguerre approximation to N can be found. It will suffice to find a Laguerre system approximation to the mean-square operator N_1 shown in Fig. 6(b), since passing its output through a polynomial $q(\cdot)$ which approximates the squareroot operator will yield a Laguerre system approximation of the overall operator N, as in the previous subsection.

Consider the system $\hat{N}_{\mathrm{lag}}$ shown in Fig. 7, where the readout polynomial p is homogeneous of degree two, that is

$$p(x_1,\cdots,x_m)=\sum_{i,j=1}^{M}\beta_{ij}x_ix_j.$$

This $\hat{N}_{\mathrm{lag}}$ can be transformed to a Laguerre system via the change of coordinates $\bar{x}=Tx$, where T is the (constant, invertible) matrix such that

$$T\begin{bmatrix}(1+s)^{-1}\\ \vdots\\ (M-1)!(1+s)^{-M}\end{bmatrix}=\sqrt{2}\begin{bmatrix}(1+s)^{-1}\\ \vdots\\ (1-s)^{M-1}(1+s)^{-M}\end{bmatrix}.$$

$\hat{N}_{\mathrm{lag}}$ is a Volterra series operator whose only nonzero kernel is

$$\hat{h}_2(\tau_1,\tau_2)=\sum_{i,j=1}^{M}\beta_{ij}\tau_1^{i-1}\tau_2^{j-1}e^{-(\tau_1+\tau_2)}.$$

We will now show that by proper choice of p (that is, M and the β_{ij}'s) $\hat{N}_{\mathrm{lag}}$ approximates N on K. Define

$$q(\tau_1,\tau_2)=19^{-1}\left(e^{1.9\min\{\tau_1,\tau_2\}}-1\right)e^{-(\tau_1+\tau_2)/2}$$

so that $h_2(\tau_1,\tau_2)=q(\tau_1,\tau_2)\exp-(\tau_1+\tau_2)/2$. Since $q\in L^2(\mathbb{R}_+^2)$ and the span of the functions $\tau_1^i\tau_2^j\exp-(\tau_1+\tau_2)/2$ is dense in $L^2(\mathbb{R}_+^2)$, we can find M and β_{ij}'s such that

$$\left\|q(\tau_1,\tau_2)-\sum_{i,j=1}^{M}\beta_{ij}\tau_1^{i-1}\tau_2^{j-1}e^{-(\tau_1+\tau_2)/2}\right\|_2\le\frac{\epsilon}{M_1}. \tag{7.2.1}$$

Now we claim that for $u\in K$ we have $\|N_1u-\hat{N}_{\mathrm{lag}}u\|\le\epsilon$. To see this,

$$\begin{aligned}N_1u(t)-\hat{N}_{\mathrm{lag}}u(t)&=\int_0^\infty\int_0^\infty\left(h_2(\tau_1,\tau_2)-\hat{h}_2(\tau_1,\tau_2)\right)\\&\quad\cdot u(t-\tau_1)u(t-\tau_2)\,d\tau_1\,d\tau_2\\&=\int_0^\infty\int_0^\infty\left(q(\tau_1,\tau_2)-\sum_{i,j=1}^{M}\beta_{ij}\tau_1^{i-1}\tau_2^{j-1}e^{-(\tau_1+\tau_2)/2}\right)\\&\quad\cdot\left(e^{-(\tau_1+\tau_2)/2}u(t-\tau_1)u(t-\tau_2)\right)d\tau_2\,d\tau_2\end{aligned}$$

so by (7.2.1) and the Cauchy–Schwarz inequality:

$$|N_1u(t)-\hat{N}_{\mathrm{lag}}u(t)|\le\frac{\epsilon}{M_1}\|e^{-(\tau_1+\tau_2)/2}\cdot u(t-\tau_1)u(t-\tau_2)\|_2\le\epsilon$$

since

$$\|e^{-(\tau_1+\tau_2)/2}u(t-\tau_1)u(t-\tau_2)\|_2\le M_1.$$

Thus for $u\in K$ we have $\|N_1u-\hat{N}_{\mathrm{lag}}u\|\le\epsilon$.

VIII. Further Discussion of Fading Memory

We have seen that the notion of fading memory is quite useful in establishing various approximation theorems. In this section we discuss briefly two other topics which involve fading memory.

