28

SIAM J. APPL. MATH.
Vol. 33, No. 2, September 1977

THE VOLTERRA REPRESENTATION AND THE WIENER EXPANSION: VALIDITY AND PITFALLS*

G. PALM AND T. POGGIO†

Abstract. Volterra and Wiener series provide a general representation for a wide class of nonlinear systems. In this paper we derive rigorous results concerning

(a) the conditions under which a nonlinear functional admits a Volterra-like integral representation,

(b) the class of systems that admit a Wiener representation and the meaning of such a representation,

(c) some sufficient conditions providing a connection between the Volterra-like and the Wiener representations,

(d) the mathematical validity of the method of Lee and Schetzen for identifying a nonlinear system.

1. Introduction. Representation and experimental identification of the input–output behavior of a "system" is one of the basic problems in the neurosciences. Methods of linear system theory have been used extensively in the past to study biological systems. Although they are not useful in obtaining information about the structure and the mechanisms of a biological "black-box", they provide explicit representations for the functional properties of a system and experimental identification techniques, typically based on the use of sinusoidal test stimuli. Despite many failures, linear-system theory was successfully used to obtain (not many!) interesting contributions to neurophysiology. An outstanding example is represented by the analysis of the lateral inhibition network in the eye of Limulus [28].

Biological systems, however, are rarely linear, even for "small" inputs. Nonlinearities are often essential, especially in nervous subsystems that solve information processing problems. In recent years, linear-system theory has been extended to encompass a large class of nonlinear systems through the introduction of functional polynomial series representations. Volterra theory [32] can be regarded as a forerunner of the modern functional approach. Wiener's work [33] stimulated interest in functional series representations especially as a tool for identifying nonlinear biological systems. Brownian motion test inputs play the same role in Wiener's theory as sinusoidal test signals in linear-system theory. The basis of the method is that Brownian motion inputs "fill" the neighborhood of every possible input signal with a nonzero probability and therefore test efficiently the input space (see Appendix B).

L. Stark [31] applied the Wiener method and Katzenelson and Gould [12] a variation of it, to characterize the pupillary control system. A modification of the Wiener method, proposed by Lee and Schetzen [14], has recently led (together with the availability of cheap computing power) to an outburst of applications. Marmarelis [18], Marmarelis and Naka [19], [20], Naka et al. [24], McCann [21], Krausz and Friesen [13], Fishman [8], Lipson [15], von Seelen [30] applied the

* Received by the editors May 1, 1976, and in revised form August 19, 1976.

† Max-Planck-Institut für biologische Kybernetik, D 7400 Tübingen, Germany.

Lee–Schetzen or related techniques to a variety of biological systems. At the same time—and independently from Wiener's approach—Volterra-like series have been widely used to represent the input–output relations of nonlinear systems in the areas of communication (Bedrosian and Rice [3], Brilliant [5], Zames [34], Barrett [2], Bussgang [6]) and neuroscience (Poggio and Torre [27]).

However, several of these attempts, especially concerning white-noise system identification, have been limited by the lack of a satisfactory mathematical treatment. In particular, questions about range of validity and connections of the two functional representations have been left essentially unanswered. Furthermore, confusions in the literature have occasionally obscured the mathematical foundations of these functional approaches. The purpose of this paper is to attempt a mathematical clarification of some of these problems. A forthcoming note will discuss on a more general level advantages and disadvantages of these identification methods for an understanding of nervous systems.

In this paper we will try to answer the following questions:

1. Which class of nonlinear systems (mappings) admits a Volterra-like integral series representation and for which class of input functions?

2. Which class of nonlinear systems admits a Wiener representation? And in which sense is this a *valid* representation of a system?

3. What is the connection between the two representations?

4. Can the white-noise identification method of Lee and Schetzen [14] be rigorously justified?

Practical implications of these problems and of their solutions will be discussed later. We will not deal here with probabilistic questions concerning Wiener's theory [33]. A recent account of some of the problems involved has been given by McKean [23]. Extensions of the theorems of this paper to more general stochastic inputs than Brownian motion (see footnote 4) will be presented elsewhere [25].

The organization of the paper is as follows. In § 2 we introduce the functional model of a nonlinear system. Section 3 deals with the Taylor series expansion of a functional and with the existence of integral ("Volterra-like") representations. The Wiener theory is introduced in § 4, and § 5 presents a few heuristic remarks about it. Convergence and validity of the Wiener series are clarified through a few theorems (§ 6). Section 7 discusses the Lee and Schetzen identification method and a new alternative identification method. Section 8 is devoted to an interpretation of our results concerning practical applications.

We shall use the following notation concerning function spaces (compare Gel'fand and Vilenkin [10], Schwartz [29]):

$\mathbf{C}=\mathbf{C}([0,1])$ denotes the Banach space of all real-valued continuous functions on the closed real interval $[0,1]$ with the norm $\|f\|:=\sup\{|f(t)|: t\in[0,1]\}$.

$\mathbf{C}_0=\mathbf{C}_0([0,1])$ denotes the closed subspace of $\mathbf{C}([0,1])$, containing all functions which vanish at 0.

$L^2=L^2([0,1])$ denotes the Banach space of all (equivalence classes modulo null functions of) square integrable, real-valued functions on the real interval $[0,1]$ with norm $\|f\|:=(\int f^2(t)\,dt)^{1/2}$. More generally, if (X,Σ,p) is a probability space, we denote by $L^2(X,p)$ the space of all real-valued functions on X, square integrable with respect to p, i.e. $\|f\|:=(\int f^2(x)\,dp(x))^{1/2}<\infty$.

$\mathbf{D}([0, 1]^n)$ denotes the space of all real-valued infinitely differentiable functions on the subinterval $[0, 1]^n$ of $\mathbb{R}^n$, vanishing on the boundary of the interval $[0, 1]^n$.

$\mathbf{D}'([0, 1]^n)$ denotes the space of all continuous linear functionals on $\mathbf{D}([0, 1]^n)$, where the topology of $\mathbf{D}([0, 1]^n)$ is determined by the seminorms

$$p_k(f) := \sup\{|f^{(k)}(t)| : t \in [0, 1]^n\} \qquad (k = 0, 1, 2, \cdots).$$

In general, $E = E([0, 1])$ denotes any topological vector space containing $\mathbf{D}([0, 1])$ as a dense subspace, such that the usual topology on $\mathbf{D}([0, 1])$ is finer than the topology induced by E. Examples are L^2 and $\mathbf{D}'$.

The following inclusions hold:

$$\mathbf{D}([0, 1]) \subseteq \mathbf{C}_0([0, 1]) \subseteq \mathbf{C}([0, 1]) \subseteq L^2([0, 1]) \subseteq \mathbf{D}'([0, 1]).$$

2. The functional representation. The functional approach characterizes a given system S as a mapping or operator between two function spaces, the elements of which represent input and output signals. The mapping associates to each input signal a corresponding output signal (see Fig. 1). This paper is confined to the class of time- (or space-) invariant mappings with a finite memory. In this case the output value $y(t)$ at time t is given by a real-valued functional S of the history up to t of the input variable x

$$y(t) = S\{x^t\} \tag{2.1}$$

Here x^t is a "restricted" function $x^t : [0, 1] \to \mathbb{R}$ with the property that $x^t(s)$ equals the values of x at time $t - s$

$$x^t(s) = x(t - s), \qquad 0 \leqq s < 1, \tag{2.2}$$

where the characteristic memory of the system is normalized to 1 and x is the input function. Thus, for any given time t the output $y(t)$ is a functional of x^t and because of the time invariance this functional is the same for all t. Therefore, a time invariant system can be completely characterized by a functional

$$S : E([0, 1]) \to \mathbb{R}, \tag{2.3}$$

where E denotes some vector space of real-valued "functions" on $[0, 1]$. A system operating on complex valued inputs will be characterized by a functional $S : E_{\mathbb{C}}([0, 1]) \to \mathbb{C}$.

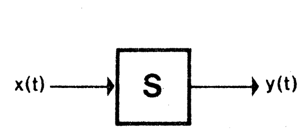

FIG. 1. *Redrawn from* [16]

3. Functional Taylor series and integral representations. Physical systems often have a characteristic smoothness property. Systems which do not respond critically to certain changes in input show a smooth dependence of the output on the input. This notion of smoothness can be mathematized by assuming that the functional $S:E([0,1])\to\mathbb{R}$ representing the system is $N+1$ times Fréchet differentiable (and that E is a Banach space). In this case

$$(3.1)\qquad Sx=\sum_{n=0}^{N} K_n x+O(\|x\|^{N+1})\quad \text{for sufficiently small } \|x\|,$$

where K_n are bounded, homogeneous polynomial functionals of degree n.[1]

In this section, we assume that $S:E([0,1])\to\mathbb{R}$ can be represented as

$$(3.2)\qquad Sx=\sum_{n=0}^{\infty} K_n x\quad \text{for sufficiently "small" } x,$$

where E is any suitable space of input functions and K_n is a bounded homogeneous functional of degree n.

