IMA Journal of Mathematical Control & Information (1984) **1**, 243–282

Analytical Foundations of Volterra Series*

STEPHEN BOYD, L. O. CHUA AND C. A. DESOER
Department of Electrical Engineering and Computer Sciences, and the Electronics Research Laboratory, University of California, Berkeley 94720

[Received 21 March 1984 and in revised form 30 May 1984]

In this paper we carefully study the *analysis* involved with Volterra series. We address system-theoretic issues ranging from bounds on the gain and incremental gain of Volterra series operators to the existence of Volterra series operator inverses, and mathematical topics such as the relation between Volterra series operators and Taylor series. The proofs are complete, and use only the basic facts of analysis.

We prove a general Steady-state theorem for Volterra series operators, and then establish a general formula for the spectrum of the output of a Volterra series operator in terms of the spectrum of a periodic input.

This paper is meant to complement recent work on Volterra series expansions for dynamical systems.

1. Introduction

A VOLTERRA SERIES OPERATOR with *kernels* h_n is one of the form

$$Nu(t) = \sum_{n=1}^{\infty} y_n(t), \tag{1.1a}$$

$$y_n(t) = \int \cdots \int h_n(\tau_1, \tau_2, \ldots, \tau_n) u(t-\tau_1) u(t-\tau_2) \cdots u(t-\tau_n)\, d\tau_1\, d\tau_2 \cdots d\tau_n, \tag{1.1b}$$

and is a generalization of the convolution description of linear time-invariant (LTI) operators to time-invariant (TI) nonlinear operators. These operators are important because many TI nonlinear operators occurring in engineering either have this form or can be approximated, in some sense, by operators of this form (Boyd & Chua, 1984). Volterra series have been the object of much recent study. The focus has primarily been on proofs that the input/output (I/O) operators of dynamical systems, and various generalizations, have a Volterra series representation, and the relationship between the Volterra kernels and the geometry of the dynamical system. Brockett (1976) established the existence of a Volterra series representation for many dynamical systems, and Sandberg (1983*a*) has recently extended this to include a very wide class of systems, e.g. delay–differential systems. Fliess, Lamnabhi, & Lamnabhi-Lagarrique have found a simple and

* Research supported in part by the Office of Naval Research under contract N00014-76-C-0572, the National Science Foundation under grants ECS 80-20-640 and ECS 81-19-763, and the Fannie and John Hertz Foundation.

elegant formula for the kernels of a dynamical system in terms of various Lie derivatives.

In contrast our focus is on the *analysis* involved with Volterra series. We first carefully address the basic issues of the formal Volterra series (1.1) above: what are the kernels (functions, distributions, . . .?) and when do the integrals and sums in (1.1) make sense? In the remainder of Section 2 we examine the elementary properties of Volterra series operators, both system-theoretic (e.g. bounds on their gain and incremental gain) and mathematical (e.g. their relation to Taylor series).

In Section 3 we use the methods of Section 2 to prove some well-known formulas for the kernels of various 'system interconnections'. We give an elementary and complete proof of the Inversion theorem for Volterra series, and work through an illustrative example.

In Section 4 we explore some frequency domain topics. We start by proving the Steady-state theorem for Volterra series operators. We then establish the validity of a general formula for the spectrum of the output in terms of the spectrum of a periodic input.

In the appendix we present more advanced (and esoteric) material: Volterra-like series, the incremental gain theorem for L^p, Taylor series which are not Volterra series, conditions under which the frequency domain formula of Section 4 holds, and almost periodic inputs.

The results we present range from 'well-known' (e.g. the Uniqueness theorem) to new (e.g. the material in Section 4 and the appendix). Some of the material overlaps or extends the work of other researchers, notably Sandberg's (1983*b*, 1984) work on Volterra-like series and almost periodic inputs respectively.

In order to keep the paper interesting and accessible to a wide audience, we have used only the basic tools of real analysis, in a few places developing some necessary background material. We do *not* present the results in their full generality: we have limited the scope of the paper to single-input-, single-output-(SISO-) stable TI Volterra series in order to do a more thorough job on this important case. Extensions to other cases will be presented in a future paper.

The references we give are not meant to be complete but only representative. More complete bibliographies can be found in our references, for example Sandberg (1983*a*) or Fliess *et al.* (1983). Rugh (1981) contains a very complete annotated bibliography up to 1980.

2. Formulation

2.1 *What are the Kernels?*

In most treatments the kernels $h_n(\tau_1, \ldots, \tau_n)$ in equation (1.1) are interpreted as *functions* from $\mathbb{R}^n$ to $\mathbb{R}$. Unfortunately this interpretation rules out some operators common in engineering. We start with two examples:

EXAMPLE 1

$$\dot{x} = -x + u$$
$$y = x^2$$

and $x(0)=0$. Then for $t \geq 0$

$$y(t)=\left(\int_0^t e^{-\tau}u(t-\tau)\,d\tau\right)^2=\iint 1_H(\tau_1)1_H(\tau_2)e^{-(\tau_1+\tau_2)}u(t-\tau_1)u(t-\tau_2)\,d\tau_1\,d\tau_2$$

where

$$1_H(t):=\begin{cases}1 & (t\geq 0),\\ 0 & (t<0),\end{cases}$$

is Heaviside's unit function. Hence this operator has a Volterra series description with just one nonzero kernel,

$$h_2(\tau_1,\tau_2)=1_H(\tau_1)1_H(\tau_2)e^{-(\tau_1+\tau_2)}.$$

This kernel h_2 is an ordinary function $:\mathbb{R}^2\to\mathbb{R}$.

EXAMPLE 2

$$\dot{x}=-x+u^2,\qquad y=x,\qquad \text{and } x(0)=0.$$

Here

$$y(t)=\int_0^t e^{-\tau}u(t-\tau)^2\,d\tau$$

$$=\iint 1_H(\tau_1)1_H(\tau_2)e^{-\tau_1}\,\delta(\tau_1-\tau_2)u(t-\tau_1)u(t-\tau_2)\,d\tau_1\,d\tau_2$$

where δ is Dirac's 'derivative' of 1_H if you will condone the notation. So here the kernel h_2 is not a *function* as it was in Example 1 but a *measure* supported on the line $\tau_1=\tau_2$, informally given by

$$h_2(\tau_1,\tau_2)=1_H(\tau_1)1_H(\tau_2)\,\delta(\tau_1-\tau_2)e^{-\tau_1}$$

These examples are typical: in general the Volterra series of dynamical systems with the vector field *affine* in the input u (e.g. in *bilinear systems*) have kernels which are ordinary functions whereas in other cases more general measures may be necessary (Brockett, 1976; d'Alessandro, Isideri & Ruberti, 1974; Lesiak & Krener, 1978; Sandberg, 1983*b*). In the latter case Sandberg has called the series 'Volterra-like': Appendix A1 contains an in-depth discussion of Volterra-like series.

A less exotic but widely occurring nonlinear operator whose description requires kernels which are measures is the *memoryless operator*

$$y(t)=f[u(t)]$$

where $f:\mathbb{R}\to\mathbb{R}$ is analytic near 0.

We will allow our kernels to be measures. We will see that the analysis is no harder, and the resulting theory then includes *all* the examples above.

2.2 *When the Series Converges*

Recall that the ordinary power series $g(x)=\sum_{n=0}^{\infty}a_n x^n$ converges absolutely for $|x|<\rho$, where the radius of convergence is given by $\rho=(\limsup_{n\to\infty}|a_n|^{1/n})^{-1}$.

Similarly a radius of convergence ρ can be associated with a formal Volterra series

$$Nu(t)=y(t)=\sum_{n=1}^{\infty}\int\cdots\int h_n(\tau_1,\ldots,\tau_n)u(t-\tau_1)\cdots u(t-\tau_n)\,d\tau_1\cdots d\tau_n \quad (2.1)$$

such that the series will converge for input signals with $|u(t)|<\rho$.

More precisely, let $\mathcal{B}^n$ be the bounded measures on $\mathbb{R}^n_{\geq 0}$,† with $\|\mu\|=\int d|\mu|$. For convenience we will write elements of $\mathcal{B}^n$ as if they were absolutely continuous ("Physicists' style"), e.g. $h_2(\tau_1,\tau_2)=\delta(\tau_1-\tau_2)e^{-\tau_1}$. For signals $\|\cdots\|$ will denote the ∞-norm, i.e. $\|u\|=\|u\|_\infty$.‡

Definition By a *Volterra series operator* we will henceforth mean an operator given by equation (2.1) above and satisfying assumptions

(A1) $h_n\in\mathcal{B}^n$ and

(A2) $\limsup_{n\to\infty}\|h_n\|^{1/n}<\infty$, that is, $[\|h_n\|^{1/n}]_{n=1}^{\infty}$ is bounded.

Our first task is to determine for which u equation (2.1) makes sense.

Definition If N is a Volterra series operator with kernels h_n, we define the *gain bound function* of N to be, for $x\geq 0$, $f(x):=\sum_{n=1}^{\infty}\|h_n\|\,x^n$ (with extended values, that is, $f(x)$ may be ∞). The *radius of convergence* of N is defined by $\rho=\operatorname{rad}N:=(\limsup_{n\to\infty}\|h_n\|^{1/n})^{-1}$.

Assumption (A2) implies that $\rho>0$ and that the gain bound function f is analytic at 0, with normal radius of convergence ρ. Since all the terms in the series for f are positive ρ is also given by $\rho=\inf\{x: f(x)=\infty\}$, a formula which will be useful in Section 3. We can now say when (2.1) makes sense.

Theorem 2.2.1 (Gain Bound Theorem) *Suppose N is a Volterra series operator with kernels h_n, gain bound function f, and radius of convergence ρ. Then*

(i) *the integrals and sum in equation* (2.1) *above converge absolutely for inputs with $\|u\|<\rho$, that is, in B_ρ, the zero-centred ball of radius ρ in* L^∞.

(ii) *N satisfies $\|Nu\|\leq f(\|u\|)$ and consequently N maps B_ρ into* L^∞.

(ii) is partial justification for naming f the gain bound function, we will soon see more. Theorem 2.2.1 is well known in various forms (Rugh, 1981; Barrett 1963; Brockett 1976, 1977; Christensen, 1968; Lesiak & Krener, 1978; Rao, 1970; Sandberg, 1983).

Proof.

$$\int\cdots\int|h_n(\tau_1,\ldots,\tau_n)u(t-\tau_1)\cdots u(t-\tau_n)|\,d\tau_1\cdots d\tau_n\leq\|h_n\|\cdot\|u\|^n.$$

† We thus consider only *causal* operators, but in fact all of the following holds for kernels which are bounded measures on $\mathbb{R}^n$.

‡ An excellent reference on bounded measures and these norms (and analysis in general) is Rudin's (1974) book.

In particular, the integrals make sense. If $\|u\|<\rho$, then

$$\sum_{n=1}^{\infty}\left|\int\cdots\int h_n(\tau_1,\ldots,\tau_n)u(t-\tau_1)\cdots u(t-\tau_n)\,d\tau_1\cdots d\tau_n\right|$$
$$\leqslant \sum_{n=1}^{\infty}\int\cdots\int|h_n(\tau_1,\ldots,\tau_n)u(t-\tau_1)\cdots u(t-\tau_n)|\,d\tau_1\cdots d\tau_n$$
$$\leqslant \sum_{n=1}^{\infty}\|h_n\|\cdot\|u\|^n=f(\|u\|)<\infty$$

which establishes absolute convergence of the series and the gain bound in (ii). □

For convenience we adopt the notational convention that throughout this paper N will denote a Volterra series operator with kernels h_n, gain bound function f, and radius of convergence ρ.

The Gain Bound theorem has many simple applications. For example, the tail of the gain bound function gives a bound on the truncation error for a Volterra series.

COROLLARY 2.2.2 (Error bound for truncated Volterra series) *The truncated Volterra series operator defined by*

$$N^{(k)}(t):=\sum_{n=1}^{k}\int\cdots\int h_n(\tau_1,\ldots,\tau_n)u(t-\tau_1)\cdots u(t-\tau_n)\,d\tau_1\cdots d\tau_n$$

satisfies

$$\|Nu-N^{(k)}u\|\leqslant\sum_{n=k+1}^{\infty}\|h_n\|\cdot\|u\|^n$$

which is $o(\|u\|^k)$.

2.3 *Elementary Properties: Continuity*

We will now show that N is continuous on B_ρ and Lipschitz continuous on any B_r, $r<\rho$.

LEMMA 2.3.1 *Suppose* $\|u\|+\|v\|<\rho$. *Then*

$$\|N(u+v)-N(u)\|\leqslant f(\|u\|+\|v\|)-f(\|u\|)\leqslant f'(\|u\|+\|v\|)\,\|v\|.$$

Proof. Assume $\|u\|+\|v\|<\rho$. Then $\|u+v\|<\rho$ so $N(u+v)$ makes sense and

$$|N(u+v)(t)-N(u)(t)| \tag{2.2a}$$
$$\leqslant\sum_{n=1}^{\infty}\int\cdots\int\left|h_n(\tau_1,\ldots,\tau_n)\left(\prod_{i=1}^{n}(u+v)(t-\tau_i)-\prod_{i=1}^{n}u(t-\tau_i)\right)\right|d\tau_1\cdots d\tau_n \tag{2.2b}$$
$$\leqslant\sum_{n=1}^{\infty}\|h_n\|\sum_{j=0}^{n-1}\binom{n}{j}\|u\|^j\,\|v\|^{n-j} \tag{2.2c}$$
$$=\sum_{n=1}^{\infty}\|h_n\|\,[(\|u\|+\|v\|)^n-\|u\|^n] \tag{2.2d}$$
$$=f(\|u\|+\|v\|)-f(\|u\|). \tag{2.2e}$$

This technique will recur so careful explanation is worthwhile. In (2.2b) the first product, when expanded, has 2^n terms; the second product is precisely the first term in the expansion. Replacing the remaining $2^n - 1$ terms by their norms and integrating yields (2.2c).

The final inequality in Lemma 2.3.1 follows from the mean-value theorem, since

$$f(\|u\|+\|v\|)-f(\|u\|)=f'(\zeta)\,\|v\|$$

where $\|u\| \leq \zeta \leq \|u\|+\|v\|$ and f' is increasing. Thus f' can be interpreted as an *incremental gain bound function* for N. □

THEOREM 2.3.2 (Incremental Gain theorem) *Let B_r be the zero-centred ball of radius r in L^∞, and suppose $r<\rho$. Then*

(i) *$N: B_r \to B_{f(r)}$ is Lipschitz continuous,*

(ii) *$N: B_\rho \to L^\infty$ is continuous.*

Proof. Suppose u and v are in B_r. From the Gain Bound theorem

$$\|Nu-Nv\| \leq f(\|u\|)+f(\|v\|). \tag{2.3}$$

We claim that

$$\|Nu-Nv\| \leq f(\|u-v\|+\|v\|)-f(\|v\|) \tag{2.4}$$

For $\|u-v\|+\|v\|<\rho$ (2.4) is simply Lemma 2.3.1; for $\|u-v\|+\|v\| \geq \rho$ (2.4) is true since its right-hand side is ∞. From (2.3) and (2.4) we deduce

$$\|Nu-Nv\| \leq \min\{f(\|u-v\|+\|v\|)-f(\|v\|),\, f(\|u\|)+f(\|v\|)\}$$
$$\leq \|u-v\| \min\left\{\frac{f(r+\|u-v\|)-f(r)}{\|u-v\|}, \frac{2f(r)}{\|u-v\|}\right\} \leq K\,\|u-v\|$$

where K is the supremum of the expression $\min\{\cdots\}$ for $0 \leq \|u-v\| \leq 2r$ and is finite. (K is in fact $2f(r)/\{f^{-1}[3f(r)]-r\}$: see Fig. 1). This establishes (i); since (i) is true for any $r<\rho$ (ii) follows. □

We will soon see that N is much more than merely continuous; for example, N has Frechet derivatives of all orders on B_ρ. But before moving on, we present an extension of the last theorem which will be important in Section 4.