8.1. Linear Time-Invariant Operators and Fading Memory

There is a folk theorem that every LTI causal continuous operator has a convolution representation. Unfortunately this folk theorem is *false*, since there are LTI causal continuous operators which have no convolution representation. But in fact these operators are unlikely to occur in engineering; for example they do not have fading memory (see Section A3 for an example of such an operator).

However, if "continuous" is strengthened to "FM", our folk theorem becomes true.

Theorem 5 (Convolution Theorem):

(I) A: $\boldsymbol{C}(\mathbb{R})\to\boldsymbol{C}(\mathbb{R})$ is LTI FM iff A has a convolution representation

$$Au(t)=\int_0^\infty u(t-\tau)h(d\tau) \tag{8.1.1}$$

where h is a bounded measure on $\mathbb{R}_+$.

(II) A: $\boldsymbol{l}^\infty\to\boldsymbol{l}^\infty$ is LTI FM iff A has a convolution representation

$$Au(n)=\sum_0^\infty h(k)u(n-k) \tag{8.1.2}$$

where $h\in\boldsymbol{l}^1(\mathbb{Z}_+)$.

Remark: Equation (8.1.1) may be more familiar to the reader in the form

$$Au(t)=\int_0^{\infty}h(\tau)u(t-\tau)\,d\tau$$

where in this equation h is to be interpreted as a measure, e.g., may contain δ-functions.

The proof of Theorem 5 is in Section A4. Theorem 5 shows that for LTI causal systems, having a fading memory is equivalent to having a convolution representation.

8.2. Fading Memory and Unique Steady State in Dynamical Systems

The notion of fading memory is strictly an input/output property, that is, it refers only to the operator N which maps inputs into outputs; the realization of N (there need not even be one) is irrelevant. But if N does have a realization as a dynamical system, then the fading memory property is related to the *unique steady-state property* for dynamical systems [21]. In this section we elaborate this point.

Consider the system

$$\dot{x}=f(x,u) \qquad (8.2.1)$$

$$x(0)=0 \qquad (8.2.2)$$

where $x(t)\in\mathbb{R}^n$, $u\in C(\mathbb{R}_+)$, and f: $\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}^n$. Suppose f is such that (8.2.1) and (8.2.2) define an operator N: $C(\mathbb{R}_+)\to C(\mathbb{R}_+)^n$ given by $x=Nu$.

Theorem 6: Suppose N has FM on $K\subset C(\mathbb{R}_+)$, where K is closed under concatenation. Let X denote the set of all states reachable with inputs in K, that is,

$$X=\{Nu(t)|t\geq 0,\ u\in K\}.$$

Then the system (8.2.1) , (8.2.2) has a unique steady state, for inputs in K and initial conditions in X.

More precisely, let x_0, $\tilde{x}_0\in X$, and let x and $\tilde{x}$ denote the solutions (8.2.1), but with initial conditions x_0 and $\tilde{x}_0$, respectively. Then

$$\lim_{t\to\infty}\|x(t)-\tilde{x}(t)\|=0.$$

Thus the fading memory assumption implies that the state will be "asymptotically independent" of the initial condition, to use Wiener's phrase.

The proof of Theorem 6 is in Section A5. We have presented Theorem 6 only to demonstrate that there is a connection between the ideas of fading memory and unique steady state; far stronger theorems can be proved.

The conditions under which a dynamical system has a fading memory is a very important topic itself. To mention perhaps the simplest condition, if an equilibrium point is well behaved (meaning, the vector field is continuously differentiable there and the linearized system is exponentially stable and controllable) then for inputs small enough the input-to-state map will have a fading memory.

IX. A Mathematical Formulation

It is possible to generalize the results of this paper to a clean and simple mathematical form, at the cost of some engineering intuition. First, we extend our definition of fading memory to:

> N has fading memory if it is continuous with respect to the compact-open topology.

For continuous-time systems, this is the topology of uniform convergence on compact sets; for discrete-time systems, this is the topology of pointwise convergence. The definition of fading memory given in this chapter, in terms of a weighting function $w(\cdot)$, implies fading memory in this sense. Our Lemma 1 of Section IV can be generalized to:

> A closed bounded equicontinuous subset of $C(\mathbb{R}_-)$ is compact in the compact-open topology.

For the discrete-time case:

> A closed bounded subset of l^{∞} is compact in the compact-open topology.