A natural question is whether the various terms K_n of (3.2) can be written as Volterra-like integrals. Balakrishnan [1] and others have maintained that for $E=L^2([0,1])$ a representation like (3.5) is possible with $k_n\in L^2([0,1]^n)$, through an extension (by induction!) of the familiar Riesz theorem for linear functionals. This statement is incorrect[2] as a simple counterexample shows:

$$(3.3)\qquad K_2:x\to\int_0^1 (x(s))^2\,ds\quad \text{cannot be expressed that way.}$$

However, the use of distributions—namely Dirac's δ-function—allows one to write the functional K_2x as

$$(3.4)\qquad K_2x=\iint_0^1 \delta(s-r)x(s)x(r)\,ds\,dr$$

This is a special example of the following general result.

THEOREM 1. *Let $K_n:E([0,1])\to\mathbb{R}$ be a bounded homogeneous polynomial functional of degree n on a topological vector space E containing $\mathbf{D}([0,1])$ as a dense subspace*[3]*, such that the usual topology on $\mathbf{D}$ is finer than the topology induced by E. Then, for every $x\in E$, K_nx can be written in terms of a symbolic integral*

$$(3.5)\qquad K_nx=\int_0^1\cdots\int k_n(t_n,\cdots,t_n)x(t_1)\cdots x(t_n)\,dt_1\cdots dt_n,$$

where the kernel k_n is in $\mathbf{D}'([0,1]^n)$.

[1] In particular it is well known (e.g. Hille and Phillips, [11, pp. 112, 769]) that a Fréchet-differentiable functional $S:E_{\mathbb{C}}([0,1])\to\mathbb{C}$ (where $E_{\mathbb{C}}$ is a complex Banach space) is analytic and has a Taylor expansion $Sx=\sum_{n=0}K_nx$ which converges uniformly in a neighborhood of 0.

[2] B. Coleman mentioned this point to one of us (T.P.).

[3] In Theorem 1 the restriction to the interval $[0,1]$ is not essential: it holds true for topological vector spaces E containing $\mathbf{D}(\mathbb{R})$ or $\mathbf{S}(\mathbb{R})$ as a dense subspace. In this sense Theorem 1 holds for systems with arbitrary memory and anticipation. However, in the following sections the restriction to the finite memory case seems to be important.

Proof. Since the functional K_n occurring in (3.2) can be restricted to the subspace $\mathbf{D}([0, 1])$ of $E([0, 1])$ (and being continuous on $E([0, 1])$ is also continuous on $\mathbf{D}([0, 1])$), it has the form (3.5) (Gel'fand and Vilenkin [10, Vol. IV, Thm. 5', p. 20]). In other words it can be represented by an element k_n of $\mathbf{D}'([0, 1]^n)$. This symbolic integral representation, which at first only makes sense for input functions $x \in \mathbf{D}([0, 1])$, can be extended in a nonambiguous way to input functions $x \in E([0, 1])$, since the functionals K_n are continuous in the E sense and $\mathbf{D}$ is dense in E.

An integral representation like (3.5) is called a *symbolic integral representation* (see Appendix A), because eventually the k_n are distributions. Let us now consider the special case $E([0, 1]) = L^2([0, 1])$. Since the functionals K_n are bounded and continuous in the L^2 norm it can be shown that the "kernels" k_n are the derivatives of some functions (see Gel'fand and Vilenkin [10, Vol. IV, p. 16]). Consequently, an alternative representation for the functionals K_n is given by a *Stieltjes integral representation*

$$K_n x = \int_0^1 \cdots \int \mathbf{k}_n(t_1, \cdots, t_n)\, dx(t_1) \cdots dx(t_n), \tag{3.6}$$

where the "kernels" $\mathbf{k}_n$ are L^2 functions. Equations (3.6) and (3.5) are in fact equivalent since for $x \in \mathbf{D}$ equation (3.5) can be derived from (3.6) through partial integration:

$$\begin{aligned} K_n x &= \int_0^1 \cdots \int \mathbf{k}_n(t_1, \cdots, t_n)\, dx(t_1) \cdots dx(t_n) \\ &= \int_0^1 \cdots \int \mathbf{k}_n(t_1, \cdots, t_n)\dot{x}(t_1)\, dt_1 \cdots \dot{x}(t_n)\, dt_n \\ &= \int_0^1 \cdots \int (-1)^n \frac{\partial^n}{\partial t_1 \cdots \partial t_n} \mathbf{k}_n(t_1, \cdots, t_n) x(t_1) \cdots x(t_n)\, dt_1 \cdots dt_n, \end{aligned} \tag{3.7}$$

showing that the following relationship connects the *symbolic kernels* k_n to the *Stieltjes kernels* $\mathbf{k}_n$

$$k_n(t_1, \cdots, t_n) = (-1)^n \frac{\partial^n}{\partial t_1 \cdots \partial_n} \mathbf{k}_n(t_1, \cdots, t_n). \tag{3.8}$$

The Stieltjes representation, (3.6), equivalent to (3.5) for input functions $x \in \mathbf{D}([0, 1])$, can also be extended in a unique way to input functions $x \in L^2([0, 1])$, since $\mathbf{D}$ is dense in L^2. While (3.5) poses no difficulties, it is important to stress that (3.6) could lead to some misinterpretations when $\mathbf{k}_n$ and x are both differentiable and continuous. Consider, for instance, the simple expression

$$K_1 x = \int_0^1 \mathbf{k}_1(t)\, dx(t) \quad \text{for } x(t) \equiv 1 \quad \text{in } [0, 1]; \tag{3.9}$$

a superficial interpretation would lead to

$$K_1 x = \mathbf{k}_1(1)x(1) - \mathbf{k}_1(0)x(0) - \int_0^1 \dot{\mathbf{k}}_1(t)x(t)\, dt = 0. \tag{3.10}$$

However, if the integral in (3.9) is interpreted as an extension from $\mathbf{D}$ to L^2, an approximation of $x(t) \equiv 1$ through a sequence of functions in $\mathbf{D}$ leads to the limit

$$K_1 x = \mathbf{k}_1(0) - \mathbf{k}_1(1). \tag{3.11}$$

This may also be seen by interpreting $x(t) \equiv 1$ on [0, 1] as $\Theta(t) - \Theta(t-1)$

$$\left(\Theta(t) = \begin{cases} 1 & \text{for } t > 0 \\ 0 & \text{for } t < 0 \end{cases}\right):$$

$$\begin{aligned} K_1 x &= \int \mathbf{k}_1(t)\, dx(t) \\ &= \int \mathbf{k}_1(t)[\delta(t) - \delta(t-1)]\, dt \\ &= \mathbf{k}_1(0) - \mathbf{k}_1(1). \end{aligned} \tag{3.12}$$

In summary, we have shown that every functional $S : E([0, 1]) \to \mathbb{R}$, which can be written as a "Taylor series" like equation (3.2), has a "Volterra-like" integral representation

$$Sx = \sum_{n=0}^{\infty} \int_0^1 k_n(t_1, \cdots, t_n) x(t_1) \cdots x(t_n)\, dt_1 \cdots dt_n, \tag{3.13}$$

$$\text{with } k_n \in \mathbf{D}'([0, 1]^n),$$

which we call *symbolic integral Taylor expansion.* One can also write

$$Sx = \sum_{n=0}^{\infty} \int_0^1 \mathbf{k}_n(t_1, \cdots, t_n)\, dx(t_1) \cdots dx(t_n), \tag{3.14}$$

which we will refer to as *Stieltjes integral Taylor expansion.*

4. The Wiener representation of nonlinear systems. Wiener [33] introduced a canonical expansion of a nonlinear functional of the Brownian motion into a series of mutually orthonormal polynomial functionals. One of the main interests in Wiener's work lies in the possibility of physically measuring the various terms of the Wiener series for a nonlinear system. Thus, a complete characterization of nonlinear systems through input-output experiments should be possible, in analogy with harmonic input methods for linear systems. In this section we will discuss the basic concepts of Wiener's theory, its relationships with the Taylor series expansion outlined in the previous paragraphs, and convergence problems associated to the Wiener representation.

The Wiener development of a nonlinear functional is somewhat analogous to the Fourier theory as the following theorem[4] shows.