Recall that for linear systems $y = h_1 \star u$ we have the result $\|y\|_p \leq \|h_1\| \cdot \|u\|_p$, for $1 \leq p \leq \infty$ (Desoer & Vidyasagar, 1975). It turns out that when *properly reformulated* the Gain Bound theorem and the Incremental Gain theorem are also true with general p-norms. First some warnings for $p<\infty$: a Volterra series operator need not be defined on any open subset of L^p, e.g. $Nu(t)=u(t)/[1-u(t)]$, and even when it is, it need not map L^p back into L^p, e.g. $Nu(t)=u(t)^2$. For more details and discussion we refer the reader to Appendix A2.

THEOREM 2.3.3 (Gain Bound theorem for L^p) *For $1 \leq p \leq \infty$*

$$\|Nu\|_p \leq \|u\|_p \frac{f(\|u\|)}{\|u\|}$$

(unmarked norms are ∞-norms).

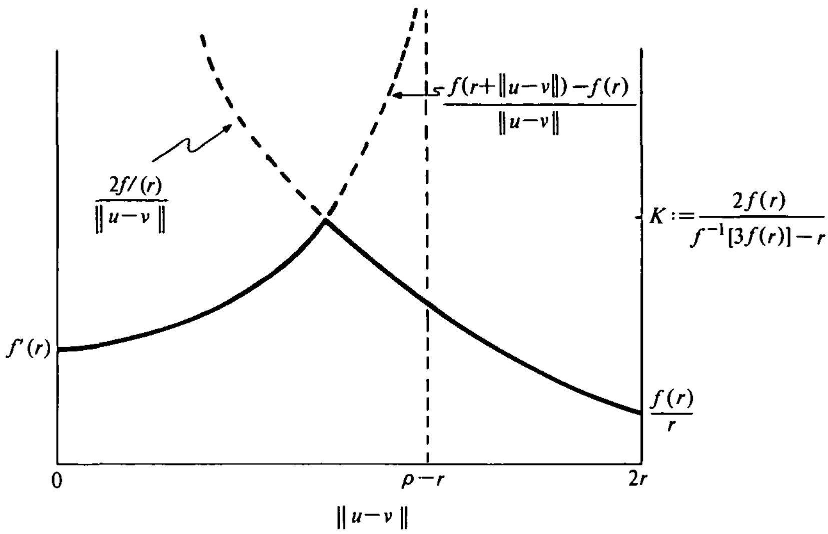

FIG. 1.

Even though our next theorem is stronger, we give the proof here to demonstrate the basic argument.

Proof.

$$|y_n(t)| \leq \int \cdots \int |h_n(\tau_1, \ldots, \tau_n)u(t-\tau_1) \cdots u(t-\tau_n)| \, d\tau_1 \cdots d\tau_n$$

$$\leq \|u\|^{n-1} \int \left[\int \cdots \int |h_n(\tau_1, \ldots, \tau_n)| \, d\tau_2 \cdots d\tau_n \right] |u(t-\tau_1)| \, d\tau_1. \qquad (2.5)$$

Now the bracketed expression in (2.5) is a *measure* in τ_1 with norm $\|h_n\|$, hence using the result for *linear systems* cited above we have (Desoer & Vidyasagar, 1975)

$$\|y_n\|_p \leq \|u\|^{n-1} \, \|h_n\| \, \|u\|_p.$$

Thus

$$\|Nu\|_p \leq \sum_{n=1}^{\infty} \|y_n\|_p \leq \|u\|_p \sum_{n=1}^{\infty} \|h_n\| \, \|u\|^{n-1} = \|u\|_p \frac{f(\|u\|)}{\|u\|}$$

which establishes Theorem 2.3.3. □

LEMMA 2.3.4

$$\|(N(u+v) - Nu)\|_p \leq \|v\|_p \frac{f(\|u\| + \|v\|) - f(\|u\|)}{\|v\|} \leq \|v\|_p f'(\|u\| + \|v\|).$$

The proof combines the proof above with the proof of the Incremental Gain theorem and is in Appendix A2.

THEOREM 2.3.5 (Incremental Gain theorem for L^p) *Let B_r be the zero-centred ball of radius r in* L^∞, *with $r<\rho$. Then there is a K such that*

$$\|Nu - Nv\|_p \leq K\,\|u - v\|_p$$

The proof is identical to that of Theorem 2.3.2 and so is omitted.

2.4 *Multilinear and Polynomial Mappings*

This section contains mostly background material for Section 2.5 and may be omitted by those familiar with the topic. There are many good references on this material, both in mathematics (Balakrishnan, 1976; Dieudonne, 1969) and engineering (Halme, Orava, & Blomberg, 1971; Sontag, 1979).

Note that the nth term y_n in a Volterra series is homogeneous of degree n in the input u. Indeed much more is true; it is a *polynomial* mapping in u.

DEFINITION Let V and W be vector spaces over $\mathbb{R}$. Then $M: V^n \to W$ is said to be *multilinear* or n-linear if it is linear in each argument separately, i.e. if

$$M(v_1, \ldots, v_i + \alpha w, \ldots, v_n) = M(v_1, \ldots, v_i, \ldots, v_n) + \alpha M(v_1, \ldots, w, \ldots, v_n).$$

EXAMPLE 1 $V = \mathbb{R}^n$ *and* $M(v_1, v_2) = v_1^T A v_2$ where $A \in \mathbb{R}^{n \times n}$.

EXAMPLE 2 $V = W = L^\infty$, $h \in \mathcal{B}^2$, and

$$M(u_1, u_2) = \int\int h(\tau_1, \tau_2) u_1(t-\tau_1) u_2(t-\tau_2)\, d\tau_1\, d\tau_2.$$

DEFINITION Let $M: V^n \to W$ be n-linear. Then a map $P: V \to W$ of the form

$$P(v) = M(v, \ldots, v)$$

is said to be an n-order *polynomial mapping.*

EXAMPLE 3 $V = W = L^\infty$, $h \in \mathcal{B}^2$, and

$$P(u) = \int\int h(\tau_1, \tau_2) u(t-\tau_1) u(t-\tau_2)\, d\tau_1\, d\tau_2,$$

and in general the nth term of a Volterra series operator is an n-order polynomial mapping in the input u.

THEOREM 2.4.1 *An n-order polynomial mapping is homogeneous of degree n, but the converse is not true.*

Proof. $P(\alpha v) = M(\alpha v, \ldots, \alpha v) = \alpha^n M(v, \ldots, v) = \alpha^n P(v)$.
To see that the converse is not in general true, let $V = \mathbb{R}^2$, $W = \mathbb{R}$, and consider

$$F(x_1, x_2) = (|x_1| + |x_2|)^2 = x_1^2 + x_2^2 + 2\,|x_1 x_2|.$$

F is homogeneous of degree two but is not a polynomial mapping, since a second order polynomial mapping satisfies $P(x_1 + x_2) + P(x_1 - x_2) = 2P(x_1) + 2P(x_2)$; F does not.

This distinction between a homogeneous mapping and a polynomial mapping is like the difference between a general norm and a norm which comes from an inner product. To bring the discussion home to engineering consider the nonlinear TI operator $\tilde{N}$ given by

$$\tilde{N}u(t) = F[u(t), u(t-1)]$$

$\tilde{N}$ is homogeneous of degree two. We will see later that the response of a second order Volterra series operator to the input $u(t) = \cos t$ has, at most, two components: one at D.C. and one at $2\,\text{rad}\,\text{sec}^{-1}$. $\tilde{N}(\cos t)$, however, has infinitely many harmonics. □

We need just a few more definitions.

DEFINITION An n-linear map M is said to be *symmetric* if for any permutation $\sigma \in S^n$

$$M(v_{\sigma 1}, \ldots, v_{\sigma n}) = M(v_1, \ldots, v_n).$$

Thus the bilinear map of Example 1 is symmetric iff $A = A^T$, and the bilinear map of Example 2 is symmetric iff $h(\tau_1, \tau_2) = h(\tau_2, \tau_1)$.

DEFINITION sym M is the multilinear mapping defined by

$$\operatorname{sym} M(v_1, \ldots, v_n) := \frac{1}{n!} \sum_{\sigma \in S^n} M(v_{\sigma 1}, \ldots, v_{\sigma n})$$

and similarly if h_n is a function or measure, we define

$$\operatorname{sym} h_n(\tau_1, \ldots, \tau_n) := \frac{1}{n!} \sum_{\sigma \in S^n} h_n(\tau_{\sigma 1}, \ldots, \tau_{\sigma n}).$$

sym M derives its importance from the following theorem.

THEOREM 2.4.2 *Suppose the polynomial maps P_1 and P_2 are induced by multilinear maps M_1 and M_2, respectively. Then $P_1 = P_2$ iff* $\operatorname{sym} M_1 = \operatorname{sym} M_2$.

Thus two bilinear maps of the form of Example 1 induce the same polynomial map if and only if $A_1 + A_1^T = A_2 + A_2^T$.

Proof. First note that sym M and M always induce the same polynomial map, since

$$\operatorname{sym} M(v, \ldots, v) = \frac{1}{n!} \sum_{\sigma \in S^n} M(v, \ldots, v) = M(v, \ldots, v).$$

The 'if' part follows. In the next section we will prove more than the 'only if' part, so here we will give just an informal sketch of how the 'only if' proof goes. The key is the formula

$$\frac{1}{n!} \left[\frac{\partial^n}{\partial \alpha_1 \cdots \partial \alpha_n} \right]_{\alpha=0} P\left(\sum_{i=1}^{n} \alpha_i v_i \right) = \operatorname{sym} M(v_1, \ldots, v_n) \tag{2.6}$$

so that $P_1 = P_2$ implies $\operatorname{sym} M_1 = \operatorname{sym} M_2$. To 'establish' the formula, note that

$$P\left(\sum_{i=1}^{n} \alpha_i v_i \right) = \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} \alpha_{i_1} \cdots \alpha_{i_n} M(v_{i_1}, \ldots, v_{i_n}).$$

The only terms which contribute to

$$\left[\frac{1}{n!}\frac{\partial^n}{\partial\alpha_1\cdots\partial\alpha_n}\right]_{\alpha=0}$$

are the $n!$ terms where $[i_j]_{j=1}^n$ is a permutation of $[1, 2, \ldots, n]$, and the resulting sum is sym $M(v_1, \ldots, v_n)$. Of course we do not know yet that these derivatives exist, but we will see later that if the multilinear operators are bounded, then these derivatives can be interpreted as honest Frechet derivatives.

This process of determining sym M from the polynomial map P induced by M is known as *polarization*. In fact, we could replace the formula (2.6) above involving partial derivatives with a purely algebraic one; for example for $n = 2$ we have the polarization formula

$$\text{sym } M(u_1, u_2) = P\left(\frac{u_1+u_2}{2}\right) - P\left(\frac{u_1-u_2}{2}\right).$$

We gave the formula (2.6) because it generalizes to whole Volterra Series; the algebraic identities do not.

Let us now assume that V and W are Banach spaces. Then an n-linear map $M: V^n \to W$ is *bounded* if

$$\sup_{\|v_i\|\leq 1} \|M(v_1, \ldots, v_n)\| < \infty \tag{2.7}$$

in which case we call the left-hand side of (2.7) the norm of M as a multilinear operator and denote it $\|M\|_{\text{ml}}$. The bilinear operator of Example 1 is bounded, with $\|M\|_{\text{ml}} = \bar{\sigma}(A)$.† The bilinear operator in Example 2 is bounded with norm at most $\|h_2\|$.‡

We now quickly review derivatives in Banach space (Dieudonne, 1969; Balakrishnan, 1976). Recall that $\mathscr{L}(V, W)$ denotes the Banach space of bounded linear maps from V into W, with the operator norm $\|A\| := \sup\{\|Av\| : \|v\| \leq 1\}$. A map $N: G \to W$, where G is an open subset of V, is said to have a *Frechet* or *strong* derivative $DN(u_0) \in \mathscr{L}(V, W)$ at $u_0 \in G$ if

$$\|N(u_0 + u) - N(u_0) - DN(u_0)u\| = o(\|u\|).$$

If the map $u_0 \mapsto DN(u_0)$ has a Frechet derivative, we say N has a second Frechet derivative $D^{(2)}N(u_0)$ and it is an element of $\mathscr{L}[V, \mathscr{L}(V, W)]$. Fortunately this space can be identified with $\mathscr{L}_2(V, W)$, the space of bounded bilinear maps $: V^2 \to W$, with the norm $\|\cdot\cdot\|_{\text{ml}}$ defined above. Similarly the nth Frechet derivative, if it exists, can be thought of as a bounded n-linear map $: V^n \to W$. It can be shown that $D^{(n)}N(u_0)$ is symmetric, e.g. if $D^{(n+1)}N(u_0)$ exists.

2.5 *Relation to Taylor Series; Uniqueness of Volterra Series*

We will now see that Volterra series operators *are* Taylor series of operators from some open ball in L^∞ into L^∞.

† $\bar{\sigma}(A)$ means the largest singular value of A; here we assume the Euclidean norm on $\mathbb{R}^n$.
‡ The actual norm, rather than this upper bound, is hard to compute; see Appendix 3.

THEOREM 2.5.1 (Frechet derivatives of Volterra series operators) *On B_ρ, N has Frechet derivatives of all orders with*

$$D^{(k)}N(u_0)(u_1,\ldots,u_k)(t)= \tag{2.8a}$$

$$\sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)\int\cdots\int \operatorname{sym} h_n(\tau_1,\ldots,\tau_n)\prod_{i=1}^{k} u_i(t-\tau_i)\,d\tau_i \prod_{i=k+1}^{n} u_0(t-\tau_i)\,d\tau_i. \tag{2.8b}$$

Thus $\|D^{(k)}N(u_0)\|\leq f^{(k)}(\|u_0\|)$ and $(n!)^{-1}D^{(n)}N(0)$ is the n-linear mapping associated with the nth *term of the Volterra series and given by:*

$$\frac{1}{n!}D^{(n)}N(0)(u_1,\ldots,u_n)(t)=$$

$$\int\cdots\int h_n(\tau_1,\ldots,\tau_n)u_1(t-\tau_1)\cdots u_n(t-\tau_n)\,d\tau_1\cdots d\tau_n. \tag{2.8c}$$

Remark. (2.8c) of Theorem 2.5.1 tells us that the Volterra series we have considered so far are in fact Taylor series of operators $:L^\infty\to L^\infty$. The reader may wonder whether the Volterra series constitute *all* of the Taylor series of TI nonlinear maps $:L^\infty\to L^\infty$. In Appendix 3 we show that this is *not* true, but that the Taylor series left out are not important in engineering.