Since in l^{∞} the compact-open topology is the weak-* topology, this last assertion is just an instance of a classic theorem of functional analysis: the closed unit ball is weak-* compact [22].

With these extended definitions, all of the approximation theorems presented still hold.

X. Conclusion

We have shown that any operator with fading memory can be approximated in a strong sense by a (finite) Volterra series operator which can be realized as a finite dimensional linear dynamical system with a polynomial readout map. For discrete-time systems, the approximating operator can simply be a *nonlinear moving-average operator*. The approximation holds over any *bounded* set of signals K; in the continuous-time case we must add a *slew-rate* limitation as well. The approximation is in the sense of *peak error*, *worst case* for all signals in K.

Since the original work of Volterra there has been much research on this topic, but none has yielded the strong approximations presented here. The reason is related to a remark in Section 2.1 concerning the difference between TI causal operators and functionals on $C(\mathbb{R}_-)$. Intuitively it would seem that this correspondence implies that an approximation of a *functional* (perhaps, via the Stone–Weierstrass theorem) should also yield an approximation of the corresponding *TI causal operator*. This is true, if the set of signals $K\subset C(\mathbb{R}_-)$ over which the approximation holds is also time-invariant, i.e., $U_tK=K$ for all $t\geq 0$. But here's the catch: TI subsets of $C(\mathbb{R}_-)$ are generally *not compact*,[8] and hence the Stone–Weierstrass theorem cannot be used to approximate the functional. Our solution to this problem was to observe that while a set such as K_-, while not compact, should "appear" compact to an operator whose memory fades with elapsed time.

We close with some remarks concerning the practical application of the material presented here. While the approximations are certainly strong enough to be useful in applications like macro-modeling of complicated systems

[8] For example, if K contains at least one compactly supported element, then it is not compact. There are TI compact subsets of $C(\mathbb{R}_-)$, for example $\{U_tf|t\geq 0\}$, where f is almost periodic.

or in universal nonlinear system identifiers, we know of no general procedure, based only on input/output measurements, by which an approximation can be found. Perhaps an adaptive scheme can be made to work in practice.

Appendix

A1. Proof of Lemma 1

We must show that

$$K_- = \{u \in C(\mathbb{R}_-) \mid |u(t)| \le M_1,$$
$$|u(s)-u(t)| \le M_2(s-t) \quad \text{for} \quad t \le s \le 0\}$$

is compact with the weighted norm $\|\cdot\|_w$ in $C(\mathbb{R}_-)$. Let u_n, $n=1,2,\cdots$ be any sequence in K_-. We will find a $u_0 \in K_-$ and a subsequence of $\{u_n\}$ converging in the $\|\cdot\|_w$ norm to u_0, which will establish Lemma 1.

Let $K_-[-n,0]$ denote K_- restricted to $[-n,0]$, that is

$$K_-[-n,0] \triangleq \{u \in C[-n,0] \mid |u(t)| \le M_1,$$
$$|u(s)-u(t)| \le M_2(s-t)$$
$$\text{for } -n \le t \le s \le 0\}.$$

For each n, $K_-[-n,0]$ is uniformly bounded (by M_1) and equicontinuous (by the slew-limit M_2), hence compact in $C[-n,0]$ by the Arzela–Ascoli theorem (see e.g. Dieudonne [15]). Since $K_-[-1,0]$ is compact in $C[-1,0]$, we can find a $u_0^{(1)} \in K_-[-1,0]$ and an infinite subset $\mathbb{N}_1 \subset \mathbb{N}$ such that

$$\sup_{-1 \le t \le 0} |u_n(t) - u_0^{(1)}(t)| \to 0 \quad \text{as} \quad n \to \infty, \qquad n \in \mathbb{N}_1.$$

Viewing $\{u_n \mid n \in \mathbb{N}_1\}$ as a sequence in $K_-[-2,0]$, we conclude that there is a $u_0^{(2)} \in K_-[-2,0]$ and an infinite subset $\mathbb{N}_2 \subset \mathbb{N}_1$ such that

$$\sup_{-2 \le t \le 0} |u_n(t) - u_0^{(2)}(t)| \to 0 \quad \text{as} \quad n \to \infty, \qquad n \in \mathbb{N}_2.$$

Clearly $u_0^{(2)}$ extends $u_0^{(1)}$, that is, $u_0^{(1)}(t) = u_0^{(2)}(t)$ for $-1 \le t \le 0$.