THEOREM 2. (Compare Cameron and Martin [7] and Wiener [33].) *Let* $x(t)$ ($t \in [0, 1]$) *be a Brownian motion on* [0, 1] (*with* $x(0) \equiv 0$). *It is known that* $x(t)$ *is* (*with probability* 1) *continuous as a function of* t; *therefore the Brownian motion is determined by a probability measure* w—*called Wiener measure*—*on* $\mathbf{C}_0([0, 1])$. *Then the space* $L^2(\mathbf{C}_0([0, 1]), w)$ *contains a complete orthonormal set of Wiener*

[4] The theorem can be extended to a very large class of stochastic inputs, containing Brownian motion and other continuous as well as discrete time processes. This and related results will be published elsewhere [25].

functionals H_{nk}, which are polynomials of degree n, i.e.

$$H_{nk} = \sum_{i=0}^{n} G_{nki}, \tag{4.1}$$

where G_{nki} is homogeneous of degree i ($i \leqq n$) and in fact

$$G_{nki}x = \int_0^1 \cdots \int \mathbf{g}_{nki}(\tau_1, \cdots, \tau_i)\, dx(\tau_1) \cdots dx(\tau_i) \tag{4.2}$$

with $\mathbf{g}_{nki} \in L^2([0,1]^i)$.

Thus a functional $S \in L^2(\mathbf{C}_0, w)$ can be expanded in the Wiener series

$$S = \sum_{n,k=0}^{\infty} (S, H_{nk}) H_{nk} = \sum_{n,k=0}^{\infty} \sum_{i=0}^{n} (S, H_{nk}) G_{nki}, \tag{4.3}$$

where $(S, H_{nk}) = \int Sx \cdot H_{nk}x\, dw(x)$ represents the inner product in $L^2(\mathbf{C}_0, w)$ and the order of summation on the indices n and k is immaterial. Formal rearrangements of the summation indices provide various polynomial functionals which have appeared in the literature. For instance, the functionals H_n

$$H_n = \sum_{i=0}^{n} G_{ni} = \sum_{i=0}^{n} \sum_{k=0}^{\infty} (S, H_{nk}) G_{nki} \tag{4.4}$$

correspond to the orthogonal functionals used, with a different notation, by Lee and Schetzen [14] and Marmarelis and Naka [20] among others. Exhaustive definitions and correspondences between the different notations are given in Appendix A. The formal connection of the Taylor functionals K_i, introduced in (3.2), with the elementary functionals G_{nki} is given by

$$K_i = \sum_{n=i}^{\infty} \sum_{k=0}^{\infty} (S, H_{nk}) G_{nki}. \tag{4.5}$$

At this point, the right-hand side of (4.5) is only a formal series. Later we will show under which conditions (4.5) makes sense.

The Stieltjes representation for the homogeneous functionals G_{nki}, (4.2), can be alternatively rewritten (see Gel'fand and Vilenkin [10, Vol. IV, Chap. III]) in terms of the generalized stochastic process $\dot{x}(t)$, i.e. "white noise" ($x(t)$ is Brownian motion, $x(t) \in \mathbf{C}_0([0, 1])$, $\dot{x}(t) \in \mathbf{D}'([0, 1])$), as

$$G_{nki}x = \int_0^1 \mathbf{g}_{nki}(\tau_1, \cdots, \tau_i)\dot{x}(\tau_1) \cdots \dot{x}(\tau_i)\, d\tau_1 \cdots d\tau_i, \tag{4.6}$$

or as

$$G_{nki}x = \int_0^1 g_{nki}(\tau_1, \cdots, \tau_i)x(\tau_1) \cdots x(\tau_i)\, d\tau_i \cdots d\tau_i, \tag{4.7}$$

where

$$g_{nki} = (-1)^n \frac{\partial^i}{\partial\tau_1 \cdots \partial\tau_i} \mathbf{g}_{nki}(\tau_1, \cdots, \tau_i), \tag{4.8}$$

showing that the functionals G_{nki} also have a symbolic integral representation in terms of the symbolic kernels g_{nki}, which are derivatives (in a distributionary sense) of L^2-functions as the k_i in § 3.

Wiener [33] has discussed practical methods to measure the coefficients (S, H_{nk}). However, we are faced with three difficulties concerning the Wiener representation.

(a) The first problem concerns the convergence of the sum $S = \sum_{n,k=0} (S, H_{nk})H_{nk}$. According to Theorem 2, the convergence is in the $L^2(\mathbf{C}_0, w)$ sense, that is

$$\int \left| \sum_{nk=0}^{p,q} (S, H_{nk})H_{nk}x - Sx \right|^2 dw(x) \to 0.$$

Therefore, the number $\sum_{n,k=0}^{\infty} (S, H_{nk})H_{nk}x$ is only determined for almost every realization of the Brownian notion. Moreover, the sum $\sum_{n,k=0}^{\infty} (S, H_{nk})H_{nk}x$ may not converge for any particular realization x. Thus, the actual meaning of the Wiener representation of a system for any single input function is not clear (see question 2 in the Introduction).

(b) The convergence of the formal power series introduced earlier (see (4.3), (4.4), (4.5)) also represents a problem which has to be investigated in detail. One aspect of the question concerns the point-wise convergence of those series; another aspect has to do with the possibility of rearranging the order of summation of the indices in the various series. For instance, while the order of summation on the indices n and k in (4.3) can be interchanged, the index i presents a much more difficult problem. We will discuss this point in the next section.

(c) The main interest in the Wiener representation is stimulated by its application to actual physical systems. However, Brownian motion or "white noise" inputs are not physically available. Physical devices can only give approximations, which are much "smoother". What are the implications for the Wiener representation? We will not touch the rather critical stochastic problems associated with the practical presentation of Brownian motion or "white noise" inputs. For a recent review of some of these problems compare McKean [23].

Section 6 provides a few theorems which deal with difficulties (a) and (b). Before digressing into the mathematics, we briefly present some heuristic remarks on these three points.

5. Remarks on the Wiener representation and its problems. Let us first consider problem (c). If one approximates a realization $x(t) \in \mathbf{C}_0([0, 1])$ of the Brownian motion (i.e. a continuous, not differentiable function) by some "physical" function $z(t)$, it is necessary to require of the system S that Sz is "near" to Sx. Therefore, the mapping $S: \mathbf{C}_0([0, 1]) \to \mathbb{R}$ has to be continuous. In addition, if approximated "white noise" inputs are considered, a necessary condition is that even the mapping $S: \mathbf{D}'([0, 1]) \to \mathbb{R}$ must be continuous (the "white noise" realizations are in $\mathbf{D}'([0, 1])$), which represents a much stronger requirement on the system S. More generally: the larger the input spaces allowed, the stronger are the conditions on the system. In fact, one of the basic disadvantages of the Wiener representation is that the Wiener measure is not the *natural* measure for many applications: Brownian motion and white noise inputs cannot be generated physically and their approximation presents deep problems.

We consider now the problem (a), raised in the previous section: what does the $L^2(\mathbf{C}_0, w)$ convergence of the Wiener representation $S=\sum_{nk}^{\infty}(S, H_{nk})H_{nk}$ mean for a single input function $x(t)\in \mathbf{C}_0([0, 1])$? A standard measure-theoretic interpretation runs as follows: it is known that the $L^2(\mathbf{C}_0, w)$ convergence implies the convergence in probability, that is for every ε, $\delta>0$ there is an integer m, such that for $p, q\geq m$

$$w\left(\left\{x: \left|\sum_{n,k=0}^{pq}(S, H_{nk})H_{nk}x - Sx\right|>\delta\right\}\right)<\varepsilon. \tag{5.1}$$

In other words, given any particular input x and two small numbers ε, $\delta>0$, for sufficiently large p, q we are sure that the error associated to the truncated (at p, q) Wiener representation is less than δ, with a "confidence level" of ε.

This heuristic interpretation shows that—at least in a statistical sense—the Wiener representation is valid for single specific inputs. Moreover, the statistical argument can be made precise under some additional assumptions. This will be proven in Theorem 4.