Proof. Let M_k denote the multilinear map given in (2.8b) above. We will show that

$$N(u_0+u)-\sum_{k=0}^{n}\frac{1}{k!}M_k(u,\ldots,u)=o(\|u\|^{n+1})$$

which will prove $M_k=D^{(k)}N(u_0)$ as claimed. First note that

$$\|M_k\|\leq\sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)\,\|h_n\|\cdot\|u_0\|^{n-k}=f^{(k)}(\|u_0\|).$$

Now using the symmetry of sym h_n

$$N(u_0+u)=\sum_{n=1}^{\infty}\int\cdots\int \operatorname{sym} h_n(\tau_1,\ldots,\tau_n)\sum_{k=0}^{n}\binom{n}{k}\prod_{i=1}^{k}u(t-\tau_i)\,d\tau_i\prod_{i=k+1}^{n}u_0(t-\tau_i)\,d\tau_i. \tag{2.9}$$

For $\|u\|$ small enough ($\|u\|+\|u_0\|<\rho$ will do) the entire right-hand side of equation (2.9) is absolutely convergent so we may rewrite it as:

$$\sum_{k=0}^{\infty}\frac{1}{k!}\sum_{n=k}^{\infty}n(n-1)\cdots(n-k+1)\int\cdots$$

$$\cdots\int \operatorname{sym} h_n(\tau_1,\ldots,\tau_n)\prod_{i=1}^{k}u_i(t-\tau_i)\,d\tau_i\prod_{i=k+1}^{n}u_0(t-\tau_i)\,d\tau_i=\sum_{k=0}^{\infty}\frac{1}{k!}M_k(u,\ldots,u).$$

Thus we have

$$\left\| N(u_0+u) - \sum_{k=0}^{n} \frac{1}{k!} M_k(u,\ldots,u) \right\| \leq \sum_{k=n+1}^{\infty} \frac{1}{k!} \|M_k\| \cdot \|u\|^k$$
$$\leq \sum_{k=n+1}^{\infty} \frac{1}{k!} f^{(k)}(\|u_0\|)\,\|u\|^k$$
$$= f(\|u_0\|+\|u\|) - \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(\|u_0\|)\,\|u\|^k$$

which is indeed $o(\|u\|^n)$. □

THEOREM 2.5.2 (Uniqueness theorem for Volterra series) *Suppose N and M are Volterra series operators with kernels h_n and g_n, respectively. Then $N = M$ iff* $\operatorname{sym} h_n = \operatorname{sym} g_n$ *for all n.*

Note that $N = M$ asserts equality of maps from some ball in L^∞ into L^∞, whereas the conclusion asserts equality of a sequence of measures.

Proof. The 'if' part is clear, (see Theorem 2.4.2). To show the 'only if' part we will show that the measures $\operatorname{sym} h_n$ are determined by the operator N. A measure $\mu \in \mathcal{B}^n$ is determined by its integral over all n-rectangles in $\mathbb{R}^n$, i.e. by the integrals

$$\int \cdots \int \mu(\tau_1,\ldots,\tau_n) u_1(-\tau_1) \cdots u_n(-\tau_n)\, d\tau_1 \cdots d\tau_n \tag{2.10}$$

where each u_i is the characteristic function of an interval. Now by Theorem 2.5.1 we have

$$\int \cdots \int \operatorname{sym} h_n(\tau_1,\ldots,\tau_n) u_1(-\tau_1) \cdots u_n(-\tau_n)\, d\tau_1 \cdots d\tau_n = \frac{1}{n!} D^{(n)} N(u_1,\ldots,u_n)(0)$$

so that N determines the integrals in (2.10) and hence the measure $\operatorname{sym} h_n$. A more explicit formula for these integrals is:

$$\int \cdots \int \operatorname{sym} h_n(\tau_1,\ldots,\tau_n) u_1(-\tau_1) \cdots u_n(-\tau_n)\, d\tau_1 \cdots d\tau_n$$
$$= \frac{1}{n!} \left[\frac{\partial}{\partial\alpha_1 \cdots \partial\alpha_n} \right]_{\alpha=0} N\left(\sum_{i=1}^{n} \alpha_i u_i \right)(0)$$

which is the formula mentioned in the previous section. □

The Uniqueness theorem tells us that we may as well choose our kernels h_n to be symmetric, and from now on we will assume that all kernels are symmetric. Of course other canonical forms are possible and in some cases more convenient. For example the *triangular* kernels satisfy

$$h_{\text{tri}\,n}(\tau_1,\ldots,\tau_n) = 0 \quad \text{unless} \quad 0 \leq \tau_1 \leq \cdots \leq \tau_n$$

and the Volterra series is then

$$Nu(t)=\sum_{n=1}^{\infty}\int_0^t\int_0^{\tau_n}\cdots\int_0^{\tau_2}h_{\text{tri}\,n}(\tau_1,\ldots,\tau_n)u(t-\tau_1)\cdots u(t-\tau_n)\,\mathrm{d}\tau_1\cdots\mathrm{d}\tau_n.$$

One point worth mentioning: the triangle inequality implies

$$\|\text{sym}\,h_m\|=\|h_{\text{tri}\,n}\|\leq\|h_n\|.$$

Thus using the symmetric (or triangular) kernels can only *decrease* the gain bound function f and hence increase the radius of convergence ρ. In the sequel we will refer to the gain bound function and radius of convergence computed from the symmetric kernels as *the* gain bound function and radius of convergence of N.

2.6 *Final Comments on the Formulation*

The formulation we have given is by no means the only possible. For example, we could interpret the norms on input signals and kernels as L^2 norms, leaving the norm on output signals (i.e. $y=Nu$) an L^∞ norm. Input signals and kernels would thus be L^2 functions with

$$\|h_n\|:=\|h_n\|_2=\left(\int\cdots\int h_n(\tau_1,\ldots,\tau_n)^2\,\mathrm{d}\tau_1\cdots\mathrm{d}\tau_n\right)^{\frac{1}{2}}.$$

Then with the exception of the L^p material of Section 2.3 *all the preceding results hold.* This is essentially the Fock space framework proposed by De Figueiredo (1983).†

3. Applications to systems theory

In this section we apply the ideas of the previous section to give simple rigorous proofs of some well-known theorems. We show that the sum, pointwise product, and composition of two Volterra series operators have Volterra series and we bound their gain functions. We proceed to find the condition under which a Volterra series operator has an inverse and compute its kernels. This is applied to show that the I/O operator of a simple dynamical system is given by a Volterra series.

This program of working out the Volterra series of various 'system interconnections' was first carried out at MIT in the late 1950s (Barrett, 1963; Brilliant, 1958), but none of this work is rigorous. This constructive approach is not really a fully modern approach, where one powerful general theorem would prove all these theorems (and more) (Sandberg, 1983*a*). Unfortunately this one powerful theorem may be so general and abstract that the underlying simplicity of the formulas may be lost. In this section we want to demonstrate two things: first, that supplying the analytical details in the MIT work is relatively straightforward; and

† Our Volterra series with radii exceeding r would be almost all of the Fock space with weights $N!\,r^n$.

second, that the resulting formulas, though complicated, are just simple extensions of the same formulas for ordinary power series. This of course should be expected in view of Theorem 2.5.1.

The notation for this section is as follows: A and B will denote Volterra series operators with kernels a_n and b_n, gain bound functions f_A and f_B, and radii of convergence ρ_A and ρ_B, respectively.

3.1 *Sum and Product Operators*

The pointwise product of A and B is defined by

$$[A \cdot B]u(t) = [Au](t)[Bu](t).$$

DEFINITION If $a \in \mathcal{B}^n$, $b \in \mathcal{B}^k$ then the *symmetric tensor product* $a \vee b \in \mathcal{B}^{n+k}$ is defined by:

$$[a \vee b](\tau_1, \ldots, \tau_{n+k}) := \text{sym}\,[a(\tau_1, \ldots, \tau_n)b(\tau_{n+1}, \ldots, \tau_{n+k})].$$

By the product we mean the normal product measure. (Thus $h(\tau)g(\tau)$ does not necessarily make sense, but $h(\tau_1)g(\tau_2)$ does.) Note that

$$\begin{aligned}\|a \vee b\| &= \int \cdots \int |\text{sym}\,[a(\tau_1, \ldots, \tau_n)b(\tau_{n+1}, \ldots, \tau_{n+k})]|\, d\tau_1 \cdots d\tau_{n+k} \\ &\leq \frac{1}{(n+k)!} \sum_{\sigma \in S^{n+k}} \int \cdots \int |a(\tau_{\sigma 1}, \ldots, \tau_{\sigma n})|\, |b(\tau_{\sigma n+1}, \ldots, \tau_{\sigma n+k})|\, d\tau_1 \cdots d\tau_{n+k} \\ &= \|a\| \cdot \|b\|.\end{aligned}$$

THEOREM 3.1.1 (Product operator) *$A \cdot B$ is a Volterra series operator with kernels*

$$h_n = \sum_{k=1}^{n-1} a_k \vee b_{n-k}$$

and characteristic gain function $f_{A \cdot B} \leq f_A f_B$. In particular $\rho_{A \cdot B} \geq \min\{\rho_A, \rho_B\}$.

Remark. If we write a Volterra series as a formal sum

$$a_1 + \cdots + a_n + \cdots$$

then we can write the formal symmetric tensor product of $a_1 + \cdots$ and $b_1 + \cdots$ as

$$(a_1 + \cdots) \vee (b_1 + \cdots) = (a_1 \vee b_1) + (a_1 \vee b_2 + a_2 \vee b_1) + \cdots$$

so the Volterra series of $A \cdot B$ is the formal symmetric tensor product of the Volterra series of A and B. Note the similarity with the formula for the coefficients of the product of two power series.

Proof. Let $\|u\| < \min\{\rho_A, \rho_B\}$. Then Au and Bu make sense and

$$A \cdot Bu(t) = \left(\sum_{m=1}^{\infty} \int \cdots \int a_m(\tau_1, \ldots, \tau_m) \prod_{i=1}^{m} u(t-\tau_i)\, d\tau_i\right) \cdot \left(\sum_{n=1}^{\infty} \int \cdots \int b_n(\tau_1, \ldots, \tau_n) \prod_{i=1}^{n} u(t-\tau_i)\, d\tau_i\right) \tag{3.1a}$$

$$= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \int \cdots \int a_m(\tau_1, \ldots, \tau_m) b_n(\tau_{m+1}, \ldots, \tau_{m+n}) \prod_{i=1}^{m+n} u(t-\tau_i)\, d\tau_i \tag{3.1b}$$

$$= \sum_{n=1}^{\infty} \int \cdots \int \left(\sum_{k=1}^{n-1} a_k \vee b_{n-k} \right) u(t-\tau_1) \cdots u(t-\tau_n)\, d\tau_1 \cdots d\tau_n. \tag{3.1c}$$

All of the changes in the order of summation and integration in equations (3.1) are justified by the Fubini theorem, since

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \int \cdots \int |a_m(\tau_1, \ldots, \tau_m) b_n(\tau_{m+1}, \ldots, \tau_{m+n})| \prod_{i=1}^{m+n} |u(t-\tau_i)|\, d\tau_i$$
$$\leq \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \|a_n\| \cdot \|b_m\| \cdot \|u\|^{m+n} = f_A(\|u\|) f_B(\|u\|) < \infty.$$

Since equation (3.1) holds for *any* u with $\|u\| < \min\{\rho_A, \rho_B\}$, the Uniqueness theorem tells us that $\sum_{k=1}^{n-1} a_k \vee b_{n-k}$ are *the* kernels of $A \cdot B$. Now

$$f_{A \cdot B}(x) = \sum_{n=1}^{\infty} \|h_n\| x^n = \sum_{n=1}^{\infty} \sum_{k=1}^{n-1} \|a_k \vee b_{n-k}\| x^n$$
$$\leq \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \|a_k\| \cdot \|b_{n-k}\| x^{n-k}$$
$$= \left(\sum_{n=1}^{\infty} \|a_n\| x^n \right) \left(\sum_{n=1}^{\infty} \|b_n\| x^n \right) = f_A(x) f_B(x).$$

The final conclusion $\rho_{A \cdot B} \geq \min\{\rho_A, \rho_B\}$ follows from $f_{A \cdot B} \leq f_A f_B$ and $\rho_{A \cdot B} = \inf\{x : f_{A \cdot B} = \infty\}$. □

THEOREM 3.1.2 (Sum operator) *$A + B$ is a Volterra series operator with kernels*

$$h_n(\tau_1, \ldots, \tau_n) = a_n(\tau_1, \ldots, \tau_n) + b_n(\tau_1, \ldots, \tau_n)$$

and gain bound function $f_{A+B} \leq f_A + f_B$. Thus $\rho_{A+B} \geq \min\{\rho_A, \rho_B\}$.

The proof is left to the reader.

3.2 *Composition Operator*

The composition of A and B, which we denote by the juxtaposition AB, is defined by

$$[AB]u(t) := A(Bu)(t)$$

To motivate the formula for the kernels of AB, recall that the nth coefficient of the composition of the ordinary power series $\sum_{l=1}^{\infty} a_l x^l$ and $\sum_{l=1}^{\infty} b_l x^l$ is given by

$$\sum_{k=1}^{n} \left[\sum_{\substack{i_1, \ldots, i_k \geq 1 \\ i_1 + \cdots + i_k = n}} \right] a_k b_{i_1} \cdots b_{i_k}. \tag{3.2}$$

THEOREM 3.2.1 (Composition theorem) *AB is a Volterra series operator with kernels*

$$h_n(t_1,\ldots,t_n)=\text{sym}\sum_{k=1}^{n}\left[\sum_{\substack{i_1,\ldots,i_k\geq 1\\ i_1+\cdots+i_k=n}}\right]\int\cdots$$

$$\cdots\int a_k(\tau_1,\ldots,\tau_k)b_{i_1}(t_1-\tau_1,\ldots,t_{i_1}-\tau_1)\cdots b_{i_k}(t_{n-i_k+1}-\tau_k,\ldots,t_n-\tau_k)\,d\tau_1\cdots d\tau_k. \tag{3.3}$$

Moreover $f_{AB}(x)\leq f_A[f_B(x)]$. *Thus* $\rho_{AB}\geq\min\{\rho_B, f_B^{-1}(\rho_A)\}$.