Continuing in this way we find a $u_0 \in K_-$ and a sequence of decreasing infinite subsets $\mathbb{N} \supset \mathbb{N}_1 \supset \cdots$ such that for each k

$$\sup_{-k \le t \le 0} |u_n(t) - u_0(t)| \to 0 \quad \text{as} \quad n \to \infty, \qquad n \in \mathbb{N}_k. \tag{A1.1}$$

We now choose any increasing subsequence n_k such that $n_k \in \mathbb{N}_k$. Then from (A1.1) we have for each k_0

$$\sup_{-k_0 \le t \le 0} |u_{n_k}(t) - u_0(t)| \to 0 \quad \text{as} \quad k \to \infty$$

that is, the sequence u_{n_k} converges to u_0 uniformly on compact subsets.

Now we claim that u_{n_k} converges to u_0 in the weighted norm, that is, $\lim_{k\to\infty} \|u_{n_k} - u_0\|_w = 0$. To prove our claim, let $\epsilon > 0$. Since $w(t) \to 0$ as $t \to \infty$, we can find $k_0 \in \mathbb{N}$ such that $w(k_0) < \epsilon / 2M_1$; since u_{n_k}, $u_0 \in K_-$ we have

$$\sup_{t \le -k_0} |u_{n_k}(t) - u_0(t)| w(-t) \le 2M_1 w(k_0) < \epsilon. \tag{A1.2}$$

Now find k_1 such that

$$\sup_{-k_0 \le t \le 0} |u_{n_k}(t) - u_0(t)| \le \epsilon, \qquad \text{for} \quad k \ge k_1. \tag{A1.3}$$

From (A1.2), (A1.3), and $w(t) \le 1$ we conclude

$$\|u_{n_k} - u_0\|_w \le \epsilon, \qquad \text{for} \quad k \ge k_1$$

which concludes the proof of Lemma 1.

A2. Proof of NLMA Approximation Theorem

We start with the analog of Lemma 1:

Lemma A1:

$$K_- \triangleq \{u \in l^\infty(\mathbb{Z}_-) \mid \|u\| \le M_1\}$$

is compact with the weighted norm $\|\cdot\|_w$ given by

$$\|u\|_w \triangleq \sup_{k \le 0} |u(k)| w(-k).$$

Proof: We give an abbreviated proof since it is similar to, and in fact simpler than, the proof of Lemma 1 given in Section A1.

Let $\{u^{(n)}\}$ be a sequence in K_-. Since $|u^{(n)}(0)| \le M_1$, find a subsequence along which $u^{(n)}(0)$ converges; let us call the limit $u^{(0)}(0)$. Now find a subsequence of this subsequence along which $u^{(n)}(-1)$ converges; call this limit $u^{(0)}(-1)$.

Just as in proof of Lemma 1 we continue this process, defining the element $u^{(0)} \in K_-$ as we go. Take a diagonal subsequence n_k; $u^{(n_k)}$ converges pointwise to $u^{(0)}$ as $k \to \infty$, and exactly as in Lemma 1 we can show

$$\|u^{(n_k)} - u^{(0)}\|_w \to 0 \quad \text{as} \quad k \to \infty$$

which proves that K_- is compact.

Now consider the set of functionals

$$G \triangleq \{G_0, G_1, \cdots\}$$

where $G_k u \triangleq u(-k)$, that is, G_k is the functional associated with the k-delay operator U_k (transfer function z^{-k}).

It is easy to verify that the G_k's are continuous with respect to the weighted norm $\|\cdot\|_w$ and that G separates points in $l^\infty(\mathbb{Z}_-)$. Applying the Stone–Weierstrass theorem as in Theorem 1 yields an approximation by a NLMA operator.

A3. Causal Continuous LTI Operator with no Convolution Representation

Here is a brief description of one such operator (see Kantorovich [23, p. 58] for details). It is possible to find a linear functional LIM: $l^\infty \to \mathbb{R}$ such that

$$|LIMu| \leqslant \|u\|$$

and if $\lim_{k\to-\infty} u(k)$ exists, then $LIMu = \lim_{k\to-\infty} u(k)$. Thus LIM assigns a "pseudo-limit" $LIMu$ to every element of l^∞ (the vast majority of which do not converge as $k \to -\infty$). Consider the operator A: $l^\infty \to l^\infty$ given by

$$Au(n) = LIMu.$$

Thus for every $u \in l^\infty$, Au is the constant sequence $LIMu$.