We finally turn to problem (b) of the previous section, namely the order of summation with respect to the different indices n, k, i. Clearly, in the orthogonal expansion (equation (4.3))

$$S=\sum_{n,k}(S, H_{nk})H_{nk}\quad\left(\sum_{n,k}(S, H_{nk})^2<\infty,\quad H_{nk}=\sum_{i=0}^{n}G_{nki}\right)$$

the summation with respect to n and k can be interchanged, the convergence being in the $L^2(\mathbf{C}_0, w)$ sense. However, it is not clear under which conditions on $S=\sum_{n,k,i}(S, H_{nk})G_{nki}$ the summation with respect to k and i can be interchanged. To settle the question whether

$$\sum_{k=0}^{\infty}\sum_{i=0}^{n}(S, H_{nk})G_{nki}=\sum_{i=0}^{n}\sum_{k=0}^{\infty}(S, H_{nk})G_{nki}, \tag{5.2}$$

we have to show, that all the sums $\sum_{k=0}^{\infty}(S, H_{nk})G_{nki}$ $(0\leqq i\leqq n)$ and $\sum_{k=0}^{\infty}(S, H_{nk})H_{nk}$ converge. The last sum converges in $L^2(\mathbf{C}_0, w)$. For $\sum_{k=0}^{\infty}(S, H_{nk})G_{nkn}$ we can consider the corresponding Stieltjes-kernels $(\mathbf{g}_{nkn})_{k=0,1,2,\cdots}$ which are orthogonal in $L^2([0, 1]^n)$. Therefore $\sum_{k=0}^{\infty}(S, H_{nk})\mathbf{g}_{nkn}$ converges in $L^2([0, 1]^n)$. As for the other sums $\sum_{k=0}^{\infty}(S, H_{nk})G_{nki}$ $(0\leqq i\leqq n-1)$: their convergence can in general not be assured. In the next section rather strong conditions will be given, which imply the pointwise convergence of all the sums occurring in (5.2). Furthermore, in Theorem 4 we will show that the interchange of the order of summation with respect to all three indices is immaterial.

6. Additional requirements for pointwise convergence. In this section we will characterize some conditions under which the Wiener representation also makes sense for any specific input. Moreover, the connection of the Wiener representation with the Taylor (or "Volterra-like") expansion of § 2 will be derived. The basic result is the following theorem.

THEOREM 3. *Let E be a topological (vector) space, containing $\mathbf{C}_0([0,1])$ as a dense subspace, such that the sup-norm topology of $\mathbf{C}_0$ is finer than the topology induced by E. Consider a sequence of mappings $S_n : E \to \mathbb{R}$, equicontinuous on some neighborhood $V \subseteq E$ of zero, and a mapping $S : E \to \mathbb{R}$, which is continuous on V.*

If every S_n belongs to $L^2(\mathbf{C}_0, w)$ and $S_n \to S$ in the $L^2(\mathbf{C}_0, w)$-norm, then $S_n x \to Sx$ for every $x \in V$.

Proof. Suppose for some $x_0 \in V$ $S_n x_0$ does not converge to Sx_0. Then there is a $\delta > 0$ and a subsequence (S_{n_k}) of (S_n) such that $|S_{n_k}x_0 - Sx_0| > 3\delta$. Since (S_{n_k}) are equicontinuous we can find an open neighborhood $U \subseteq V$ of x_0 such that $|S_{n_k}x - S_{n_k}x_0| < \delta$ for every k and $|Sx - Sx_0| < \delta$; hence

$$(*) \qquad |S_{n_k}x - Sx| > \delta \quad \text{for every } x \in U.$$

Now $U' := U \cap \mathbf{C}_0([0,1]) \neq \emptyset$ is open in $\mathbf{C}_0[0,1]$, and therefore $w(U') \neq 0$ (see Appendix B). But on the other hand $S_n \to S$ converges in the Wiener sense; hence for every ε, $\delta > 0$ there is $m \in \mathbb{N}$ such that for every $n \geqq m$

$$w(\{x : |S_n x - Sx\| > \delta\}) < \varepsilon \quad \text{(see (5.1))}.$$

If we call $\{x : |S_n x - Sx| > \delta\} =: M_n$, this means $w(M_n) \to 0$. In $(*)$ we have shown that $U' \subseteq M_{n_k}$, and therefore $w(U') \leqq w(M_{n_k}) \to 0$, contradicting $w(U') \neq 0$.

We translate Theorem 3 into the language of Theorem 2, using a Banach space E in Theorem 4 and the space $L^2([0,1])$in Theorem 5.

THEOREM 4. *Let E be a Banach space containing $\mathbf{C}_0([0,1])$ as a dense subspace, such that the sup-norm topology on $\mathbf{C}_0$ is finer than the topology induced by E. Let $S : E \to \mathbb{R}$ be a continuous functional and $S \in L^2(\mathbf{C}_0, w)$. Consider the following conditions on the functionals G_{nki} of the Wiener expansion:*

(a) $\|\sum_{n=i}^{n_1} \sum_{k=0}^{k_1} (S, H_{nk}) G_{nki}\| \leqq c^{-i} d$, *for every n_1, k_1, i and some constant $d > 0$,*

(b) $\sum_{n=i}^{\infty} \sum_{k=0}^{\infty} |(S, H_{nk})| \, \|G_{nki}\| \leqq c^{-i} d$, *for every i and some constant $d > 0$,*

(c) $G_{ni}x := \sum_{k=0}^{\infty} (S, H_{nk}) G_{nki}x$ *converges*[5] *for every x and* $\|\sum_{n=0}^{n_1} G_{ni}\| \leqq c^{-i} d$ *for every n_1, i and some $d > 0$,*

(d) $G_{ni}x := \sum_{k=0}^{\infty} (S, H_{nk}) G_{nki}x$ *converges*[5] *for every x and* $\sum_{n=0}^{\infty} \|G_{ni}\| \leqq c^{-i} d$ *for every i and some $d > 0$.*

Any of these conditions imply that

(i) $\sum_{nk=0}^{\infty} \sum_{i=0}^{n} (S, H_{nk}) G_{nki}x = Sx$ *holds for* $\|x\| < c$. *Moreover, if* $K_i x := \sum_{n=i}^{\infty} \sum_{k=0}^{\infty} (S, H_{nk}) G_{nki}x$ *converges,*

(ii) $\sum_{i=0}^{\infty} K_i x = Sx$ *holds and is the Taylor expansion of S for $\|x\| < c$ (therefore S is infinitely often Fréchet-differentiable).*

Proof. (a)$\Rightarrow$(i). From Theorem 2 we know, that (i) holds in $L^2(\mathbf{C}_0, w)$. Now we apply Theorem 3, to show the pointwise convergence (i) on

$$V := \{x \in E : \|x\| < c\}.$$

[5] Except for g_{nn}, the corresponding symbolic kernels g_{ni} are not necessarily the Lee–Schetzen kernels (see § 7 and discussion).

We only have to verify the equicontinuity of the functionals

$$F_{n_1k_1} := \sum_{n=0}^{n_1} \sum_{k=0}^{k_1} \sum_{i=0}^{n} (S, H_{nk}) G_{nki}:$$

Let $x, x_0 \in V$ and $s := \sup\{\|x\|, \|x_0\|\} < c$.

$$|F_{n_1k_1}x - F_{n_1k_1}x_0| \leqq \sum_{i=0}^{n_1} \left| \sum_{k=0}^{k_1} \sum_{n=i}^{n_1} (S, H_{nk})(G_{nki}x - G_{nki}x_0)\right|$$

$$\leqq \sum_{i=1}^{n_1} \left\| \sum_{k=0}^{k_1} \sum_{n=i}^{n_1} (S, H_{nk}) G_{nki} \right\| \cdot i \cdot s^{i-1} \cdot \|x - x_0\|$$

(Compare Hille and Phillips [11, p. 764])

$$\leqq \sum_{i=1}^{n_1} c^{-i} \cdot d \cdot i \cdot s^{i-1} \|x - x_0\| \quad \text{(by (a))}$$

$$< \sum_{i=1}^{\infty} i \left(\frac{s}{c}\right)^{i-1} \cdot \frac{d}{c} \|x - x_0\| = \frac{dc}{(c-s)^2} \|x - x_0\|.$$

(a)$\Rightarrow$(ii). Now we know, that (i) holds, i.e. $\sum_{n,k=0}^{\infty} (S, H_{nk}) H_{nk} x = Sx$ for $\|x\| < c$, and we assume that $\sum_{n,k} (S, H_{nk}) G_{nki} x = K_i x$ converges, for some $x \in V$. Let $\varepsilon > 0$, then we choose i_1 so large, that

$$(\alpha) \qquad \frac{c \cdot d}{c - \|x\|} \left(\frac{\|x\|}{c}\right)^{i_1} < \frac{\varepsilon}{2},$$

and prove $|\sum_{i=0}^{i_1} K_i x - Sx| < \varepsilon$:

Choose n_1, k_1 so large, that

$$\left| Sx - \sum_{n=0}^{n_1} \sum_{k=0}^{k_1} H_{nk} x \right| < \frac{\varepsilon}{2(i_1 + 2)}$$

and

$$\left| K_i x - \sum_{n=i}^{n_1} \sum_{k=0}^{k_1} (S, H_{nk}) G_{nki} x \right| < \frac{\varepsilon}{2(i_1+2)} \qquad (0 \leqq i \leqq i_1).$$