Proof. Let h_n be defined by the formula (3.3) above. First note that

$$\|h_n\|\leq\sum_{k=1}^{n}\left[\sum_{\substack{i_1,\ldots,i_k\geq 1\\ i_1+\cdots+i_k=n}}\right]\|a_k\|\cdot\|b_{i_1}\|\cdot\cdots\cdot\|b_{i_k}\| \tag{3.4}$$

and the right-hand side of (3.4) is the nth coefficient of $f_A[f_B(\cdots)]$ so $f_H(x)\leq f_A[f_B(x)]$. This computation justifies the changes of order of integration and summation in the following. Suppose $f_A[f_B(\|u\|)]<\infty$. Then Bu makes sense and $\|Bu\|\leq f_B(\|u\|)$ so ABu makes sense and:

$$ABu(t)=\sum_{k=1}^{\infty}\int\cdots\int a_k(\tau_1,\ldots,\tau_k)Bu(t-\tau_1)\cdots Bu(t-\tau_k)\,d\tau_1\cdots d\tau_k$$

$$=\sum_{k=1}^{\infty}\int\cdots\int a_k(\tau_1,\ldots,\tau_k)\cdot$$

$$\prod_{i=1}^{k}\left(\sum_{m=1}^{\infty}\int\cdots\int b_m(t_1,\ldots,t_m)u(t-\tau_i-t_1)\cdots u(t-\tau_i-t_m)\,dt_1\cdots dt_m\right)d\tau_i$$

$$=\sum_{k=1}^{\infty}\int\cdots\int a_k(\tau_1,\ldots,\tau_k)\left[\sum_{i_1,\ldots,i_k\geq 1}\right]b_{i_1}(t_1,\ldots,t_{i_1}-\tau_1)\cdots$$

$$\cdots b_{i_k}(t_{i_1+\cdots+i_{k-1}},\ldots,t_{i_1+\cdots+i_k})u(t-\tau_1-t_1)\cdots u(t-\tau_1-t_{i_1})\cdots$$

$$\cdots u(t-\tau_k-t_{i_1+\cdots+i_{k-1}+1})\cdots u(t-\tau_k-t_{i_1+\cdots+i_k})\,dt_1\cdots dt_{i_1+\cdots+i_k}\,d\tau_1\cdots d\tau_k.$$

We now collect terms by degree in u to get:

$$ABu(t)=\sum_{n=1}^{\infty}\int\cdots\int\Bigg(\sum_{k=1}^{n}\left[\sum_{\substack{i_1,\ldots,i_k\geq 1\\ i_1+\cdots+i_k=n}}\right]a_k(\tau_1,\ldots,\tau_k)\cdot$$

$$b_{i_1}(t_1,\ldots,t_{i_1})\cdots b_{i_k}(t_{n-i_k+1},\ldots,t_n)u(t-\tau_1-t_1)\cdots u(t-\tau_1-t_{i_1})\cdots$$

$$\cdots u(t-\tau_k-t_{n-i_k+1})\cdots u(t-\tau_k-t_n)\Bigg)\,dt_1\cdots dt_n\,d\tau_1\cdots d\tau_k.$$

Finally, we change the t_i variables:

$$ABu(t)=\sum_{n=1}^{\infty}\int\cdots\int\ \sum_{k=1}^{n}\left[\sum_{\substack{i_1,\ldots,i_k\geq 1\\ i_1+\cdots+i_k=n}}\right]\int\cdots\int a_k(\tau_1,\ldots,\tau_k)\cdot$$

$$b_{i_1}(t_1-\tau_1,\ldots,t_{i_1}-\tau_1)\cdots b_{i_k}(t_{n-i_k+1}-\tau_k,\ldots,t_n-\tau_k)\,d\tau_1\cdots d\tau_k\ \ u(t-\tau_1)\cdots$$

$$\cdots u(t-\tau_n)\,dt_1\cdots dt_n$$

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

and the uniqueness theorem tells us h_n are the kernels of AB. Equation (3.4) establishes the bound $f_{AB}\leq f_A f_B$, and the lower bound on the radius of convergence of AB follows. □

3.3 *Inverses of Volterra Series Operators*

We now ask the question: when does a Volterra series operator have a local inverse near 0 given by a Volterra series operator and what are its kernels? Whole papers have been written on this important topic (Halme, 1971; Halme & Orava 1972). Just like ordinary power series, the condition is just that the first term be invertible.

THEOREM 3.3.1 (Inversion theorem for Volterra series) *A has a local inverse at* 0 *if and only if its first kernel a_1 is invertible in $\mathcal{B}^1$, i.e. there exists a measure $b_1\in\mathcal{B}^1$ with $a_1 \star b_1=\delta$.*†

Remark. Since the Frechet derivative of A at 0 is given by convolution with a_1 (Theorem 2.5.1), the Inversion theorem can be thought of as a generalized *Inverse-function theorem.* We will not pursue this idea further: instead we take a constructive approach.

Proof. To see the 'only if' part, suppose A has a local inverse B, that is

$$AB=BA=I \tag{3.5}$$

where I is the identity operator ($I_1=\delta$, $I_n=0$ for $n>1$). Using the composition theorem to compute the first kernel of the operators in equation (3.5) yields

$$a_1\star b_1=b_1\star a_1=\delta. \tag{3.6}$$

Thus a_1 is invertible in $\mathcal{B}^1$.

The proof of the 'if' part will proceed as follows: we first construct a right inverse for A under the assumption that the first kernel is just δ. Using this we show that A has a right inverse in the general case, a_1 invertible in $\mathcal{B}^1$. We finish the proof by showing that the right inverse constructed is in fact also a left inverse for A.

Special case. Assume for now that $a_1=\delta$. To motivate what follows, consider an ordinary power series $a(x):=\sum_{n=1}^{\infty}a_nx^n$ with $a_1=1$. Since $a'(0)=1$, $a(\cdot)$ has an analytic inverse $b(x):=\sum_{n=1}^{\infty}b_nx^n$ near 0. Using formula (3.3) for the coefficients of the composition $a[b(x)]=x$ yields $b_1=1$ and the following recursive

† Since the convolution of measures in $\mathcal{B}^1$ is commutative, $a_1\star b_1=\delta$ implies $b_1\star a_1=\delta$.

formula for b_n:

$$b_n = -\sum_{k=2}^{n}\left[\sum_{\substack{i_1,\ldots,i_k\geq 1\\ i_1+\cdots+i_k=n}}\right] a_k b_{i_1}\cdots b_{i_k}. \tag{3.7}$$

Note that since the index k starts at two, the right-hand side of (3.7) refers only to $b_1, \ldots, b_{n-1}$. Incidentally this process of recursively computing the coefficients of the inverse of an analytic function is known as *reversion* of a power series (Bromwich, 1908).

We now use the same construction for Volterra series. Let $b_1 = \delta$ and for $n > 1$ define measures $b_n \in \mathscr{B}^n$ recursively by

$$\begin{aligned} b_n(t_1, \ldots, t_n) = -\mathrm{sym} \sum_{k=2}^{n}\left[\sum_{\substack{i_1,\ldots,i_k\geq 1\\ i_1+\cdots+i_k=n}}\right]\int\cdots\int a_k(\tau_1, \ldots, \tau_k). \\ b_{i_1}(t_1-\tau_1, \ldots, t_{i_1}-\tau_1)\cdots b_{i_k}(t_{n-i_k+1}-\tau_k, \ldots, t_n-\tau_k)\,\mathrm{d}\tau_1\cdots\mathrm{d}\tau_k. \end{aligned} \tag{3.8}$$

As in (3.7) above this comes directly from the composition formula and $(AB)_n = 0$, $n > 1$. We now have to show that the b_n, as defined in (3.8) above, are actually the kernels of a Volterra series operator: we must verify that assumptions (A1) and (A2) hold.

We establish (A1) by induction. First note that $b_1 = \delta \in \mathscr{B}^1$. Assuming that $b_j \in \mathscr{B}^j$ for $j = 1, \ldots, n-1$ (3.8) implies that $b_n \in \mathscr{B}^n$, with

$$\|b_n\| \leq \sum_{k=2}^{n}\left[\sum_{\substack{i_1,\ldots,i_k\geq 1\\ i_1+\cdots+i_k=n}}\right]\|a_k\|\cdot\|b_{i_1}\|\cdot\cdots\cdot\|b_{i_k}\|. \tag{3.9}$$

We now establish (A2). Let $g(x) := 2x - f_A(x)$. Since $g'(0) = 1$ (recall that $a_1 = \delta$) g has an analytic inverse $h(x) := \sum_{n=1}^{\infty}\alpha_n x^n$ near 0. We claim that $f_B(x) \leq h(x)$ and thus $\rho_B \geq \mathrm{rad}\, h(\cdot)$. The coefficients α_n are given by formula (3.7): $\alpha_1 = 1$ and for $n > 1$

$$\alpha_n = \sum_{k=2}^{n}\left[\sum_{\substack{i_1,\ldots,i_k\geq 1\\ i_1+\cdots+i_k=n}}\right]\|a_k\|\,\alpha_{i_1}\cdots\alpha_{i_k}. \tag{3.10}$$

By induction we now show

$$\|b_n\| \leq \alpha_n \tag{3.11}$$

for all n. (3.11) is true for $n = 1$; suppose (3.11) has been established for $n < m$. Then (3.9), (3.10), and the inductive hypothesis establish (3.11) for $n = m$ and hence for all n. Consequently

$$f_B(x) = \sum_{n=1}^{\infty}\|b_n\|\,x^n \leq \sum_{n=1}^{\infty}\alpha_n x^n = h(x)$$

which proves our claim above that the measures b_n do satisfy assumption (A2) and hence are the kernels of a Volterra series operator which we naturally enough call B. From the formula (3.8) for b_n we conclude

$$AB = I$$

B is thus a right inverse for A. This concludes the proof for the *special case.*

General case. Suppose now that a_1 is invertible in $\mathscr{B}^1$. We will use the proof of the *special case* presented above to prove the *general case.* Let $b_1 \in \mathscr{B}^1$ satisfy $a_1 \star b_1 = \delta$. Let A_{lin} be the Volterra series operator with first kernel a_1 and other kernels zero. A_{lin} is invertible, with inverse A_{lin}^{-1} (which has first kernel b_1 and other kernels zero). Consider the operator $A_{\text{lin}}^{-1}A$ whose kernels we could easily compute with the composition theorem. Its first kernel is δ, so using the construction above find a local right inverse C to $A_{\text{lin}}^{-1}A$. Then $B = CA_{\text{lin}}^{-1}$ is the local right inverse of A, since

$$AB = A_{\text{lin}}A_{\text{lin}}^{-1}ACA_{\text{lin}}^{-1} = A_{\text{lin}}A_{\text{lin}}^{-1} = I. \tag{3.12}$$

Our final task is to show that the right inverse B is also a *left* inverse for A. Since the first kernel of B is invertible (indeed it has inverse a_1) we can find a right inverse D for B. Then we have

$$A = AI = A(BD) = (AB)D = ID = D. \tag{3.13}$$

(3.13) and $BD = I$ shows

$$BA = I$$

which with (3.12) proves that B really is the local inverse of A at 0 and completes the proof of Theorem 3.4.1. □

Remark 1. If $a_1 \in \mathscr{A}$, the subalgebra of $\mathscr{B}^1$ of those measures lacking singular continuous part, then we have the criterion (Desoer & Vidyasagar, 1975)

$$A \quad \text{is invertible iff} \inf_{\operatorname{Re} s \geq 0} |\hat{a}_1(s)| > 0$$

where $\hat{a}_1(s)$ denotes the Laplace transform of a_1.

Remark 2. It is interesting to note that the special case considered above has the interpretation of *unity feedback around a strictly nonlinear operator,* which is an important system-theoretic topic in its own right.

3.4 *Dynamical System Example*

To illustrate the theorems of this section we now work an example.

EXAMPLE Consider the one-dimensional dynamical system:

$$\dot{x} = f(x) + g(u), \tag{3.14a}$$

$$x(0) = 0, \tag{3.14b}$$

and

$$y = q(x). \tag{3.14c}$$

Suppose f, g, and q are analytic near 0, $f(0) = g(0) = q(0) = 0$, and $f'(0) < 0$. Then the system is exponentially stable at 0, and for $\|u\|$ small there is a unique state trajectory x satisfying (3.14). We will now show that the I/O map $:u \mapsto y$ is a Volterra series operator.

Proof. We first use a *loop transformation* to reexpress equations (3.14a) and (3.14b) in terms of Volterra series operators. (3.14a) and (3.14b) are equivalent to

$$x = e^{f'(0)t} \star [f_{snl}(x) + g(u)]$$

where $f_{snl}(x) := f(x) - f'(0)x$, the *strictly nonlinear* part of f: see Fig. 2. Let H_{lin} be the Volterra series operator with first kernel $1_H(\tau)e^{f'(0)\tau}$ and other kernels 0. Let F_{snl}, G, and Q be the memoryless Volterra series operators associated with the functions f_{snl}, g, and q, respectively, e.g. $Q_n(\tau_1, \ldots, \tau_n) = n!^{-1}q^{(n)}(0)\delta(\tau_1) \cdots \delta(\tau_n)$. Then the system equations (3.14) are equivalent to

$$x = H_{lin}[F_{snl}(x) + G(u)] \tag{3.15a}$$

and

$$y = Q(x) \tag{3.15b}$$

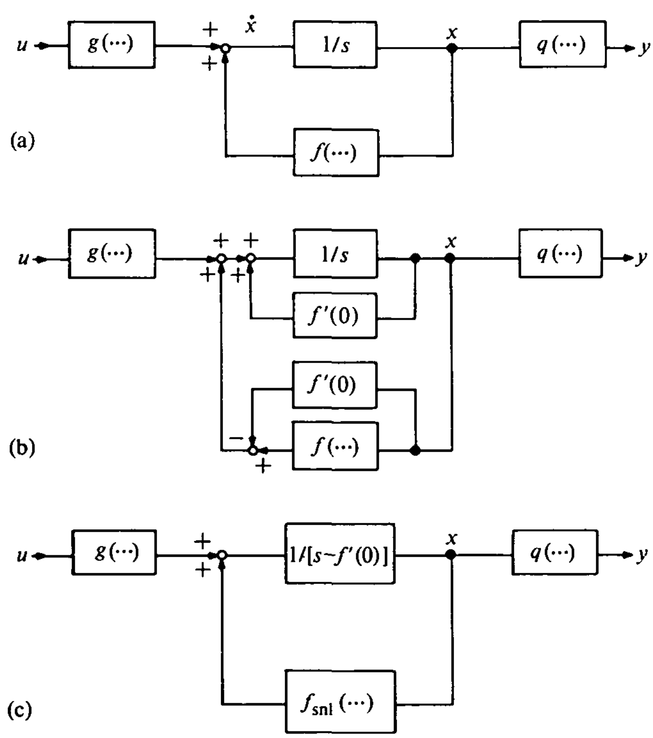

FIG. 2.

Since H_{lin} is linear

$$(I - H_{\text{lin}} F_{\text{snl}})x = H_{\text{lin}} Gu. \tag{3.16}$$

By the sum and composition theorems $I - H_{\text{lin}} F_{\text{snl}}$ is a Volterra series operator with first kernel δ. By the inversion theorem $I - H_{\text{lin}} F_{\text{snl}}$ has a Volterra series local inverse $(I - H_{\text{lin}} F_{\text{snl}})^{-1}$ near 0. Since as mentioned above (3.16) has only one solution x when $\|u\|$ is small, it must be

$$x = (I - F_{\text{snl}} H_{\text{lin}})^{-1} Gu. \tag{3.17}$$

Thus for $\|u\|$ small, the output y is given by a Volterra series operator in u:

$$y = Q(I - F_{\text{snl}} H_{\text{lin}})^{-1} Gu. \tag{3.18}$$

A few comments are in order. (3.14) may have *multiple equilibria* when $u \equiv 0$, for example if $f(x) = -\sin x$; or even a finite escape time for some inputs u, for example if $f(x) = -x + x^2$. We have shown that as long as $\|u\|$ is small enough, say less than K, then the state x and the output y are given by a Volterra series in u. In particular $\|u\| < K$ must keep the state x from leaving the domain of attraction of 0, for otherwise the Steady-state theorem (see Section 4.1) or the Gain Bound theorem would be violated.

4. Frequency-domain topics

In this section we consider frequency-domain topics, concentrating on the simplest case: periodic inputs. Even in this case the analysis is not simple. Nevertheless we show that an intuitive formula for the output spectrum in terms of the input spectrum holds in essentially all engineering contexts.

Before starting our topic proper, we prove the following theorem.