A is LTI causal continuous, but has no convolution representation since its response to a unit sample is zero, and yet it is not the zero operator. Note that A is a LTI causal operator which does not have fading memory. Of course, an operator like A is not likely to occur in engineering.

A4. Proof of Theorem 5 (Convolution Theorem)

We will prove (II), and then indicate some of the changes necessary to prove the continuous-time version (I).

First suppose $Au = h*u$ where $h \in l^1(\mathbb{Z}_+)$. We will show that A has fading memory (that it is LTI causal is clear). Consider the weighting function

$$w(n) \triangleq \|h\|_1^{-1/2}\left\{\sum_{k=n}^{\infty}|h(k)|\right\}^{1/2}. \quad \text{(A4.1)}$$

We claim that A has a w-fading memory. As in (5.1.2) we need only establish

$$S \triangleq \sum_{n=0}^{\infty}|h(n)|w(n)^{-1} < \infty.$$

In fact $S \leqslant 2$, which we now prove. Define

$$\theta(n) \triangleq \sum_{k=n}^{\infty}|h(k)|$$

so that

$$\begin{aligned} S &= \theta(0)^{-1/2}\sum_{n=0}^{\infty}\frac{\theta(n)-\theta(n+1)}{\theta(n)^{1/2}} \\ &= 1+\theta(0)^{-1/2}\sum_{n=0}^{\infty}\theta(n+1) \\ &\quad \cdot\left(\theta(n+1)^{-1/2}-\theta(n)^{-1/2}\right). \end{aligned} \quad \text{(A4.2)}$$

Since $0 \leqslant \theta(n+1) \leqslant \theta(n)$ we have

$$\theta(n+1)\left(\theta(n+1)^{-1/2}-\theta(n)^{-1/2}\right) \leqslant \theta(n)^{-1/2}-\theta(n+1)^{-1/2} \quad \text{(A4.3)}$$

(the ratio of the two is $\sqrt{\theta(n+1)/\theta(n)} \leqslant 1$). From (A4.2) and (A4.3)

$$S \leqslant 1+\theta(0)^{-1/2}\sum_{n=0}^{\infty}\left(\theta(n)^{-1/2}-\theta(n+1)^{-1/2}\right) = 2$$

which proves that A has a w-fading memory.

Remark: If h happens to be exponentially decaying then we may use the weight $w(n) = (1+n)^{-1}$, but of course not all $h \in l^1(\mathbb{Z}_-)$ are exponentially decaying, and then the more complicated weight (A4.1) is necessary.

Now we prove the converse. Let A be any LTI operator with, say, a w-fading memory.[9] Let h be the response of A to a unit sample, i.e., $h(n) \triangleq Ae(n)$ where $e(n) = \delta_{n0}$.

We will show (1) $h \in l^1(\mathbb{Z}_+)$ (at the moment we know only $h \in l^\infty(\mathbb{Z}_+)$), and (2) $Au = h*u$ for all $u \in l^\infty$.

Let F be the functional associated with A via (2.1.1). Using linearity and FM we conclude there is an $M < \infty$ such that for all $u \in l^\infty(\mathbb{Z}_-)$.

$$|Fu| \leqslant M\|u\|_w. \quad \text{(A4.4)}$$

[9]This w has nothing to do with the w defined in (A4.1).

Now for any $u: \mathbb{Z} \to \mathbb{R}$ define

$$u_N(k) = \begin{cases} u(k), & -N \leqslant k \leqslant 0 \\ 0, & k < -N. \end{cases}$$

We now use a standard argument. From time-invariance and linearity we have

$$Fu_N = \sum_{k=0}^{N} h(k)u_n(-k) = \sum_{k=0}^{N} h(k)u(-k). \quad \text{(A4.5)}$$

Consider $u(k) \triangleq w(k)^{-1}\operatorname{sign} h(k)$; from (A4.4) and (A4.5) we conclude

$$\sum_{k=0}^{N} w(k)^{-1}|h(k)| \leqslant M$$

for all N and thus $hw^{-1} \in l^1(\mathbb{Z}_+)$, which implies $h \in l^1(\mathbb{Z}_+)$.