Then

$$\left| \sum_{i=0}^{i_1} K_i x - \sum_{i=0}^{i_1} \sum_{n=i}^{n_1} \sum_{k=0}^{k_1} (S, H_{nk}) G_{nki} x \right|$$

$$\leqq \sum_{i=0}^{i_1} \left| K_i x - \sum_{n=i}^{n_1} \sum_{k=0}^{k_1} (S, H_{nk}) G_{nki} x \right| < \frac{\varepsilon}{2} \frac{i_1 + 1}{i_1 + 2},$$

and

$$\left| \sum_{i=0}^{i_1} \sum_{n=i}^{n_1} \sum_{k=0}^{k_1} (S, H_{nk}) G_{nki} x - \sum_{i=0}^{n_1} \sum_{n=i}^{n_1} \sum_{k=0}^{k_1} (S, H_{nk}) G_{nki} x \right|$$

$$= \left| \sum_{i=i_1}^{n_1} \sum_{n=i}^{n_1} \sum_{k=0}^{k_1} (S, H_{nk}) G_{nki} x \right|$$

$$\leqq \sum_{i=i_1}^{\infty} \left| \sum_{n=i}^{n_1} \sum_{k=0}^{k_1} (S, H_{nk}) G_{nki} x \right| < \sum_{i=i_1}^{\infty} c^{-i} d \|x\|^i \quad \text{(by (a))}$$

$$= \frac{cd}{c - \|x\|} \left(\frac{\|x\|}{c} \right)^{i_1} < \frac{\varepsilon}{2} \quad \text{(by } (\alpha)).$$

Since

$$\sum_{n=0}^{n_1} \sum_{k=0}^{k_1} H_{nk} x = \sum_{n=0}^{n_1} \sum_{k=0}^{k_1} \sum_{i=0}^{n} (S, H_{nk}) G_{nki} x$$

$$= \sum_{i=0}^{n_1} \sum_{n=i}^{n_1} \sum_{k=0}^{k_1} (S, H_{nk}) G_{nki} x$$

we can combine the obtained inequalities to

$$\left| \sum_{i=0}^{i_1} K_i x - Sx \right| = \left| \sum_{i=0}^{i_1} K_i x - \sum_{i=0}^{i_1} \sum_{n=i}^{n_1} \sum_{k=0}^{k_1} (S, H_{nk}) G_{nki} x \right|$$

$$+ \left| \sum_{i=0}^{i_1} \sum_{n=i}^{n_1} \sum_{k=0}^{k_1} (S, H_{nk}) G_{nki} x - \sum_{n=0}^{n_1} \sum_{k=0}^{k_1} H_{nk} x \right|$$

$$+ \left| \sum_{n=0}^{n_1} \sum_{k=0}^{k_1} H_{nk} x - Sx \right|$$

$$< \frac{\varepsilon}{2} \cdot \frac{i_1 + 1}{i_1 + 2} + \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \cdot \frac{1}{i_1 + 2} = \varepsilon.$$

We still have to prove, that (ii) is the Taylor-expansion of S for $\|x\| < c$, and S is infinitely often Fréchet-differentiable. But these are easy consequences of [11, Thm. 3.18.1] if we can check, the "locally uniform boundedness" of the functionals $(\sum_{i=0}^{i_1} K_i)_{i_1 \in \mathbb{N}}$: For every $c' < c$ in the ball $B = \{x \in E : \|x\| \leqq c'\}$ we have

$$\left| \sum_{i=0}^{i_1} K_i x \right| \leqq \sum_{i=0}^{\infty} |K_i x| < \sum_{i=0}^{\infty} c^{-i} d \|x\|^i$$

(by (a) and the definition of K_i)

$$= \frac{dc}{c - \|x\|} < \frac{dc}{c - c'}.$$

(c)$\Rightarrow$(i). The proof runs along the lines of (a)$\Rightarrow$(i), if one considers the functionals $F_{n_1} := \sum_{n=0}^{n_1} \sum_{i=0}^{n} G_{ni}$.

(c)$\Rightarrow$(ii). Again the proof is an easier version of (a)$\Rightarrow$(ii); (b)$\Rightarrow$(a) and (d)$\Rightarrow$(c) are consequences of the triangle inequality.

In Theorem 4 the convergence properties are not changed by changes in any finite number of the functionals G_{nki}. We will make use of this fact in the following theorem, where stronger assumptions on the symbolic kernels g_{nki} actually imply that they are in $L^2([0, 1]^i)$, which is not always the case (see for instance, (3.4)). However, the conditions required in Theorem 5 meet the demands of many applications (and, of course, the case of polynomial systems).

THEOREM 5. *Let $S: L^2([0, 1]) \to \mathbb{R}$ be a continuous functional and $S \in L^2(\mathbf{C}_0, w)$. Consider the following conditions on the kernels g_{nki} occurring in the symbolic integral Wiener representation: There is a constant $d > 0$ and integers i_0, n_0, k_0 such that*

(a) $\|\sum_{n=i}^{n_1}\sum^{k_1} (S, H_{nk}) g_{nki}\|_2 \leqq c^{-i} d$ *for every* $i > i_0$, $n_1 > n_0$, $k_1 > k_0$,

(b) $\sum_{n=i}^{\infty}\sum_{k=0}^{\infty} (S, H_{nk}) \|g_{nki}\|_2 \leqq c^{-i} d$ *for every* $i > i_0$,

(c) $g_{ni} := \sum_{k=0}^{\infty} (S, H_{nk}) g_{nki}$ *converges in* $L^2([0, 1]^i)$ *and* $\|\sum_{n=i}^{n_1} g_{ni}\|_2 \leqq c^{-i} d$ *for every* $n_1 > n_0$, $i > i_0$,

(d) $g_{ni} := \sum_{k=0}^{\infty} (S, H_{nk}) g_{nki}$ *converges in* $L^2([0, 1]^i)$ *and* $\sum_{n=i}^{\infty} \|g_{ni}\|_2 \leqq c^{-i} d$ *for every* $i > i_0$,

(e) $k_i := \sum_{n,k} (S, H_{nk}) g_{nki}$ *converges in* $L^2([0, 1]^i)$ *and* $\|k_i\|_2 \leqq c^{-i} d$ *for every* $i > i_0$.

Each of these conditions imply that

(i) $\sum_{n,k=0}^{\infty}\sum_{i=0}^{n} (S, H_{nk}) G_{nki} x = Sx$ *holds for* $\|x\| < c$.

Moreover, if $\sum_{n=i}^{\infty}\sum_{k=0}^{\infty} (S, H_{nk}) G_{nki} x = K_i x$ *converges,*

(ii) $\sum_{i=0}^{\infty} K_i x = Sx$ *holds and is the Taylor expansion of S for* $\|x\| < c$ (here the kernels g_{ni}, k_i correspond to the functionals G_{ni}, K_i).

Proof. For any kernel $k \in L^2$ $([0, 1]^i)$ we denote the L^2-norm by $\|k\|_2 = (\int \cdots \int |k(t_1, \cdots, t_n)|^2 \, dt_1 \cdots dt_n)^{1/2}$ and get

$$\begin{aligned}
&\left|\int \cdots \int k(t_1, \cdots, t_i) x(t_1) \cdots x(t_i) \, dt_1 \cdots dt_i\right| \\
&\leqq \int \cdots \int \left|\int k(t_1, \cdots, t_i) x(t_1) \, dt_1\right| x(t_2) \cdots x(t_i) \, dt_2 \cdots dt_i \\
&\leqq \int \cdots \int \left[\left(\int |k(t_1 \cdots, t_i)|^2 \, dt_1\right)^{1/2} \|x\|\right] x(t_2) \cdots x(t_i) \, dt_2 \cdots dt_i \\
&= \|x\| \int \cdots \int \left[\left(\int |k(t_1 \cdots t_i)|^2 \, dt_1\right)^{1/2} x(t_2) \, dt_2\right] x(t_3) \cdots x(t_i) \, dt_3 \cdots dt_i \\
&\leqq \|x\| \int \cdots \int \left[\left(\int\int |k(t_1 \cdots t_i)|^2 \, dt_1 \, dt_2\right)^{1/2} \|x\|\right] x(t_3) \cdots x(t_i) \, dt_3 \cdots dt_i \\
&\leqq \|x\|^n \left(\int \cdots \int |k(t_1 \cdots t_i)|^2 \, dt_1 \cdots dt_i\right)^{1/2} = \|x\|^n \|k\|_2.
\end{aligned}$$

Therefore e.g.

$$\left\|\sum_{n=i}^{n_1}\sum_{k=0}^{k_1} (S, H_{nk}) G_{nki}\right\| \leqq \left\|\sum_{n=i}^{n_1}\sum_{k=0}^{k_1} (S, H_{nk}) g_{nki}\right\|_2,$$

showing that for $E = L^2([0, 1])$ the conditions (a), (b), (c), (d) of Theorem 5 imply the corresponding conditions of Theorem 4.