4.1 *The Steady-state Theorem*

THEOREM 4.1.1 (Steady-state theorem): *Let u and u_s be any signals with: $\|u\|, \|u_s\| < \rho = \text{rad}\, N$, and suppose that $u(t) \to u_s(t)$ as $t \to \infty$. Then $Nu(t) \to Nu_s(t)$ as $t \to \infty$.*

This is a very different concept from that of the continuity of N as a map from L^∞ to L^∞, which tells us e.g. that if $u_n \to u$ *uniformly* as $n \to \infty$, then $Nu_n \to Nu$ (uniformly).†

Proof. Suppose: $\|u\|, \|u_s\| < \rho$ and $u(t) \to u_s(t)$ as $t \to \infty$. Let $v = u_s - u$ so $v(t) \to 0$ as $t \to \infty$. The proof is a modification of the proof of the incremental gain theorem; we simply break the estimate into two parts, one due to the recent past only. For $T \geq 0$

$$(Nu_s - Nu)(t) = [N(u+v) - Nu](t) = I_1(t, T) + I_2(t, T)$$

† Indeed the Steady-state theorem is *false* for some pathological LTI bounded (and therefore continuous) operators from L^∞ into L^∞: see Appendix 3.

where

$$I_1(t,T):=\sum_{n=1}^{\infty}\int_{[0,T]^n}\cdots\int h_n(\tau_1,\ldots,\tau_n)\Big(\prod_{i=1}^{n}u_s(t-\tau_i)-\prod_{i=1}^{n}u(t-\tau_i)\Big)\,d\tau_1\cdots d\tau_n$$

and

$$I_2(t,T):=\sum_{n=1}^{\infty}\int_{R^n\setminus[0,T]^n}\cdots\int h_n(\tau_1,\ldots,\tau_n)\Big(\prod_{i=1}^{n}u_s(t-\tau_i)-\prod_{i=1}^{n}u(t-\tau_i)\Big)\,d\tau_1\cdots d\tau_n.$$

We now estimate I_1 and I_2 separately.

$$I_1(t,T)=\sum_{n=1}^{\infty}\int_{[0,T]^n}\cdots\int h_n(\tau_1,\ldots,\tau_n)\Big(\prod_{i=1}^{n}(u+v)(t-\tau_i)-\prod_{i=1}^{n}u(t-\tau_i)\Big)\,d\tau_1\cdots d\tau_n$$

$$=\sum_{n=1}^{\infty}\int_{[0,T]^n}\cdots\int h_n(\tau_1,\ldots,\tau_n)\sum_{k=1}^{n}\binom{n}{k}\prod_{i=1}^{k}v(t-\tau_i)\,d\tau_i\prod_{i=k+1}^{n}u(t-\tau_i)\,d\tau_i$$

using the symmetry of h_n. Thus

$$|I_1(t,T)|\leq\sum_{n=1}^{\infty}\|h_n\|\sum_{k=1}^{n}\binom{n}{k}\|v\|_{[t-T,t]}^{k}\|u\|^{n-k}\tag{4.1a}$$

$$=f(\|u\|+\|v\|_{[t-T,t]})-f(\|u\|)\tag{4.1b}$$

where $\|v\|_{[t-T,t]}$ means $\sup\{|v(\tau)|:t-T\leq\tau\leq t\}$ and f is the gain bound function of N. Note that (4.1b) may be ∞ for some t, T. But as $t-T\to\infty$, $\|v\|_{[t-T,t]}\to 0$ so (4.1b) eventually becomes and stays finite and in fact converges to zero. Thus we conclude:

$$I_1(t,T)\to 0\quad\text{as}\quad t-T\to\infty.\tag{4.2}$$

Now we estimate I_2:

$$|I_2(t,T)|\leq\sum_{n=1}^{\infty}\|h\|_{\mathbf{R}^n\setminus[0,T]^n}(\|u_s\|^n+\|u\|^n)\tag{4.3}$$

where

$$\|h_n\|_{\mathbf{R}^n\setminus[0,T]^n}=\int_{\mathbf{R}^n\setminus[0,T]^n}\cdots\int|h_n(\tau_1,\ldots,\tau_n)|\,d\tau_1\cdots d\tau_n.\tag{4.4}$$

For each n (4.4) decreases to zero as T increases to ∞, since each h_n is a bounded measure. Hence each term in the sum in (4.3) decreases to zero as T increases to ∞. The right-hand side of (4.3) is always less than $f(\|u_s\|)+f(\|u\|)$, so the dominated convergence theorem tells us that the right-hand side of (4.3), and hence $I_2(t,T)$, converges to zero as $T\to\infty$.

If we now set $T=t/2$ then as $t\to\infty$ both T and $t-T$ increase to ∞. Hence as $t\to\infty$, $Nu_s(t)-Nu(t)=I_1(t,t/2)+I_2(t,t/2)\to 0$. □

Remark. Unlike linear systems, the rate of convergence can depend on the *amplitude* of the input. For example, consider N given by

$$Nu = \sum_{k=1}^{\infty} u(t-k)^k.$$

N has radius of convergence one. Now consider step inputs of amplitude α, $0<\alpha<1$. As α increases to one, the time to convergence to within, say, one per cent of the steady state grows like $(1-\alpha)^{-1}$. For *linear systems* the time to convergence is independent of the amplitude of the input.

Although in the Steady-state theorem u_s can be any signal with $\|u_s\|<\rho$, usually u_s has the interpretation of a *steady-state* input, for example in the following theorem.

THEOREM 4.1.2 (Periodic Steady-state theorem) *If the input u is periodic with period T for $t \geq 0$ then the output Nu approaches a steady state, also periodic with period T.*

Proof. Let u_s be u extended periodically to $t=-\infty$. Clearly $u(t) \to u_s(t)$ as $t \to \infty$ (indeed $u(t)=u_s(t)$ for $t \geq 0$) so by the Steady-state theorem $Nu(t) \to Nu_s(t)$ as $t \to \infty$. Nu_s is periodic with period T since

$$[Nu_s(\cdot)](t+T) = N[u_s(\cdot+T)](t) = Nu_s(t)$$

where the first equality is due to the time-invariance of N and the second equality is due to the T-periodicity of u_s. □

Note in particular that Volterra series operators *cannot* generate *subharmonics*. The following theorem is a related application of the Steady-state theorem.

THEOREM 4.1.3 (Almost periodic Steady-state theorem) *If the input u is almost periodic for $t \geq 0$ then the output approaches an almost periodic steady state.*

Proof. The hypothesis simply means that *there is* some u_s which is almost periodic and agrees with u for $t \geq 0$. By the Steady-state theorem we know $y(t) \to y_s(t) := Nu_s(t)$ so we need only show that Volterra series operators take almost periodic inputs into almost periodic outputs. The proof of this, as well as the formula for the spectral amplitudes of the output, are in Appendix A5. This last topic has been studied by Sandberg (1984).

4.2 *Frequency-domain Volterra Kernels*

As with linear systems, it is often convenient to use the Laplace transforms of the kernels, defined by

$$H_n(s_1, \ldots, s_n) = \int \cdots \int h_n(\tau_1, \ldots, \tau_n) e^{-(s_1\tau_1 + \cdots + s_n\tau_n)} \, d\tau_1 \cdots d\tau_n.$$

We call H_n the nth *frequency-domain kernel* or (just) *kernel* of the operator N. Since $h_n \in \mathcal{B}^n$, H_n is defined *at least* in $\mathbb{C}_+^n := \{s \in \mathbb{C}^n : \forall k \ \mathrm{Re}\, s_k > 0\}$. H_n is symmetric, bounded, and uniformly continuous there; it is analytic in $\mathbb{C}_+^n$. We should mention that the *Unicity theorem* for Laplace transform tells us that two measures in $\mathscr{B}^n$ are equal ($h_n = g_n$) if and only if their Laplace transforms are equal ($H_n = G_n$).

The formulas of Section 3 are somewhat simpler in the frequency domain. The following theorem uses the notational convention that C_n denotes the nth frequency-domain kernel of a Volterra series operator C.

THEOREM 4.2.1 *Suppose A and B are Volterra series operators. Then the frequency-domain kernels of $A+B$, $A \cdot B$ (pointwise product), and AB (composition) are given by:*

$$(A+B)_n(s_1, \ldots, s_n) = A_n(s_1, \ldots, s_n) + B_n(s_1, \ldots, s_n),$$

$$(A \cdot B)_n = \sum_{k=1}^{n-1} A_k \vee B_{n-k} := \text{sym} \sum_{k=1}^{n-1} A_k(s_1, \ldots, s_k) B_{n-k}(s_{k+1}, \ldots, s_n),$$

and

$$(AB)_n(s_1, \ldots, s_n) = \text{sym} \sum_{k=1}^{n} \left[\sum_{\substack{i_1, \ldots, i_k \geq 1 \\ i_1 + \cdots + i_k = n}} \right] A_k(s_1 + \cdots + s_{i_1}, \ldots, s_{n-i_k+1} + \cdots + s_n) \cdot B_{i_1}(s_1, \ldots, s_{i_1}) \cdots B_{i_k}(s_{n-i_k+1}, \ldots, s_n)$$

respectively.

These well-known formulas follow easily from the formulas of Section 3.

4.3 *Multitone Inputs; the Fundamental Frequency-domain Formula*

We start with a simple calculation. Suppose that $u(t)$ is a *trigonometric polynomial*, that is

$$u(t) = \sum_{k=-l}^{l} \alpha_k e^{j\omega kt}$$

where $\alpha_{-k} = \alpha_k^*$.† Suppose also that $\|u\| < \rho = \text{rad } N$. Then

$$y(t) = Nu(t) = \sum_{n=1}^{\infty} \int \cdots \int h_n(\tau_1, \ldots, \tau_n) \prod_{i=1}^{n} \sum_{k=-l}^{l} \alpha_k e^{j\omega(t-\tau_i)} \, d\tau_i \tag{4.5a}$$

$$= \sum_{n=1}^{\infty} \Bigg(\sum_{-l \leq k_1, \ldots, k_n \leq l} \Bigg) \alpha_{k_1} \cdots \alpha_{k_n} H_n(j\omega k_1, \ldots, j\omega k_n) e^{j(\omega k_1 + \cdots + \omega k_n)t} \tag{4.5b}$$

The term $\alpha_{k_1} \cdots \alpha_{k_n} H_n(j\omega k_1, \ldots, j\omega k_n) e^{j(\omega k_1 + \cdots + \omega k_n)t}$ is often called an nth-order $(\omega k_1, \ldots, \omega k_n)$ intermodulation product. Since it is proportional to $H_n(j\omega k_1, \ldots, j\omega k_n)$ this suggests the interpretation of $H_n(j\omega k_1, \ldots, j\omega k_n)$ as a *measure* of the $(\omega k_1, \ldots, \omega k_n)$ intermodulation distortion of N.

Now we have already seen that the *first* sum in (4.5b) is an ℓ^1 sum, i.e. absolutely convergent. In fact for each t,

$$\sum_{n=1}^{\infty} \left| \Bigg(\sum_{-l \leq k_1, \ldots, k_n \leq l} \Bigg) \alpha_{k_1} \cdots \alpha_{k_n} H_n(j\omega k_1, \ldots, j\omega k_n) e^{j(\omega k_1 + \cdots + \omega k_n)t} \right| \leq f(\|u\|)$$

† Thus u is real. Complex signals are easily handled but less useful in the study of Volterra series operators than of linear operators.

where f is the gain bound function of N. Consequently we may evaluate the mth Fourier coefficient of y

$$\hat{y}(m) := \frac{\omega}{2\pi}\int_0^{2\pi\omega^{-1}} y(t)e^{-j\omega mt}\,dt$$

inside the *first* sum in (4.5b) as:

$$\hat{y}(m) = \sum_{n=1}^{\infty}\frac{\omega}{2\pi}\int_0^{2\pi\omega^{-1}}\Big(\sum_{-l\le k_1,\ldots,k_n\le l}\Big)\alpha_{k_1}\cdots$$
$$\cdots\alpha_{k_n}H_n(j\omega k_1,\ldots,j\omega k_n)e^{j(\omega k_1+\cdots+\omega k_n)t}e^{-j\omega mt}\,dt.$$

Each integral is easily evaluated (the integrands are *trigonometric polynomials*) yielding

$$\hat{y}(m) = \sum_{n=1}^{\infty}\Big(\sum_{k_1+\cdots+k_n=m}\Big)\hat{u}(k_1)\cdots\hat{u}(k_n)H_n(j\omega k_1,\ldots,j\omega k_n) \qquad (4.6)$$

since $\hat{u}(k)=\alpha_k$ for $|k|\le l$ and 0 for $|k|>l$ (and thus the inner sum in (4.6) is *finite*). We call (4.6) the *fundamental frequency-domain formula* since it expresses the output spectrum in terms of the input spectrum. Of course we have only established it for inputs which are trigonometric polynomials, but we will see that it is true for more general periodic inputs, and an analogous formula holds for *almost periodic* inputs as well (see Appendix A5).

Remark. Suppose a trigonometric polynomial signal u is passed through a unit nth power law device so that $y(t)=u(t)^n$. Then

$$\hat{y}(m) = \hat{u}^{\star n}(m) := \Big(\sum_{k_1+\cdots+k_n=m}\Big)\hat{u}(k_1)\cdots\hat{u}(k_n),$$

where $\hat{u}^{*n}$ means the n-fold convolution $\hat{u}\star\hat{u}\star\cdots\star\hat{u}$ (the sum in the convolution is finite here!). The first equality makes sense: it is just the *dual* of the correspondence between *convolution in the time domain* and *multiplication in the frequency domain.* The second equality makes the fundamental formula (4.6) seem quite natural; the nth of (4.6) can be thought of as an n-fold convolution power of $\hat{u}$, weighted by $H_n(j\omega k_1,\ldots,j\omega k_n)$.

Before establishing the fundamental formula for more general periodic inputs, we have to examine carefully the question of whether it even *makes sense* for more general periodic inputs. Despite its resemblance to the composition formula and the fact that every sum and integral encountered so far has converged absolutely, we have a surprise!

4.4 *The Fundamental Formula does not Converge Absolutely*

Remarkably the fundamental formula is *not* absolutely convergent even for u as simple as a two-tone input signal! That is

$$\sum_{n=1}^{\infty}\Big[\sum_{k_1+\cdots+k_n=m}\Big]|\hat{u}(k_1)\cdots\hat{u}(k_n)H_n(j\omega k_1,\ldots,j\omega k_n)|$$

can equal ∞ even in the case considered above: u a trigonometric polynomial (but our calculation was correct).

Remark. Practically, this means that we cannot arbitrarily rearrange the terms in the sum above. We must first perform the inner (bracketed) sum (which in this case is a finite sum), and then perform the outer sum over n.

EXAMPLE Let $u=\frac{5}{9}(\cos t+\sin 2t)$. Then $\|u\|$ can be shown to be

$$\frac{5}{9}\left(\frac{15+\sqrt{33}}{32}\right)^{\frac{1}{2}}\left(\frac{3+\sqrt{33}}{4}\right)$$

which is about $0{\cdot}978<1$. Let N be the memoryless operator with $H_n=1$ for all n, that is $y(t)=u(t)/[1-u(t)]$. Then $\rho=1$ so $y(t)$ makes sense and satisfies $\|y\|\leq\|u\|/(1-\|u\|)$ (which is about 45). According to the fundamental formula (4.6) of the last subsection, the D.C. term of y is given by:

$$\hat{y}(0)=\sum_{n=1}^{\infty}\Bigg(\sum_{k_1+\cdots+k_n=0}\Bigg)\hat{u}(k_1)\cdots\hat{u}(k_n). \tag{4.7}$$

Now we claim that (4.7) does not converge absolutely. To see this,

$$\sum_{n=1}^{\infty}\Bigg(\sum_{k_1+\cdots+k_n=0}\Bigg)|\hat{u}(k_1)\cdots\hat{u}(k_n)|\geq\sum_{n\text{ even}}\Bigg(\sum_{k_1+\cdots+k_n=0}\Bigg)|\hat{u}(k_1)\cdots\hat{u}(k_n)|$$

$$=\sum_{n\text{ even}}\Bigg(\sum_{k_1+\cdots+k_n=0}\Bigg)\hat{v}(k_1)\cdots\hat{v}(k_n)$$

where $v(t):=\frac{5}{9}(\cos t+\cos 2t)$ so that $|\hat{u}(k)|=\hat{v}(k)$ for all k

$$=\sum_{n\text{ even}}\frac{1}{2\pi}\int_0^{2\pi}v(t)^n\,dt\geq\sum_{n\text{ even}}\frac{1}{4\pi}=\infty$$

since $v(t)\geq 1$ for $-0{\cdot}25\leq t\leq 0{\cdot}25$. Thus the fundamental formula is not absolutely convergent in this simple case.