Now (2): for any $u \in l^\infty(\mathbb{Z}_-)$ we have from (A4.4)

$$\begin{aligned} |Fu - Fu_N| &\leqslant M\|u-u_N\|_w \\ &\leqslant Mw(N+1) \to 0 \quad \text{as } N \to \infty. \end{aligned}$$

Thus (noting that $h(\cdot)u(-\cdot) \in l^1(\mathbb{Z}_+)$)

$$Fu = \lim_{N\to\infty} Fu_N = \sum_{k=0}^{\infty} h(k)u(-k)$$

which finishes our proof.

To show that a LTI operator $A: C(\mathbb{R}) \to C(\mathbb{R})$ which has a convolution representation (8.1.1) has a fading memory, we use the weight

$$w(t) \triangleq \left\{\int |h(d\tau)|\right\}^{-1/2}\left\{\int_t^\infty |h(d\tau)|\right\}^{1/2}.$$

Then by a change of variables we have

$$\int_0^\infty |h(dt)|w(t)^{-1} = 2$$

so that A has a w-fading memory.

To prove that a LTI FM operator has a convolution representation is technically more involved since we cannot directly apply an impulse input $\delta(t)$. But the idea is the same.

A5. Proof of Theorem 6

Assume the hypotheses of Theorem 6. Since x_0 and $\tilde{x}_0$ are reachable with inputs in K, let $T \in \mathbb{R}$ and $u_s, \tilde{u}_s \in K$ be such that

$$Nu_s(T) = x_0 \qquad N\tilde{u}_s(T) = \tilde{x}_0.$$

Thus u_s and $\tilde{u}_s$ steer x from 0 to x_0 and $\tilde{x}_0$, respectively, over the interval $[0, T]$.

Define

$$v(t) \triangleq \begin{cases} u_s(t), & 0 \leqslant t \leqslant T \\ u(t+T), & t > T \end{cases}$$

and similarly,

$$\tilde{v}(t) \triangleq \begin{cases} \tilde{u}_s(t), & 0 \leqslant t \leqslant T \\ \tilde{u}(t+T), & t > T. \end{cases}$$

Since K is closed under concatenation, $v, \tilde{v} \in K$. In fact $x(t) = Nv(t+T)$ and $\tilde{x}(t) = N\tilde{v}(t+T)$, so it will suffice to prove $v(t) \to \tilde{v}(t)$ as $t \to \infty$.

Let $\epsilon > 0$. Using our fading memory assumption, there is a $\delta > 0$ such that for all $t \in \mathbb{R}$

$$\sup_{0 \leqslant \tau \leqslant t} |v(t) - \tilde{v}(t)| w(t-\tau) < \delta \to \|Nv(t) - N\tilde{v}(t)\| < \epsilon. \tag{A5.1}$$

Since $v(t) = \tilde{v}(t)$ for $t \geqslant T$,

$$\sup_{0 \leqslant \tau \leqslant t} |v(t) - \tilde{v}(t)| w(t-\tau) \leqslant 2M_1 w(t-T).$$

Using $w(t) \to 0$ as $t \to \infty$, find $T_0 \geqslant T$ such that $w(T_0) < \delta/(3M_1)$. Then for $t \geqslant T_0$ the right-hand side of (A5.1) is satisfied and hence

$$\|x(t) - \tilde{x}(t)\| < \epsilon, \qquad \text{for } t \geqslant T_0$$

which proves Theorem 6.

Acknowledgment

The authors would like to thank Professors C. A. Desoer, S. S. Sastry, and Dr. A. Heunnis for checking a draft of this paper. Prof. M. Fliess suggested the note on approximation by bilinear systems (§5.3). S. Boyd gratefully acknowledges the support of the Fannie and John Hertz foundation.

References

[1] I. W. Sandberg, "Series expansions for nonlinear systems" *Circuits, Systems, and Signal Processing*, vol. 2, no. 1, pp. 77–87, 1983.

[2] S. Boyd, L. O. Chua, and C. A. Desoer, "Analytical foundations of Volterra series," *IMA J. Inform. Contr.*, vol. 1, pp. 243–282, 1984.

[3] V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations*. New York: Dover, 1959.

[4] M. Frechet, "Sur les fonctionelles continues," *Ann. l'Ecole Normale Sup.*, vol. 27, 1910.

[5] N. Wiener, *Nonlinear Problems in Random Theory*, MIT Press Cambridge, MA, 1958.