in this case the modified method provides the correct system identification:

$$\begin{aligned} \mathbf{g}_{00} &= E\{y(t)\} \\ &= E\left\{\int [1+(\sigma-\tau)\Theta(\sigma-\tau)+(\tau-\sigma)\Theta(\tau-\sigma)]\dot{x}'(\sigma)\dot{x}'(\tau)\,d\tau\,d\sigma\right\} \\ &= \int [1+(\sigma-\tau)\Theta(\sigma-\tau)+(\tau-\sigma)\Theta(\tau-\sigma)]E\{\dot{x}'(\sigma)\dot{x}'(\tau)\}\,d\tau\,d\sigma \qquad (7.4)\\ &= \int [1+(\sigma-\tau)\Theta(\sigma-\tau)+(\tau-\sigma)\Theta(\tau-\sigma)]\delta(\sigma-\tau)\,d\sigma\,d\tau = 1; \end{aligned}$$

$$\begin{aligned} \mathbf{g}_{11} &= E\{y(t)\dot{x}'(s)\} \\ &= E\left\{\int [1+(\sigma-\tau)\Theta(\sigma-\tau)+(\tau-\sigma)] \cdot \dot{x}'(\sigma)\dot{x}'(\tau)\,d\sigma\,d\tau\,\dot{x}'(s)\right\} \qquad (7.5)\\ &= 0; \end{aligned}$$

$$\mathbf{g}_{10} = 0; \qquad (7.6)$$

$$\begin{aligned} 2\mathbf{g}_{22}(r,s) &= E\{y(t)\dot{x}'(r)\dot{x}'(s)\} \\ &= E\left\{\int [1+(\sigma-\tau)\Theta(\sigma-\tau)+(\tau-\sigma)\Theta(\tau-\sigma)] \cdot \dot{x}'(\sigma)\dot{x}'(\tau)\,d\sigma\,d\tau\,\dot{x}'(r)\dot{x}'(s)\right\} \qquad (7.7a)\\ &= \int [1+(\sigma-\tau)\Theta(\sigma-\tau)+(\tau-\sigma)\Theta(\tau-\sigma)] \cdot E\{\dot{x}'(\sigma)\dot{x}'(\tau)\dot{x}'(r)\dot{x}'(s)\}\,d\sigma\,d\tau \\ &= \int [1+(\sigma-\tau)\Theta(\sigma-\tau)\Theta(\tau-\sigma)][\delta(\sigma-\tau)\delta(r-s)+\delta(\sigma-r)\delta(\tau-s) \\ &\qquad +\delta(\sigma-s)\delta(\tau-r)]\,d\sigma\,d\tau \\ &= \delta(r-s)+1+(r-s)\Theta(r-s)+(s-r)\Theta(s-r)+1+(s-r)\Theta(s-r) \\ &\qquad +(r-s)\Theta(r-s) \\ &= 2\{1+(r-s)\Theta(r-s)+(s-r)\Theta(s-r)\}+\delta(r-s), \qquad (7.7b) \end{aligned}$$

where the δ term is not to be taken into account (Lee and Schetzen, [14]). Practically this requires the other term (i.e. $\mathbf{g}_{22}$) to be sufficiently smooth in a neighborhood of the diagonal. Furthermore we obtain

$$\mathbf{g}_{21} \equiv 0, \qquad (7.8)$$

$$\mathbf{g}_{20} = -\int \mathbf{g}_{22}(r,r)\,dr = -1. \qquad (7.9)$$

The main difficulty of the Lee–Schetzen method, which cannot be overcome by a modified prescription, consists of the evaluation of the "diagonal integrals," like $\mathbf{g}_{20}$ in (7.9). The crosscorrelation method of Lee–Schetzen leads in fact to delta functions along the "diagonals," for all kernels of degree two or higher (compare (7.7b)). Therefore, the experimental errors in evaluating for instance $\mathbf{g}_{22}(\tau, \tau)$ may be large, leading to comparable large errors in the corresponding "diagonal integral" and finally in the second order Wiener functional. This practical difficulty has also a mathematical counterpart, namely question (b) of § 4 (see also the end of § 5 and the footnote 5). The "diagonal integrals" are, in general, ill-defined, since the kernels $\mathbf{g}_{nn}$ belong to $L^2([0, 1]^n)$ and their "diagonals" are sets of zero measure. As a matter of fact, Wiener noticed this problem (Wiener [33, p. 36]) which is circumvented by his method (Theorem 2). Let us illustrate this difficulty with a slight modification of our previous example. Assume that the system is

$$y(t) = Sx' = \iint [\Theta(\sigma-\tau)+\Theta(\tau-\sigma)]\, dx'(\sigma)\, dx'(\tau) + \int 2\Theta(\tau-\tau)\, d\tau. \tag{7.10}$$

It is not difficult to show that $S \in L^2(\mathbf{C}_0([0, 1]), w)$ is well defined, since $\Theta(\sigma-\tau) \in L^2([0, 1]^2)$ (compare Wiener, [33, p. 36]). However, the last integral does not make sense. As a consequence, even the modified Lee–Schetzen method, which requires the evaluation of this integral, fails here.

8. Conclusions. We have attempted to find the connections between some versions of the Volterra and the Wiener representation for nonlinear systems. A comparison of the various approaches with the original work of Wiener has revealed that different kernels were used without a discriminating terminology. *Therefore we have introduced the notions of "symbolic" and "Stieltjes" kernels* (see Appendix A). We believe that the language introduced in Appendix A is very helpful in clarifying the relations between different notations occurring in this field.

Throughout this paper two types of conditions are always required for the validity of a specific system representation: one set of conditions concerns the input space, the other set is on the mapping (or system) itself. These two types of requirements are strictly connected: heuristically, *the larger the input space, the smaller the mapping space which is allowed.* This remark is clearly of considerable importance especially for identification problems, where the choice of the test inputs not only determines the input space but also restricts the class of systems for which the identification is valid. For instance, a specific input space E in Theorem 4 determines a corresponding continuity requirement on the mapping. The linear symbolic kernel of the Taylor representation is[6]

(i) anything in $\mathbf{D}'$ if the mapping is continuous for inputs $x \in \mathbf{D}([0, 1])$,

(ii) any finite measure, except the Dirac $\delta(\tau)$, if the mapping is continuous for inputs $x \in \mathbf{C}_0([0, 1])$,

(iii) any function in $L^2([0, 1])$ if the mapping is continuous for inputs $x \in L^2([0, 1])$,

(iv) any function in $\mathbf{D}([0, 1])$ if the mapping is continuous for inputs $x \in \mathbf{D}'([0, 1])$.

[6] The Wiener representation is valid for linear systems, *iff* the Stieltjes kernel $\mathbf{g}_{11}$ is in $L^2([0, 1])$, i.e. the symbolic kernel g_{11} is the derivative of a function in L^2. The Lee–Schetzen method for linear systems requires the symbolic kernel g_{11} to be in $L^2([0, 1])$.

For the nonlinear kernels of second or higher order it is not easy to establish similar correspondences. In particular (iii), derived from the Riesz theorem, does not hold, in general, for nonlinear kernels (see § 3). However, (i) holds for all kernels, linear and nonlinear (compare Theorem 1).

We summarize now our main results in terms of the four problems outlined in the Introduction.

The first question about the class of systems admitting an integral "Volterra-like" representation has been completely answered in § 3 (see Theorem 1). An alternative answer is provided by Theorems 4, 5, which require different conditions on the system than Theorem 1. *The class of Volterra-like representable systems is only restricted by a kind of smoothness conditions.*

The second problem was extensively dealt with in §§ 4, 5, 6. *Theorem* 3 *is the fundamental result, which insures that the* $L^2(\mathbf{C}_0, w)$ *convergence of the Wiener representation implies also its convergence for every single input.* Theorems 4, 5 give simpler conditions for the pointwise validity of the Wiener representation.

Moreover, they provide the answer to the third question: *under the same conditions, the Taylor and the Wiener representations can be obtained from each other by exchanging the order of summations* (see also Appendix A).

However, it is clear that the conditions of Theorems 4 and 5 cannot be checked by experimental measurements to validate the system identification.