It is surprising that the trouble in (4.6) occurs when the input is a simple trigonometric polynomial signal; we might expect it to give us trouble only when, say, u does not have an absolutely convergent Fourier series.

There is one obvious but rare case in which (4.6) does converge absolutely. Suppose $\hat{u}\in\ell^1$, i.e. u has an absolutely convergent Fourier series, and in addition $f(\|\hat{u}\|_1)<\infty$. Then $|\hat{u}|\in\ell^1$ and $|\hat{u}|^{\star n}\in\ell^1$ with $\|\,|\hat{u}|^{\star n}\|_1\leq\|\hat{u}\|_1^n$, thus we have the estimate

$$\sum_{n=1}^{\infty}\Bigg(\sum_{k_1+\cdots+k_n=m}\Bigg)|\hat{u}(k_1)\cdots\hat{u}(k_n)H_n(j\omega k_1,\ldots,j\omega k_n)|\leq\sum_{n=1}^{\infty}\|\hat{u}\|_1^n\|h_n\|=f(\|\hat{u}\|_1).$$

In conclusion, then, we must proceed with extreme care in establishing the fundamental formula for more general periodic input signals.

4.5 *Proof of Fundamental Formula for General Inputs*

We start with some calculations. Suppose u is any periodic input with $\|u\|<\rho$. Recall that the lth *Cesaro* sum of the Fourier series of u is defined by

$$u_l(t)=\sum_{k=-l}^{l}\left(1-\frac{|k|}{l}\right)\hat{u}(k)e^{j\omega kt}=:\sum_{k=-l}^{l}C_l(k)\hat{u}(k)e^{j\omega kt}.$$

u_l is u convolved with an approximate identity and thus satisfies $\|u_l\|\leq\|u\|$ and $\|u_l-u\|_1\to 0$ as $l\to\infty$ (Helson, 1983). From the first fact we conclude that Nu_l makes sense since $\|u_l\|\leq\|u\|<\rho=\text{rad }N$. Using the Incremental Gain theorem for L^1 (Theorem 2.3.5), we conclude that $\|Nu-Nu_l\|_1\to 0$ as $l\to\infty$. Hence $\hat{y}_l$ ($y:=Nu$) converges uniformly to $\hat{y}$ as $l\to\infty$. u_l is a trigonometric polynomial so we know the fundamental formula holds for Nu_l; putting all this together we have shown

$$\hat{y}(m)=\lim_{l\to\infty}\sum_{n=1}^{\infty}\left(\sum_{k_1+\cdots+k_n=m}\right)\hat{u}_l(k_1)\cdots\hat{u}_l(k_n)H_n(j\omega k_1,\ldots,j\omega k_n) \tag{4.8}$$

$$=\sum_{n=1}^{\infty}\lim_{l\to\infty}\left(\sum_{k_1+\cdots+k_n=m}\right)\hat{u}(k_1)\cdots\hat{u}(k_n)\prod_{i=1}^{n}C_l(k_i)H_n(j\omega k_1,\ldots,j\omega k_n). \tag{4.9}$$

The dominated convergence theorem justifies the interchange of limit and sum in (4.8) since as we have mentioned before the first sum in (4.8) and (4.9) is always absolutely convergent and $|C_l|\leq 1$. Since $\lim_{l\to\infty}C_l(k)=1$ for each k, if we knew that the inner sum also converged absolutely we could apply dominated convergence once again to conclude

$$=\sum_{n=1}^{\infty}\left(\sum_{k_1+\cdots+k_n=m}\right)\hat{u}(k_1)\cdots\hat{u}(k_n)H_n(j\omega k_1,\ldots,j\omega k_n) \tag{4.10}$$

which would establish the fundamental formula in the general case.

Unfortunately the inner sum

$$\left(\sum_{k_1+\cdots+k_n=m}\right)\hat{u}(k_1)\cdots\hat{u}(k_n)H_n(j\omega k_1,\ldots,j\omega k_n) \tag{4.11}$$

is *not* always absolutely convergent (and thus does not always make sense). In such a case formula (4.9) is as close as we can get to the fundamental formula. But in fact the inner sum *is* absolutely convergent in almost all situations arising in engineering. We now present two conditions which suffice:

LEMMA 4.5.1 *Suppose u has bounded variation over one period. Then*

$$\left(\sum_{k_1+\cdots+k_n=m}\right)|\hat{u}(k_1)\cdots\hat{u}(k_n)|<\infty.$$

In particular, the inner sum (4.11) in the fundamental formula converges absolutely. The proof is given in Appendix A4.

LEMMA 4.5.2 *Suppose that*

$$H_n(j\omega k_1, \ldots, j\omega k_n) = O\left(\frac{1}{k_1 \cdots k_n}\right).$$

Then the inner sum (4.11) is absolutely convergent.

Remark. This condition can be interpreted as: N is *strictly proper*. For example the kernels of the input/output operator of a dynamical system with vector field affine in the input have this property.

The proof is in Appendix A4. We summarize the results of this section in

THEOREM 4.5.3 (Fundamental frequency-domain formula) *Suppose* $\|u\| < \rho$ *and that either*

(i) *the input* u *has bounded variation over one period, or*
(ii) *the operator* N *is strictly proper, that is,*

$$H_n(j\omega k_1, \ldots, j\omega k_n) = O\left(\frac{1}{k_1, \ldots, k_n}\right).$$

Then the fundamental frequency-domain formula is valid, that is:

$$\hat{y}(m) = \sum_{n=1}^{\infty} \left(\sum_{k_1+\cdots+k_n=m} \right) \hat{u}(k_1) \cdots \hat{u}(k_n) H_n(j\omega k_1, \ldots, j\omega k_n)$$

Proof. Theorem 4.5.3 follows from the discussion at the beginning of this section and the lemmas above. □

5. Acknowledgement

The authors would like to thank Professors S. S. Sastry, D. J. Newman, M. Hasler, and Dr C. Flores for helpful suggestions. Mr Boyd gratefully acknowledges the support of the Fannie and John Hertz Foundation.

REFERENCES

d'ALESSANDRO, P., ISIDORI, A. & RUBERTI, A. 1974 Realization and structure theory of bilinear systems. *SIAM J. Contr.* **12,** no. 3, 517–535.

BALAKRISHNAN, A. V. 1976 *Applied Functional Analysis.* Springer-Verlag, New York.

BARRETT, J. F. 1963 The use of functionals in the analysis of nonlinear physical systems. *J. electron. Contr.* **15,** 567–615.

BOYD, S. & CHUA, L. O. 1984 Approximating nonlinear operators with Volterra series. *to appear.*

BRILLIANT, M. B. 1958 Theory of the analysis of nonlinear systems. Report RLE-345, MIT.

BROCKETT, R. W. 1976 Volterra series and geometric control theory. *Automatica* **12,** 167–76.

BROCKETT, R. W. 1977 Convergence of Volterra series on infinite intervals and bilinear approximations. in Lakshmikantham, V. (ed.) *Nonlinear Systems and Applications.* Academic Press, New York.

BROMWICH, T. J. I'A. 1908 *Introduction to the Theory of Infinite Series.* MacMillan & Co., New York.

CHRISTENSEN, G. S. 1968 On the convergence of Volterra series. *IEEE Trans. autom. Contr.* (*Correspondence*) **AC-13,** 736–7.
CORDUNEANU, C. 1968 *Almost Periodic Functions*, Interscience Publishers, New York.
DE FIGUEIREDO, R. 1983 A generalized Fock space framework for nonlinear system and signal analysis. *IEEE Trans. Circuits Syst.* **CAS-30,** 637–8.
DESOER, C. A. & VIDYASAGAR, M. 1975 *Feedback Systems: Input/Output Properties*. Academic Press, New York.
DIEUDONNE, J. 1969 *Foundations of Modern Analysis*. Academic Press, New York.
FLIESS, M., LAMNABHI, M., & LAMNABHI-LAGARRIQUE, F. 1983 An algebraic approach to nonlinear functional expansions. *IEEE Trans. Circuits Syst.* **CAS-30,** 554–571.
HALME, A. & ORAVA, J. 1972 Generalized polynomial operators for nonlinear system analysis, *IEEE Trans. Autom. Contr.* AC-, 226–8.
HALME, A., ORAVA, J., & BLOMBERG, H. 1971 Polynomial operators in nonlinear systems theory. *Int. J. Systems Sci.* **2,** 25–47.
HELSON, H. 1983 *Harmonic Analysis*. Addison Wesley, Reading, MA.
LESIAK, C. & KRENER, A. J. 1978 Existence and uniqueness of Volterra series. *IEEE Trans. Automat. Contr.* **AC-23,** 1090–5.
RAO, R. S. 1970 On the Convergence of Discrete Volterra Series. *IEEE Trans. autom. Contr.* (*Correspondence*) **AC-15,** 140–1.
RUDIN, W. 1974 *Real and Complex Analysis*. McGraw-Hill, New York.
RUGH, W. J. 1981 *Nonlinear System Theory: The Volterra/Wiener Approach*. Johns Hopkins Univ. Press, Baltimore.
SANDBERG, I. W. 1983*a* Series expansions for nonlinear systems. *Circuits, Syst. and Signal Process.* **2,** 77–87.
SANDBERG, I. W. 1983*b* Volterra-like expansions for solutions of nonlinear integral equations and nonlinear differential equations. *IEEE Trans. Circuits Syst.* **CAS-30,** 68–77.
SANDBERG, I. W. 1984 Existence and evaluation of almost periodic steady state responses of mildly nonlinear systems, *to be published.*
SCHETZEN, M. 1980 *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley, New York.
SONTAG, E. D. 1979 *Polynomial Response Maps*. Springer Verlag, Berlin.
WIENER, N. 1964 *Generalized Harmonic Analysis and Tauberian Theorems*. MIT Press, Cambridge, MA.
ZADEH, L. 1953 A contribution to the theory of nonlinear systems. *J. Franklin Inst.* **255,** 387–408.

Appendices

A1. Volterra-like series

In the study of (linear) convolution operators in engineering it is common to consider only a subalgebra of the bounded measures, for example the subalgebra of measures lacking singular continuous part (Desoer & Vidyasager, 1975). This algebra is large enough to capture all of the commonly occurring distributed systems such as distributed transmission lines, transport delays in control systems, etc. Similarly in the study of Volterra series operators only certain types of measures occur in practice; the singular kernel $1_H(\tau_1)1_H(\tau_2)e^{-\tau_1}\delta(\tau_1-\tau_2)$ of Example 2 of Section 2.1 is typical. Sandberg (1983*b*) calls series with kernels of this form *Volterra-like*; an early occurrence of this idea was in Zadeh's (1953) paper.

In a Volterra-like series we index the series not by the order n but by a multi-index $\boldsymbol{n}=(n_1,\ldots,n_k)$ $(n_i>0)$. k is called the *length* of $\boldsymbol{n}$; the *degree* of $\boldsymbol{n}$ is defined by $\partial \boldsymbol{n} = n_1 + \cdots + n_k$. We write

$$Nu(t) = \sum_{\boldsymbol{n}} y_{\boldsymbol{n}}(t)$$

where

$$y_{\boldsymbol{n}}(t) = \int \cdots \int h_{\boldsymbol{n}}(\tau_1, \ldots, \tau_k) u(t-\tau_1)^{n_1} \cdots u(t-\tau_k)^{n_k} \, d\tau_1 \cdots d\tau_k$$

in which the kernels $h_{\boldsymbol{n}}$ are now ordinary L^1 *functions* instead of bounded measures. Each Volterra-like kernel $h_{\boldsymbol{n}}$ can be turned into an equivalent Volterra kernel $h_{[\boldsymbol{n}]}$ by:

$$h_{[\boldsymbol{n}]}(\tau_1, \ldots, \tau_n) := \text{sym}\, h_{\boldsymbol{n}}(\tau_1, \tau_{n_1+1}, \ldots, \tau_{n-n_k+1}) \delta(\tau_1 - \tau_2) \cdots \\ \cdots \delta(\tau_{n_1-1} - \tau_{n_1}) \cdots \delta(\tau_{n-1} - \tau_n).$$

We call $h_{[\boldsymbol{n}]}$ the *associated Volterra kernel* of the Volterra-like kernel $h_{\boldsymbol{n}}$. Collecting the associated Volterra kernels by degree,

$$h_n := \sum_{\partial \boldsymbol{n} = n} h_{[\boldsymbol{n}]} \tag{A1.1}$$

yields a Volterra series equivalent to the Volterra-like series. Via this associated Volterra series, Volterra-like series inherit the concepts of gain bound function and radius of convergence.

Note that $h_{[\boldsymbol{n}]}$ is supported on the k-dimensional set given by†

$$C_{\boldsymbol{n}} := \{(\tau_1, \ldots, \tau_n)^T : n_1 \text{ of the } \tau \text{ are } x_1; \ldots; n_k \text{ of the } \tau \text{ are } x_k\}.$$

Thus the associated kernel is singular (with respect to Lebesgue measure) unless $\boldsymbol{n} = (1, \ldots, 1)$.

We extend the notion of sym to Volterra-like series by:

$$\text{sym}\, h_{\boldsymbol{n}}(\tau_1, \ldots, \tau_k) = \frac{1}{k!} \sum_{\sigma \in S^k} h_{\sigma \boldsymbol{n}}(\tau_{\sigma 1}, \ldots, \tau_{\sigma k})$$

where $\sigma \boldsymbol{n} = (n_{\sigma 1}, \ldots, n_{\sigma k})$; we say $h_{\boldsymbol{n}}$ is symmetric if sym $h_{\boldsymbol{n}} = h_{\boldsymbol{n}}$. This agrees with our earlier notation if we think of the old order n as the n-long multi-index $(1, \ldots, 1)$, since $\sigma(1, \ldots, 1) = (1, \ldots, 1)$. Note that sym $h_{\boldsymbol{n}}$ involves not just the Volterra-like kernel $h_{\boldsymbol{n}}$ but all Volterra-like kernels of the form $\boldsymbol{m} = \sigma \boldsymbol{n}$. We say $\boldsymbol{m}$ and $\boldsymbol{n}$ have the same *type* in this case. A Volterra-like series thus has $P(n)$ different types of kernels of degree n, where $P(n)$ is the number of partitions of n.‡ If the Volterra-like series is symmetric then the kernels of the same type have identical associated kernels and are simply related by:

$$h_{\boldsymbol{m}}(\tau_1, \ldots, \tau_k) = h_{\sigma \boldsymbol{n}}(\tau_1, \ldots, \tau_k) = h_{\boldsymbol{n}}(\tau_{\sigma 1}, \ldots, \tau_{\sigma k}).$$

This extension of sym will also be useful in the study of multi-input Volterra series.

† We appeal to the reader's intuitive notion of dimension, but it can be shown that the Hausdorff-Besicovitch dimension of $C_{\boldsymbol{n}}$ is indeed k.

‡ There is no nice formula for $P(n)$. For those interested it is asymptotic to $(4\sqrt{3}n)^{-1} \exp \pi\sqrt{(2n/3)}$.