[6] D. George, "Continuous nonlinear systems," MIT RLE Tech. Rep., no. 355, 1959.

[7] M. B. Brilliant, "Theory of the analysis of nonlinear systems," Rep. RLE-345, MIT, Mar. 1958.

[8] P. Prenter, "A Weierstrass theorem for real separable Hilbert spaces," *J. Approx. Theory*, vol. 3, pp. 341–351, 1970.

[9] W. Porter and T. Clark, "Causality structure and the Weierstrass theorem," *J. Math. Anal. Appl.*, vol. 52, pp. 351–363, 1975.

[10] W. Porter, "Approximation by Bernstein systems," *Math. Syst. Theory*, vol. 11, pp. 259–274, 1978.

[11] W. Root, "On the modeling of systems for identification, Part I: e-Representations of classes of systems," *SIAM J. Contr.*, vol. 13, pp. 927–944, 1975.

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

[13] J. F. Barrett, "The use of functionals in the analysis of nonlinear physical systems," *Int. J. Electron. Control*, vol. 15, pp. 567–615, 1963.

[14] W. L. Root, "On system measurement and identification," *System Theory*, pp. 133–157.

[15] J. Dieudonne, *Foundations of Modern Analysis*. New York: Academic, 1969.

[16] R. J. P. deFigueiredo and T. A. Dwyer, "A best approximation framework and implementation for simulation of large scale nonlinear systems," *IEEE Trans. Circuits Syst.*, vol. CAS-27, pp. 1005–1014, Nov. 1980.

[17] M. Fliess, "Fonctionnelles causales non lineares et indeterminees non commutatives," *Bull. Soc. Math. France*, vol. 109, pp. 3–41, 1981.

[18] H. J. Sussmann, "Semigroup representations, bilinear approximation of input-output maps, and generalized inputs," *Math. Syst. Theory*, no. 131, pp. 172–191, Springer-Verlag, Berlin, 1976.

[19] R. W. Brockett, "Convergence of Volterra series on infinite intervals and bilinear approximations," in *Nonlinear Systems and Applications*, Ed. V. Lakshmikantham, Academic, NY, 1977.

[20] H. Dang Van Mien and D. Normand-Cyrot, "Nonlinear state-affine identification Methods: Applications to electrical power plants," *Automatica*, vol. 20, no. 2, pp. 175–188, 1984.

[21] L. O. Chua and D. N. Green, "A qualitative analysis of the behavior of dynamic nonlinear networks: steady-state solutions of nonautonomous networks," *IEEE Trans. Circuits Syst.*, vol. CAS-23, pp. 530–550, Sept. 1976.

[22] W. Rudin, *Functional Analysis*. New York: McGraw-Hill, 1974.

[23] L. V. Kantorovich and G. P. Akilov, *Functional Analysis*, Pergamon, Oxford, England, 1982.

[24] V. Constanza, B. Dickenson, and E. F. Johnson, "Universal approximations of discrete-time control systems over finite time," *IEEE Trans. Automat. Control*, vol. AC-28, pp. 439–452, Apr. 1983.

[25] P. Gallman and K. S. Narendra, "Representations of nonlinear systems via the Stone–Weierstrass theorem," *Automatica*, vol. 12, pp. 619–622, 1976.

[26] J. Blackman, "The representation of nonlinear networks," *Syracuse Univ. Research Inst. Rep.*, Aug. 1960.

[27] S. Boyd, "Volterra Series: Engineering Fundamentals," Ph.D. dissertation, Univ. of California, Berkeley, 1985.

Stephen Boyd received the A. B. degree in mathematics, summa cum laude, from Harvard University, Cambridge, MA, in 1980. He then entered the mathematics Ph.D. program at University of California, Berkeley, and in 1981 transferred to electrical engineering. In graduate school he has held the following fellowships: Univ. of California Regents Fellowship (Mathematics), National Science Foundation Graduate Fellowship (Mathematics), Irving and Lucille Smith Fellowship (EECS) and Fannie and John Hertz Fellowship (EECS). In 1985 he joined the Information Systems Laboratory, Electrical Engineering Department, Stanford University, Stanford, CA. He is interested in nonlinear systems and control theory.

Leon O. Chua (S'60–M'62–SM'70–F'74), for a photograph and biography please see page 45 of the January 1985 issue of this TRANSACTIONS.