Section 7 was finally devoted to our fourth and last question concerning the identification method of Lee and Schetzen. Some confusion has occurred in the literature between Brownian motion and white noise and correspondingly between the mathematical justifications of the Wiener and of the Lee–Schetzen method. Our approach makes clear that this situation has arisen essentially because of confusion between Stieltjes and symbolic kernels. We were able to show, in § 7, that *the validity of the original Lee–Schetzen method is restricted to a smaller class of systems than the "Wiener" class, namely to the systems S whose "derivative" S′ belongs to the "Wiener" class.* This restriction can be removed by a modification, illustrated in Fig. 4, of the original method. However, *both* prescriptions present, in general, a rather deep mathematical difficulty concerning the "diagonal integral" appearing in the determination of the nonleading kernels g_{ni} $(0 \leqq i \leqq n-1)$. Practical consequences of this difficulty are the impossibility to measure exactly the nonlinear kernels along "diagonals" (due to the presence of delta pulses on these "diagonals") and the possible nonexistence of such "diagonal" integrals (compare Wiener [33, p. 36]).[7] Of course, there are systems, sufficiently "well-behaved", for which the Lee–Schetzen method provides the correct identification. Linear systems can be identified, as is well known, through the Lee–Schetzen method. The basic reason is that $g_{10} = 0$, i.e. no "diagonal integral" appears (Taylor and Wiener expansions coincide for linear systems). Quadratic systems already present the difficulty of the "diagonal integral" $g_{20} = \int g_{22}(\tau, \tau)\, d\tau$. However, this difficulty can be circumvented, if one knows that higher order Wiener functionals are identically zero. In this case, the following relation holds identically: $g_{00} - \int g_{22}(\tau, \tau)\, d\tau = Y_0$, where Y_0 is the system output

[7] An alternative method, proposed by French and Butz [9], which represents a translation in the frequency domain of the Lee–Schetzen method, also has similar "diagonal" difficulties.

for zero inputs (usually $Y_0 = 0$, since the system is not active). As a consequence it is not necessary to evaluate the diagonal kernel, if Y_0 is known. Interestingly, most of the practical applications of the Lee–Schetzen method have been restricted to quadratic systems (see for instance McCann and Marmarelis [22]). *However, for third or higher order systems, this problem of the Lee–Schetzen method cannot be easily avoided.*[8]

The Wiener method does not have the "diagonal integral" difficulty (and it is valid for a larger class of systems); but it is computationally cumbersome. Other identification methods, like multiple delta-pulses and multisinusoidal inputs, may often be more convenient, *if* the system has a finite order. The harmonic input method (Bedrosian and Rice [3]; Marchesini and Picci [17]) can be, for instance, interesting for the identification of the symbolic Taylor kernels since the class of allowed systems may be rather large (sinusoidal test functions are in $\mathbf{D}$).

A promising approach of a different nature, which relies on the functional and computational properties which can be associated to the various terms of a Volterra-like representation and does not strictly require a measurement of the kernels, has been recently developed and applied to some biological problems (see the review by Poggio and Reichardt [26]).

Appendix A. Polynomial functionals. Let S be a functional on a space E. An expression $Sx = \sum_{n=0}^{\infty} K_n x$ where K_n is a homogeneous polynomial functional of degree n (and $K_0 x$ is a constant), is called a "*Taylor expansion*" of S. An expression $Sx = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} H_{nk}$, where the functionals $H_{nk}x = \sum_{i=0}^{n} G_{nki}x$ (G_{nki} of degree i) are orthogonal in the Wiener sense (i.e. $\int H_{nk}xH_{n'k'}x\, dw(x) = \delta_{nn'} \cdot \delta_{kk'}$) (see Theorem 2), is called a "*Wiener expansion*" of S.

If the functionals K_i (respectively G_{nki}) of degree i can be expressed by symbolic integrals, like

$$K_i x = \int \cdots \int k_i(t_1 \cdots t_i) x(t_1) \cdots x(t_i)\, dt_1 \cdots dt_i \tag{3.5}$$

(k_i belonging to $\mathbf{D}'([0, 1]^i)$), this is called a *symbolic integral representation* of K_i (respectively G_{nki}). Similarly

$$Sx = \sum_{i=0}^{\infty} \int \cdots \int k_i(t_1 \cdots t_i) x(t_1) \cdots x(t_i)\, dt_1 \cdots dt_i$$

is called a *symbolic integral Taylor expansion*, and

$$Sx = \sum_{n,k} \sum_{i=0}^{n} \int \cdots \int g_{nki}(t_1 \cdots t_i) x(t_1) \cdots x(t_i)\, dt_1 \cdots dt_i$$

is called a *symbolic integral Wiener expansion.*

If the functionals K_i (resp. G_{nki}) of degree i can be expressed by a Stieltjes integral, like

$$K_i x = \int \cdots \int \mathbf{k}_i(t_1 \cdots t_i)\, dx(t_1) \cdots dx(t_i) \tag{3.6}$$

[8] To our knowledge, a few attempts to measure third order kernels resulted in a decrease of accuracy of the Wiener representation of the system ([18, p. 112]). The results of this paper suggest that this may be due to fundamental difficulties of the Lee–Schetzen method rather than to computational errors.

(at least for $x \in \mathbf{D}$), this is called a *Stieltjes-integral representation.* Similarly

$$Sx = \sum_{i=0}^{\infty} \int \cdots \int \mathbf{k}_i(t_1 \cdots t_i)\, dx(t_1) \cdots dx(t_i)$$

is called a *Stieltjes integral Taylor expansion*, and

$$Sx = \sum_{n,k} \sum_{i=0}^{n} \int \cdots \int \mathbf{g}_{nki}(t_1 \cdots t_i)\, dx(t_1) \cdots dx(t_i)$$

is called a *Stieltjes integral Wiener expansion.*

Wiener functionals. The Wiener expansion of S is given by

$$Sx = \sum_{nk=0}^{\infty} (S, H_{nk}) H_{nk} x = \sum_{nk=0}^{\infty} \sum_{i=0}^{n} (S, H_{nk}) G_{nki} x \tag{4.3}$$

(G_{nki} of degree i).

Therefore

(I) $\sum_{i=0}^{n} G_{nki} = H_{nk}$ are the orthogonal Wiener functionals,

(II) $\sum_{n=i}^{\infty} \sum_{k=0}^{\infty} (S, H_{nk}) G_{nki} = K_i$ are the Taylor functionals.

The method of Lee and Schetzen yields the homogeneous functionals

(III) $G_{ni} = \sum_{k=0}^{\infty} (S, H_{nk}) G_{nki}$ and therefore their expansion proceeds in terms of the orthogonal functionals

(IV) $H_n := \sum_{i=0}^{n} G_{ni} = \sum_{i=0}^{n} \sum_{k=0}^{\infty} (S, H_{nk}) G_{nki}$.

The corresponding kernels are denoted by g_{ni} resp. $\mathbf{g}_{ni}$. For instance, the orthogonal functionals G_n in the terminology of Lee and Schetzen correspond to our H_n; and the following formal correspondences exist between our g_{ni} and Lee and Schetzen's notation:

$$g_{nn} = h_n,$$

$$g_{10} = 0,$$

$$g_{21} = 0, \quad g_{20} = \int h_2(\tau_1, \tau_1)\, d\tau_1,$$

$$g_{32} = 0, \quad g_{30} = 0, \quad g_{31}(\tau_1) = \int h_3(\tau_1, \tau_2, \tau_2)\, d\tau_2.$$

Moreover, Wiener's expansion of the Lee–Schetzen h's in Laguerre functions corresponds to our "expansion" of the G_{ni} as

$$G_{ni} = \sum_{k=0}^{\infty} (S, H_{nk}) G_{nki}.$$

However, all these correspondences with the Lee–Schetzen functionals are strictly formal, since it is in general not clear that our series $\sum_{k=0}^{\infty} G_{nki} = G_{ni}$ coincides with the corresponding Lee–Schetzen functionals.

Appendix B.

THEOREM. *Every open set $U \subseteq \mathbf{C}_0([0, 1])$ contains a subset of positive Wiener measure.*

Proof. Let $x \in U$; then there is a $c > 0$, such that

$$B(x, c) := \{y \in \mathbf{C}_0 : |x(t) - y(t)| < c \,\forall t \in [0, 1]\} \subseteq U.$$

Let $\varepsilon > 0$, there is a $k \in N$ such that $|t - t'| \leqq 1/k$ implies $|x(t) - x(t')| < \varepsilon$ (x is uniformly continuous, since $[0, 1]$ is compact).

Let $I_j := [(j-1)/k, j/k]$ $(j=1,\cdots,k)$; then $\bigcup_{j=1}^{k} I_j = [0,1]$. It is known that $w(\{y : |y(t)| < 2\varepsilon\, \forall t \in I_1\}) > 0$. Now

$$|x(1/k)| = |x(1/k) - x(0)| < \varepsilon$$

and therefore

$$w(\{y : |y(t)| < 2\varepsilon\, \forall t \in I_1 \text{ and } |y(1/k) - x(1/k)| < \varepsilon/2\}) > 0$$

(since $|y(1/k) - x(1/k)| < \varepsilon/2$ implies $|y(1/k)| < 2\varepsilon$). If we define

$$M_j := \{y : |y(t) - y((j-1)/k)| < 2\varepsilon\, \forall t \in I_j \text{ and } |y(j/k) - x(j/k)| < \varepsilon/2\},$$

we have just shown $w(M_1) > 0$.