THEOREM A1.1 (Uniqueness theorem for Volterra-like series) *Suppose N and M are Volterra-like series operators with kernels h_n and g_n, respectively. Then $N = M$ iff* $\operatorname{sym} h_n = \operatorname{sym} g_n$ *for all n.*

Proof. The 'if' part is clear. By the Uniqueness theorem (Theorem 2.5.2) we know $h_n = g_n$, where h_n and g_n are the kernels of the associated Volterra series (given in (A1.1) above). We will finish the proof by showing that h_n determines the Volterra-like kernels $\operatorname{sym} h_n$.

THEOREM A1.2 (Decomposition theorem for Volterra-like series) *Suppose h_n are the kernels of the Volterra series associated with a Volterra-like series with kernels h_n. Then h_n uniquely determines the Volterra-like kernels* $\operatorname{sym} h_n$.

Thus if a Volterra series comes from a Volterra-like series, then each kernel can be uniquely decomposed into the 2^{n-1} symmetric Volterra-like kernels with which it is associated. Another way to think of the Decomposition theorem is: the (linear) map of the symmetric Volterra-like kernels into the associated Volterra kernels (given by formula A1.1) is *injective.*

Before starting the proof, let us consider a simple example which illustrates the idea. The second kernel of the associated Volterra series is:

$$h_2(\tau_1, \tau_2) = h_{[(1,1)]} + h_{[(2)]} = h_{(1,1)}(\tau_1, \tau_2) + \tfrac{1}{2}h_{(2)}(\tau_1)\delta(\tau_1 - \tau_2) + \tfrac{1}{2}h_{(2)}(\tau_2)\delta(\tau_1 - \tau_2).$$

Decomposing h_2 is easy: the terms $h_{[(2)]}$ and $h_{[(1,1)]}$ are *mutually singular measures* (The first is supported on the line $\tau_1 = \tau_2$ and the second is absolutely continuous.) To be quite explicit we have the formulas:

$$h_{(1,1)}(\tau_1, \tau_2) = h_2(\tau_1, \tau_2) \quad \text{for} \quad \tau_1 \neq \tau_2$$

and

$$h_{(2)}(\tau) = \lim_{\varepsilon \to 0} \frac{1}{2\sqrt{2}\varepsilon} \int_{\tau-\varepsilon}^{\tau+\varepsilon} \int_{\tau-\varepsilon}^{\tau+\varepsilon} h_2(\tau_1, \tau_2)\, d\tau_1\, d\tau_2.$$

The proof of the Decomposition theorem uses the same idea: the associated kernels of h_n and h_m are mutually singular unless n and m are of the same type. To prove this, note that the associated kernel of h_n has all its mass in the set

$$C_n^* = \{(\tau_1, \ldots, \tau_n)^T : n_1 \text{ of the } \tau \text{ are } x_1; \ldots; n_k \text{ of the } \tau \text{ are } x_k;\ i \neq j \Rightarrow x_i \neq x_j\}.$$

This is no more than the assertion that

$$\int \cdots \int h_n(\tau_1, \ldots, \tau_k) u(t-\tau_1)^{n_1} \cdots u(t-\tau_k)^{n_k}\, d\tau_1 \cdots$$
$$\cdots d\tau_k = \int \cdots \int h_n(\tau_1, \ldots, \tau_k) u(t-\tau_1)^{n_1} \cdots u(t-\tau_k)^{n_k}\, d\tau_1 \cdots d\tau_k$$

(remember that h_n is an L^1 function).

The sets C_n^* and C_m^* are disjoint if n and m are different type, and equal if the types are the same. This establishes the claim that the associated kernels are mutually singular unless the multi-indices are of the same type. The L^1 function

$h_n(\tau_1, \ldots, \tau_k)$ is determined by the integrals

$$\int \cdots \int h_n(\tau_1, \ldots, \tau_k) u_1(\tau_1) \cdots u_k(\tau_k) \, d\tau_1 \cdots d\tau_k \tag{A1.2}$$

where the u_i $(i = 1, \ldots, k)$ are in L^∞. According to the discussion above we have

$$\int \cdots \int h_n(\tau_1, \ldots, \tau_n) u_1(\tau_1) \cdots u_1(\tau_{n_1}) u_2(\tau_{n_1+1}) \cdots u_k(\tau_n) \, d\tau_1 \cdots$$

$$\cdots d\tau_n = K \int \cdots \int h_{\mathbf{n}}(\tau_1, \ldots, \tau_k) u_1(\tau_1) \cdots u_k(\tau_k) \, d\tau_1 \cdots d\tau_k$$

where K is the number of Volterra-like kernels with the same type as $\mathbf{n}$. Thus the integrals (A1.2), and hence the function $h_{\mathbf{n}}$, are determined by h_n. This proves the Decomposition theorem.

Remark. The Decomposition theorem is not so obvious as it might seem. For example consider the consequence that (nonzero) operators of the form

$$y(t) = \int\int h_{(2,2)}(\tau_1, \tau_2) u(t-\tau_1)^2 u(t-\tau_2)^2 \, d\tau_1 \, d\tau_2$$

can never be put in the form

$$y(t) = \int\int h_{(1,3)}(\tau_1, \tau_2) u(t-\tau_1) u(t-\tau_2)^3 \, d\tau_1 \, d\tau_2.$$

This is so even though the associated kernels are both supported on two-dimensional sets. The frequency-domain version of the example above is as follows. Suppose $H_{(2,2)}(s_1, s_2)$ and $H_{(1,3)}(s_1, s_2)$ are the Laplace transforms of symmetric functions in $L^1(\mathbb{R}^2_{\geq 0})$. The Decomposition theorem says we can extract $H_{(2,2)}$ and $H_{(1,3)}$ from the fourth-order frequency-domain kernel

$$H_4(s_1, \ldots, s_4) = \text{sym}\,[H_{(2,2)}(s_1+s_2, s_3+s_4) + H_{(1,3)}(s_1, s_2+s_3+s_4)]$$

(which has nine terms!). There are explicit formulas which effect this decomposition, but we will not give them here.

COROLLARY A1.3 *If the $h_{\mathbf{n}}$ are symmetric, then*

$$\|h_n\| = \sum_{\partial \mathbf{n} = n} \|h_{\mathbf{n}}\|.$$

Thus the gain bound function, which we originally defined via the associated Volterra series, is simply given by:

$$f(x) = \sum_{\mathbf{n}} \|h_{\mathbf{n}}\| \, x^{\partial \mathbf{n}}.$$

A2. Incremental Gain theorem for L^p

To demonstrate the difficulty of a theory of Volterra series operators for L^p $(p < \infty)$ which is unadulterated by reference to $\|u\|_\infty$, consider just the memoryless operator $Nu(t)=f[u(t)]$. If N is to be defined on any open subset of L^p then we must have rad $N=\rho=\infty$. It is not hard to show that N maps L^p back into L^p if and only if f is sector bounded, i.e. $|f(x)|\leq K|x|$. Sandberg (1984) has recently shown that if N has a Frechet derivative at 0 (as an operator from L^p into L^p) then f is in fact *linear*!

LEMMA 2.3.4

$$\|N(u+v)-Nu\|_p\leq\|v\|_p\frac{f(\|u\|+\|v\|)-f(\|u\|)}{\|v\|}\leq\|v\|_p f'(\|u\|+\|v\|).$$

(*Remember that unmarked norms are ∞-norms.*)

Proof. The conclusion is, if anything, sharpened if we assume the kernels are symmetric (see Section 2.5) so we will assume they are. Then

$$\begin{aligned}[N(u+v)-Nu](t)&\\ &=\sum_{n=1}^{\infty}\int\cdots\int h_n(\tau_1,\ldots,\tau_n)\Big(\prod_{i=1}^{n}(u+v)(t-\tau_i)-\prod_{i=1}^{n}u(t-\tau_i)\Big)\,d\tau_1\cdots d\tau_n\\ &=\sum_{n=1}^{\infty}\int\cdots\int h_n(\tau_1,\ldots,\tau_n)\sum_{k=1}^{n}\binom{n}{k}\prod_{i=1}^{k}v(t-\tau_i)\,d\tau_i\prod_{i=k+1}^{n}u(t-\tau_i)\,d\tau_i.\end{aligned}$$

Thus

$$\begin{aligned}|N(u+v)-Nu|\,(t)&\\ &\leq\sum_{n=1}^{\infty}\sum_{k=1}^{n}\binom{n}{k}\|v\|^{k-1}\|u\|^{n-k}\int\Big[\int\cdots\int|h_n(\tau_1,\ldots,\tau_n)|\,d\tau_2\cdots d\tau_n\Big]|v(t-\tau_1)|\,d\tau_1.\end{aligned}$$

As in Theorem 2.3.4 the bracketed expression is a measure in τ_1 with norm $\|h_n\|$, so we have (Desoer & Vidyasagar, 1975)

$$\|N(u+v)-Nu\|_p\leq\|v\|_p\sum_{n=1}^{\infty}\|h_n\|\sum_{k=1}^{n}\binom{n}{k}\|v\|^{k-1}\|u\|^{n-k}=\|v\|_p\frac{f(\|u\|+\|v\|)-f(\|u\|)}{\|v\|}.$$

The last inequality in the conclusion of Lemma 2.3.4 follows from the Mean-value theorem.

A3. Taylor series which are not Volterra series

In Section 2.5 we showed that the Volterra series operators are simply Taylor series of TI operators $:L^\infty\to L^\infty$, but noted that the Volterra series are not *all* of the Taylor series. In this section we discuss this point in more detail.

Much of the theory of Volterra series holds for the more general Taylor series

$$Nu=\sum_{n=1}^{\infty}P_n(u)=\sum_{n=1}^{\infty}M_n(u,\ldots,u)$$

where M_n is the bounded TI n-linear map $:L^\infty\to L^\infty$ given by $M_n=(n!)^{-1}D^{(n)}N(0)$. With the gain bound function $f(x)=\sum\|M_n\|x^n$ only notational changes are required to prove all the results of Section 3. For example, such an N has a Taylor series inverse near 0 if and only if M_1 is invertible.

The differences between our formulation of Volterra series and a more general formulation based on Taylor series are as follows.

(i) Not all bounded TI n-linear maps $:\mathrm{L}^{\infty n} \to \mathrm{L}^{\infty}$ have a convolution representation

$$M_n(u_1, \ldots, u_n) = \int \cdots \int h_n(\tau_1, \ldots, \tau_n) u_1(t-\tau_1) \cdots u_n(t-\tau_n)\, d\tau_1 \cdots d\tau_n \tag{A3.1}$$

with $h_n \in \mathscr{B}^n$.

(ii) The norm we use, $\|h_n\|$, is *not* equivalent to the norm $\|\cdot\cdot\|_{\text{ml}}$ on $\mathscr{L}_n(\mathrm{L}^{\infty}, \mathrm{L}^{\infty})$, it is stronger (larger). That is (with some abuse of notation)

$$\|h_n\|_{\text{ml}} := \sup_{\|u_i\|<1} \left\| \int \cdots \int h_n(\tau_1, \ldots, \tau_n) u(t-\tau_1) \cdots u(t-\tau_n)\, d\tau_1 \cdots d\tau_n \right\| \leq \|h_n\|$$

and the ratio of the two is not bounded away from zero. Indeed we will give an example where the ratio *is* zero.

(i) is true even for $n = 1$. We now give an example. Consider the subspace of L^{∞} of those u with a limit at $t = -\infty$, that is

$$\left\{ u \in \mathrm{L}^{\infty} : \lim_{t \to -\infty} u(t) \text{ exists} \right\}.$$

On this subspace we define $F(u)$ as $\lim_{t \to -\infty} u(t)$. F is clearly a LTI bounded functional on this subspace. Using the Hahn–Banach theorem and the Axiom of Choice, F can be extended to a LTI bounded functional on all of L^{∞}, which we denote Lim; see Rudin (1974). Lim can also be thought of as a bounded LTI operator $:\mathrm{L}^{\infty} \to \mathrm{L}^{\infty}$ (though its range is just the constants).

For any u which vanishes for $t < 0$ we have $\mathrm{Lim}\, u = 0$. This establishes that Lim is *causal*, and that Lim has no representation as a convolution with a measure. It also shows that the Steady-state theorem does not hold for Lim. To mention just one more bizarre property of Lim, it is a bounded LTI operator which maps *sinusoids* to *constants*!

Clearly this example is absurd from an engineering point of view. Lim's perfect memory of the *infinitely remote past* (and indeed, total amnesia for the *finite past*) contradicts our intuition that bounded LTI physical devices and systems should have a *decaying memory*.†

Let us now give an example of (ii). For $n > 1$, $\prod_{i=1}^{n} \mathrm{Lim}\, u_i$ furnishes an example of a bounded multilinear operator not given by a convolution as in (A3.1). Less bizarre examples can also be given for $n > 1$. For example we can have a convolution representation with h_n an *unbounded measure*.‡ Consider the kernel

$$h_2(\tau_1, \tau_2) = 1_{\mathrm{H}}(\tau_1) 1_{\mathrm{H}}(\tau_2) \frac{\cos(\tau_1 \tau_2)}{(1+\tau_1)(1+\tau_2)},$$

† Moral: don't fiddle with the Axiom of Choice.

‡ In the literature this is often stated: '$\int \cdots \int |h_n(\tau_1, \ldots, \tau_n)|\, d\tau_1 \cdots d\tau_n < \infty$ is a sufficient but not necessary condition for BIBO stability of a second-order Volterra operator.' An incorrect example is given in Schetzen (1980).

then $\|h_2\| = \int |h_2(\tau_1, \tau_2)| \, d\tau_1 \, d\tau_2 = \infty$. Nevertheless this kernel induces a bounded bilinear map $: L^{\infty 2} \to L^{\infty}$. First we have to say what we mean by the convolution since the integral in (A3.1) is not absolutely convergent with this h_2. We mean

$$M_2(u_1, u_2)(t) := \lim_{T\to\infty} \int_0^T \int_0^T h_2(\tau_1, \tau_2) u_1(t-\tau_1) u_2(t-\tau_2) \, d\tau_1 \, d\tau_2.$$

To see that this limit exists and that M_2 is bounded, we rewrite this as

$$= \lim_{T\to\infty} \int_0^\infty \left[\frac{u_2(t-\tau_2)}{1+\tau_2} 1_H(T-\tau_2)\right]\left[\mathrm{Re}\int_0^T e_0^{-j\tau_1\tau_2} \frac{u_1(t-\tau_1)}{1+\tau_1} d\tau_1\right] d\tau_2 \qquad \text{(A3.2)}$$

As $T \to \infty$ the left-hand bracketed expression in (A3.2) converges in L^2 to the L^2-function $1_H(\tau_2)u_2(t-\tau_2)/(1+\tau_2)$; by the Plancherel theorem the right-hand bracketed expression in (A3.2) converges in L^2 to the L^2-function $\mathrm{Re}\,[u_1(t-\cdot)/(1+\cdot)]^P(\tau_2)$, where by f^P we mean here the Plancherel transform of $f \in L^2$. Consequently the limit in (A3.2) exists and is bounded by

$$|M_2(u_1, u_2)(t)| \leq \left\|\frac{u_2(\cdot)}{1+(\cdot)}\right\|_2 \left\|\mathrm{Re}\left(\frac{u_1(\cdot)}{1+(\cdot)}\right)^P\right\|_2$$

$$\leq \frac{2\sqrt{(2\pi)}}{3} \|u_1\|_\infty \|u_2\|_\infty$$

which establishes $\|M_2\|_{ml} \leq 2\sqrt{(2\pi)}/3$. This example was suggested by D. J. Newman. Like the first example Lim above, it is rather forced.