If $y \in M_j$, then $|y(j/k) - x(j/k)| < \varepsilon/2$ and therefore

$$\left|x\left(\frac{j+1}{k}\right) - y\left(\frac{j}{k}\right)\right| \leqq \left|x\left(\frac{j+1}{k}\right) - x\left(\frac{j}{k}\right)\right| + \left|x\left(\frac{j}{k}\right) - y\left(\frac{j}{k}\right)\right| < \frac{3}{2}\varepsilon;$$

hence $w(M_{j+1}|M_j) > 0$ (since $|y((j+1)/k) - x((j+1)/k)| < \varepsilon/2$ implies $|y((j+1)/k) - y(j/k)| < 2\varepsilon$).

Therefore $w(\bigcap_{j=1}^{k} M_j) = w(M_1) \cdot \prod_{j=1}^{k-1} w(M_{j+1}|M_j) > 0$. But for $y \in \bigcap_{j=1}^{k} M_j$ we have

$$|y(t) - x(t)| < \left|y(t) - y\left(\frac{j-1}{k}\right)\right| + \left|y\left(\frac{j-1}{k}\right) - x\left(\frac{j-1}{k}\right)\right| + \left|x\left(\frac{j-1}{k}\right) - x(t)\right|$$

$$< 2\varepsilon + \frac{\varepsilon}{2} + \varepsilon < 4\varepsilon$$

for $t \in I_j (j=1,\cdots,k)$. If we now take $\varepsilon = c/4$, we thus find

$$\bigcap_{j=1}^{k} M_j \subseteq B(x,c) \subseteq U \quad \text{and} \quad w(\bigcap_{j=1}^{k} M_j) > 0.$$

Acknowledgments. We would like to express our gratitude to R. Nagel, E. Pfaffelhuber, W. Reichardt for helpful comments and for critically reading this paper. B. Coleman, for whom most of our results were not new, contributed with several critical remarks. Thanks also are due to C. Namislo for correcting the English manuscript, to Inge Geiss and Rosanne Brown for preparing it, and to L. Heimburger for drawing the figures.

REFERENCES

[1] A. V. BALAKRISHNAN, *A general theory of nonlinear estimation problems in control systems*, J. Math. Anal. Appl., 8 (1963), pp. 4–30.

[2] J. F. BARRETT, *The use of functionals in the analysis of nonlinear physical systems*, J. Electron. Control, 15 (1963), pp. 567–615.

[3] E. BEDROSIAN AND S. O. RICE, *The output properties of Volterra systems (nonlinear systems with memory)*, Proc. IEEE, 59 (1971), no. 12, pp. 1688–1707.

[4] A. G. BOSE, *A theory of nonlinear systems*, Tech. Rep. 309, Res. Lab. Electr., Mass. Inst. of Tech., Cambridge, MA, 1959.

[5] M. B. BRILLIANT, *Theory of the analysis of nonlinear systems*, Tech. Rep. 345, Res. Lab. Electr., Mass. Inst. of Tech., Cambridge, MA, 1959.

[6] J. J. BUSSGANG, L. EHRMAN AND J. W. GRAHAM, *Analysis of nonlinear systems with multiple inputs*, Proc. IEEE, 62 (1974), no. 8, pp. 1088–1119.

[7] R. H. CAMERON AND W. T. MARTIN, *The orthogonal development of nonlinear functionals in series of Fourier–Hermite functionals*, Ann. of Math., 48 (1947), no. 2, pp. 385–392.

[8] H. M. FISHMAN, *Rapid complex impedance measurement in squid axon membrane via input–output crosscorrelation function*, Proc. 1st Symp. on Testing and Identification of Nonlinear Systems, G. D. McCann and P. Z. Marmarelis, eds., Calif. Inst. of Tech., Pasadena, 1975.

[9] A. S. FRENCH AND E. G. BUTZ, *Measuring the Wiener kernels of a non-linear system using the fast Fourier transform algorithm*, Internat. J. Control, 17 (1973), pp. 529–539.

[10] I. M. GEL'FAND AND N. Y. VILENKIN, *Generalized Functions*, Academic Press, New York and London, 1964.

[11] E. HILLE AND R. S. PHILLIPS, *Functional Analysis and Semi-groups*. American Mathematical Society, Providence, RI, 1957.

[12] J. KATZENELSON AND L. A. GOULD, *The design of nonlinear filters and control systems, Part II*, Information and Control, 7 (1964), pp. 117–145.

[13] H. KRAUSZ AND W. O. FRIESEN, *Identification of discrete input nonlinear systems using Poisson impulse trains*, Proc. 1st Symp. on Testing and Identification of Nonlinear Systems, G. D. McCann and P. Z. Marmarelis, eds., Calif. Inst. of Tech., Pasadena, 1975, pp. 125–146.

[14] Y. W. LEE AND M. SCHETZEN, *Measurement of the Wiener kernels of a Non-linear system by crosscorrelation*, Internat. J. Control, 2 (1965), pp. 237–254.

[15] E. D. LIPSON, *White noise analysis of phycomyces light growth response system I, II, III*, Biophysical J., 15 (1975), pp. 989–1045.

[16] G. MARCHESINI AND G. PICCI, *Impiego dello sviluppo in serie di funzionali di Volterra nello studio di sistemi di controllo nonlineari*, Università Padova, Istituto di Elettrotecnica e di Elettronica, Italy, 1968.

[17] ———, *On the evaluation of the response of nonlinear systems by functional series expansion of forced and free responses*, J. Franklin Inst., 288 (1969), pp. 469–481.

[18] P. Z. MARMARELIS, *Nonlinear dynamic transfer function for certain retinal neuronal systems*, Thesis, Calif. Inst. of Tech., Pasadena, 1971.

[19] P. Z. MARMARELIS AND K. I. NAKA, *White-noise analysis of a neuron chain: An application of the Wiener theory*, Science, 175 (1972), pp. 1276–1278.

[20] ———, *Identification of multi-input biological systems*, IEEE Transactions on Biomed. Engrg., BME 21 (1974), pp. 88–101.

[21] G. D. MCCANN, *Nonlinear Identification Theory models for successive stages of visual neurons systems of flies*, J. Neurophysiology, 37 (1974), pp. 869–875.

[22] G. D. MCCANN AND P. Z. MARMARELIS, eds., *Proceedings of the First Symposium on Testing and Identification of Nonlinear Systems*, Calif. Inst. of Tech., Pasadena, 1975.

[23] H. P. MCKEAN, *Wiener's theory of nonlinear noise*, Stochastic Differential Equations, J. B. Keller and H. P. McKean, eds., Courant Institute of Math. Sciences, New York, 1972.

[24] K. I. NAKA, P. Z. MARMARELIS AND R. Y. CHAN, *Morphological and functional identifications of catfish retinal neurons. III. Functional identification*, J. Neurophysiology, 38 (1975), pp. 92–131.

[25] G. PALM AND T. POGGIO, *Stochastic identification methods for nonlinear systems: An extension of the Wiener theory*, submitted for publication (1976).

[26] T. POGGIO AND W. REICHARDT, *Visual control of orientation behavior of the fly: Towards the underlying neural interactions*, Quart. Rev. in Biophysics, 9 (1976), pp. 376–438.

[27] T. POGGIO AND V. TORRE, *A nonlinear transfer function for some neuron models*, Proc. 1st Symp. on Testing and Identification of Nonlinear Systems, G. D. McCann and P. Z. Marmarelis, eds., Calif. Inst. of Tech., 1975, pp. 292–300.

[28] F. RATLIFF, B. W. KNIGHT, JR. F. DODGE AND H. K. HARTLINE, *Fourier analysis of dynamics of excitation and inhibition in the eye of Limulus: amplitude, phase and distance*, Vision Res., 14 (1974), pp. 1155–1168.

[29] L. SCHWARTZ, *Théorie des Distributions*, Hermann, Paris, 1973.

[30] W. V. SEELEN AND K. P. HOFFMANN, *Analysis of neuronal networks in the visual system of the cat using statistical signals*, Biol. Cybernet., 22 (1975), pp. 7–20.

[31] L. STARK, *The pupillary control system: Its nonlinear adaptive and stochastic engineering design characteristics*, Automatica —J. IFAC, 5 (1969), pp. 655–676.

[32] V. VOLTERRA, *Theory of Functionals and of Integral and Integro-Differential Equations*, Dover, New York, 1930.

[33] N. WIENER, *Nonlinear Problems in Random Theory*, John Wiley, New York, 1958.

[34] G. ZAMES, *Nonlinear operations for system analysis*, Tech. Rep. 370, Res. Lab. Electr., Mass. Inst. of Tech., Cambridge, MA, 1960.