There are thus at least three costs associated with generalizing Volterra series operators to arbitrary Taylor series:

(1) we lose the concrete convolution representation (A3.1);
(2) the norm $\|h_n\| = \int \cdots \int |h_n| \, d\tau_1 \cdots d\tau_n$ is replaced by $\|M_n\|_{ml}$ which is nearly impossible to compute;
(3) we include clearly nonphysical operators such as Lim.

It is the authors' feeling, and we hope the examples above have convinced the reader, that the mathematical elegance and completeness of a general Taylor series formulation is not worth (1)–(3).

A4. Absolute convergence of the inner sum

In Section 4.5 we established the fundamental frequency-domain formula under the hypothesis that

$$\left(\sum_{k_1+\cdots+k_n=m}\right) |\hat{u}(k_1) \cdots \hat{u}(k_n) H_n(j\omega k_1, \ldots, j\omega k_n)| \qquad \text{(A4.1)}$$

is finite. In this section we give two simple conditions which ensure that (A4.1) is finite, the first a condition on the input signal u, and the second a condition on the kernel H_n.

A4.1 *Conditions on the Input Signal*

We seek conditions which ensure that

$$\left(\sum_{k_1+\cdots+k_n=m}\right)|\hat{u}(k_1)\cdots\hat{u}(k_n)| \tag{A4.2}$$

is finite. This of course implies that (A4.1) is finite, since $|H_n|\leq\|h_n\|$. Note that (A4.2) is simply (A4.1) when N is the simplest possible n-order operator: the memoryless n-power-law device $Nu(t)=u(t)^n$.

Since $u\in L^\infty$, $u\in L^2$ and so $\hat{u}\in\ell^2$. Thus for $n=2$, (A4.2) is just a convolution of two sequences in ℓ^2 and thus is finite by the Cauchy–Schwarz inequality:

$$\sum_{k_1+k_2=m}|f(k_1)g(k_2)|=\sum_{k=1}^{\infty}|f(k)|\,|g(m-k)|\leq\|f\|_2\,\|g\|_2. \tag{A4.3}$$

Since the convolution of two ℓ^2-sequences is not, in general, in ℓ^2, the finiteness of (A4.2) already is dubious for $n=3$. On the other hand if $\hat{u}\in\ell^1$, then convolution iterates of u make sense and are still in ℓ^1: (A4.2) is then bounded by $\|\hat{u}\|_1^n$.

It is a remarkable fact that for most u, (A4.2) *is* finite, even when $\hat{u}$ is not in ℓ^1. This is not true for all $u\in L^\infty$; $\cos(1/t)$ (extended periodically) is a counterexample.†

THEOREM A4.1.1 *Suppose that $\hat{u}(k)=O(1/k)$. Then (A4.2) is finite, that is*

$$\left(\sum_{k_1+\cdots+k_n=m}\right)|\hat{u}(k_1)\cdots\hat{u}(k_n)|<\infty.$$

Proof. Suppose that $\hat{u}(k)=O(1/k)$. Then there is a constant β such that $|\hat{u}(k)|\leq\beta\hat{v}(k)$ where

$$\hat{v}(k):=\begin{cases}1 & (k=0),\\ 1/|k| & (k\neq0).\end{cases}$$

Since $\hat{v}\in\ell^2$, it is indeed the Fourier series of some L^2 function which we will call, surprisingly enough, v. In fact

$$v(t)=1-\ln 2-\ln(1-\cos t)$$

the verification of which we will spare the reader.

Now

$$\left(\sum_{k_1+\cdots+k_n=m}\right)|\hat{u}(k_1)\cdots\hat{u}(k_n)|\leq\beta^n\left(\sum_{k_1+\cdots+k_n=m}\right)\hat{v}(k_1)\cdots\hat{v}(k_n) \tag{A4.4}$$

so it will suffice to show that the right-hand side of (A4.4) is finite. We break up the proof of this into three lemmas.

LEMMA 1 *Suppose f and g are in L^2. Then $\widehat{fg}=\hat{f}\star\hat{g}$.*

† D. J. Newman, personal communication.

Even though this is well known we give a short proof here for completeness.

Proof. We have already seen in equation (A4.3) that the convolution $\hat{f} \star \hat{g}$ converges absolutely. Recall that (Plancherel theorem)

$$\left\| g - \sum_{k=-l}^{l} \hat{g}(k) e^{j\omega kt} \right\|_2 \to 0 \quad \text{as} \quad l \to \infty. \tag{A4.5}$$

By the Cauchy–Schwarz inequality

$$\left| \frac{\omega}{2\pi} \int_0^{2\pi\omega^{-1}} f(t)\left(g(t) - \sum_{k=-l}^{l} \hat{g}(k) e^{j\omega kt} \right) e^{-j\omega mt}\, dt \right| \leq \|f\|_2 \left\| g - \sum_{k=-l}^{l} \hat{g}(k) e^{j\omega kt} \right\|_2 \tag{A4.6}$$

By (A4.5) the right-hand side of (A4.6), and therefore the left-hand side of (A4.6), converges to 0 as $l \to \infty$. But the left-hand side of (A4.6) is just

$$\left| \widehat{fg}(m) - \sum_{k=-l}^{l} \hat{g}(k) \hat{f}(m-k) \right|.$$

Letting $l \to \infty$ yields the conclusion. □

Lemma 2 $v(t)^n \in L^1$ *for all* n. *(That is,* $v \in L^p$ *for all* $p < \infty$.)

Proof. Clearly we need only worry about the singularity at $t = 0$, that is $v(t)^n \in L^1$ if and only if $[\ln(1 - \cos t)]^n$ is integrable near $t = 0$. This is true iff $(\ln t)^n$ is integrable near $t = 0$, which is true since

$$\int_\varepsilon^1 |\ln t|^n\, dt = \int_0^{-\ln \varepsilon} e^{-x} x^n\, dx \leq n!$$

which establishes Lemma 2. □

Lemma 3

$$\left(\sum_{k_1 + \cdots + k_n = m} \right) \hat{v}(k_1) \cdots \hat{v}(k_n) = \widehat{v^n}(m). \tag{A4.7}$$

Proof. By induction on n. Suppose we have established (A4.7) for n. By Lemma 2, v^n and v are in L^2, so applying Lemma 1 we have

$$\widehat{v^{n+1}} = \widehat{v^n} \star \hat{v};$$

using the inductive hypothesis

$$\widehat{v^{n+1}}(\tilde{m}) = \sum_m \left(\sum_{k_1 + \cdots + k_n = m} \right) \hat{v}(k_1) \cdots \hat{v}(k_n) \hat{v}(\tilde{m} - m)$$

$$= \left(\sum_{k_1 + \cdots + k_{n+1} = \tilde{m}} \right) \hat{v}(k_1) \cdots \hat{v}(k_{n+1}),$$

the change of order being valid *since the summand is positive* (Fubini theorem). This completes the proof of Lemma 3. □

We can now finish the proof of Theorem A4.1.1. From (A4.4), (A4.7), and Lemma 2 we have

$$\left(\sum_{k_1+\cdots+k_n=m}\right)|\hat{u}(k_1)\cdots\hat{u}(k_n)| \leq \beta^n \widehat{v^n}(m) \leq \|v^n\|_1 < \infty$$

establishing Theorem A4.1.1. □

One useful condition which implies $\hat{u}(n)=O(1/n)$ is that u has bounded variation over one period.

LEMMA 4.5.1 *Suppose u has bounded variation over one period. Then*

$$\left(\sum_{k_1+\cdots+k_n=m}\right)|\hat{u}(k_1)\cdots\hat{u}(k_n)|<\infty.$$

Proof. If u has bounded variation over one period then $\hat{u}(n)=O(1/n)$ (Helson, 1983). (The proof is essentially integrating by parts the formula for $\hat{u}(n)$.) Thus Theorem A4.1.1 proves Lemma 4.5.1. □

A4.2 *Conditions on the kernel H_n*

LEMMA 4.5.2 *Suppose that $H_n(j\omega k_1,\ldots,j\omega k_n)=O\left(\frac{1}{k_1\cdots k_n}\right)$. Then (A4.1) is finite, that is*:

$$\left(\sum_{k_1+\cdots+k_n=m}\right)|\hat{u}(k_1)\cdots\hat{u}(k_n)H_n(j\omega k_1,\ldots,j\omega k_n)|<\infty.$$

Proof. Suppose $H_n(j\omega k_1,\ldots,j\omega k_n)=O(1/k_1\cdots k_n)$. Then

$$[H_n(j\omega k_1,\ldots,j\omega k_n)]_{k_i\in\mathbb{N}}\in\ell^2(\mathbb{N}^n).$$

Since $\hat{u}\in\ell^2$, $[\hat{u}(k_1)\cdots\hat{u}(k_n)]_{k_i\in\mathbb{N}}\in\ell^2(\mathbb{N}^n)$ with norm $\|\hat{u}\|_2^n$ so the Cauchy–Schwarz inequality yields

$$\left(\sum_{k_1+\cdots+k_n=m}\right)|\hat{u}(k_1)\cdots\hat{u}(k_n)H_n(j\omega k_1,\ldots,j\omega k_n)|$$

$$\leq \sum_{k_1,\ldots,k_n}|\hat{u}(k_1)\cdots\hat{u}(k_n)H_n(j\omega k_1,\ldots,j\omega k_n)|$$

$$\leq \|\hat{u}(k_1)\cdots\hat{u}(k_n)\|_2\,\|H_n(j\omega k_1,\ldots,j\omega k_n)\|_2=\|\hat{u}\|_2^n\,\|H_n(j\omega k_1,\ldots,j\omega k_n)\|_2$$

which proves Lemma 4.5.2. □

A5. Almost periodic inputs

Recall that τ is said to be an ε-translation number for u if $\|u(\cdot)-u(\cdot+\tau)\|\leq\varepsilon$. u is *almost periodic* if for all $\varepsilon>0$ there is an L such that all L-long intervals contain at least one ε-translation number for u. Formally

$$\forall\varepsilon>0\,\exists L\,\forall a\,\exists\tau\;(a<\tau<a+L \text{ and } \|u(\cdot)-u(\cdot+\tau)\|\leq\varepsilon).$$

These definitions and a concise discussion can be found in Wiener (1964) or Corduneanu (1968).

THEOREM A5.1 *Suppose u is almost periodic and $\|u\| < \rho = \operatorname{rad} N$. Then Nu is almost periodic.*

This extends some results of Sandberg (1984) who established Theorem A5.1 under the assumption that u has an absolutely convergent Fourier series with small enough coefficients.

Proof. Let $\varepsilon > 0$. Choose r with $\|u\| < r < \rho$. By the Incremental Gain theorem (Theorem 2.3.2) there is a K such that on B_r, $\|Nu - Nv\| \leq K\|u - v\|$. For any τ, $\|u(\cdot + \tau)\| \leq r$, hence if τ is an ε-translation number for u then

$$\|Nu(\cdot) - Nu(\cdot + \tau)\| \leq K\|u(\cdot) - u(\cdot + \tau)\| \leq K\varepsilon$$

so τ is a $K\varepsilon$-translation number for Nu.

Now to finish the proof: since u is almost periodic find L such that all L-long intervals contain at least one ε/K-translation number for u. From the discussion above these translation numbers are ε-translation numbers for Nu, thus Nu is almost periodic. □

Remark. It is not hard to show that *any continuous time-invariant operator from* L^∞ *into* L^∞ *maps almost periodic functions into almost periodic functions.* To see this, we first give a modern (less concrete) definition of almost periodic functions: u is almost periodic iff it is continuous and the set of its translates $\{u(\cdot - t) : t \in \mathbb{R}\}$ is compact in L^∞ (Corduneanu, 1968). Now if u is almost periodic and N is time-invariant and continuous mapping L^∞ to L^∞, the set of translates of Nu is $\{Nu(\cdot - t) : t \in \mathbb{R}\}$, which, being the continuous image of a compact set, is compact. Hence Nu is almost periodic.

We will now establish the analogous fundamental formula for almost periodic inputs.

THEOREM A5.2 (Fundamental frequency-domain formula for almost periodic inputs) *Suppose that u is almost periodic and $\|u\| < \rho = \operatorname{rad} N$, and in addition*

$$\left(\sum_{\omega_{k_1} + \cdots + \omega_{k_n} = \omega}\right) |\hat{u}(\omega_{k_1}) \cdots \hat{u}(\omega_{k_n}) H_n(j\omega_{k_1}, \ldots, j\omega_{k_n})| < \infty. \tag{A5.1}$$

Then for any $\omega \in \mathbb{R}$

$$\widehat{Nu}(\omega) = \sum_{n=1}^{\infty} \left(\sum_{\omega_{k_1} + \cdots + \omega_{k_n} = \omega}\right) \hat{u}(\omega_{k_1}) \cdots \hat{u}(\omega_{k_n}) H_n(j\omega_{k_1}, \ldots, j\omega_{k_n}). \tag{A5.2}$$

Proof. Due to the similarity to the case of periodic inputs, we give a shortened proof. As in Section 4.3 we first assume that the input has the form

$$u(t) = \sum_{k=-l}^{l} \alpha_k e^{j\omega_k t}$$

We will call such a u a *multitone signal.* It is easily verified that for multitone signals

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T u(t)e^{-j\nu t}\,dt = \begin{cases}\alpha_k & (\nu=\omega_k),\\ 0 & \text{otherwise.}\end{cases} \tag{A5.3}$$

The limit in (A5.3), which can be shown to exist for any almost periodic function and any $\nu \in \mathbb{R}$, is denoted $\hat{u}(\nu)$. The same argument as in Section 4.3 establishes

$$\widehat{Nu}(\omega) = \sum_{n=1}^{\infty}\Bigg(\sum_{\omega_{k_1}+\cdots+\omega_{k_n}=\omega}\Bigg)\hat{u}(\omega_{k_1})\cdots\hat{u}(\omega_{k_n})H_n(j\omega_{k_1},\ldots,j\omega_{k_n}) \tag{A5.4}$$

for the case of u a multitone signal. We now appeal to Bohr's characterization of almost periodic functions: *they are precisely the uniform limits of multitone signals* (Wiener, 1964). Thus there is a sequence of multitone signals u_l with $\|u_l\|<\rho$ and $u_l \to u$ uniformly as $l\to\infty$. By the Incremental Gain theorem $Nu_l \to Nu$ uniformly as $l\to\infty$. Hence for any ν in R $\widehat{Nu_l}(\nu)\to\widehat{Nu}(\nu)$. Since formula (A5.4) above holds for multitone signals we have

$$\widehat{Nu}(\omega) = \sum_{n=1}^{\infty}\lim_{l\to\infty}\Bigg(\sum_{\omega_{k_1}+\cdots+\omega_{k_n}=\omega}\Bigg)\hat{u}_l(\omega_{k_1})\cdots\hat{u}_l(\omega_{k_n})H_n(j\omega_{k_1},\ldots,j\omega_{k_n}). \tag{A5.5}$$

Since $u_l \to u$ uniformly, $\hat{u}_l(\omega)\to\hat{u}(\omega)$ uniformly. Dominated convergence and hypothesis (A5.1) yield

$$= \sum_{n=1}^{\infty}\Bigg(\sum_{\omega_{k_1}+\cdots+\omega_{k_n}=\omega}\Bigg)\hat{u}(\omega_{k_1})\cdots\hat{u}(\omega_{k_n})H_n(j\omega_{k_1},\ldots,j\omega_{k_n}) \tag{A5.6}$$

which is the conclusion of Theorem A5.2. □