University of Groningen

Calculation of Volterra kernels for solutions of nonlinear differential equations

Thomas, Erik G F; van Hemmen, JL; Kistler, WM

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2000

[Link to publication in University of Groningen/UMCG research database](#)

CALCULATION OF VOLTERRA KERNELS FOR SOLUTIONS OF NONLINEAR DIFFERENTIAL EQUATIONS*

J. LEO VAN HEMMEN†, WERNER M. KISTLER†, AND ERIK G. F. THOMAS‡

Abstract. We consider vector-valued autonomous differential equations of the form $x' = f(x) + \varphi$ with analytic f and investigate the nonanticipative solution operator $\varphi \mapsto x = A(\varphi)$ in terms of its Volterra series. We show that Volterra kernels of order > 1 occurring in the series expansion of the solution operator A are continuous functions, and establish recurrence relations between the kernels allowing their explicit calculation. A practical tensor calculus is provided for the finite-dimensional case. In addition to analytically calculating the kernels, we present an algorithm to numerically obtain them from the output $x(t)$ through sampling the input space by linear combinations of delta functions. We call this "differential sampling." It is a nonlinear analogue of the classical method of impulse response. We prove a continuity theorem stating that, in the finite-dimensional case, approximate delta functions give rise to approximate Volterra kernels and that continuity holds in the sense of weak convergence. Finally, we discuss a practical implementation of differential sampling and relate it to the Wiener method.

Key words. nonlinear system, causal, nonanticipative, Volterra kernel

AMS subject classifications. Primary, 34A34, 34G20; Secondary, 47N20, 28C20

PII. S0036139999336037

1. Introduction. In the 1940s and early 1950s Hodgkin and Huxley [8, 16] analyzed the propagation of an action potential in the giant axon of the squid in the presence of what they thought to be three ion channels, viz., sodium, potassium, and chloride. Through an intricate data fit they arrived at a set of four coupled nonlinear differential equations that earned them the Nobel prize for its surprising accuracy; cf. the appendix. The equations are widely used in computational neuroscience. They are a concrete example of a more general problem,

$$(1.1) \qquad\qquad x' = f(x) + \varphi,$$

where x belongs to a Banach space E (here $E = \mathbb{R}^4$), f is an analytic function satisfying a Lipschitz condition, φ represents a time-dependent input, and the prime denotes differentiation with respect to time. One meanwhile knows that not three but many more types of ion channels (at least 25) are involved and, e.g., that in the cortex the Hodgkin–Huxley equations are not a good description of neuronal excitation. In other words, determining f is at the moment out of reach, but *sampling* the responses of the system *experimentally* is not. Since often no one knows what the appropriate differential equations are and they are likely to be very complicated, an alternative way of modeling the problem is to analyze the underlying solution operator. This is exactly what we do in the present paper.

How then should one sample a system's response? At the membrane resting potential, the system is in equilibrium and, without restriction of generality, we can

assume this is at $x = 0$. Through the input φ we can sample the neighborhood of $x = 0$. Since there exists a Volterra expansion (see below) about $x = 0$, the key question is indeed, How do we sample? Wiener's answer [40] was "through white noise." In this paper we will prove that, except for a trivial exception, the Volterra kernels are continuous, in fact, smooth, functions. We will indicate an algorithm to compute them, if f is known, and show how a more direct procedure than white-noise sampling leads to a determination of the Volterra kernels and, hence, to a characterization of the system's response to an external stimulus, if f is not known explicitly. The instrumental idea is to sample input space by Dirac delta measures directly. This is justified since the kernels are demonstrated to be continuous, even smooth. Finally, we will prove a continuity theorem stating that *approximate* delta functions give approximate kernels. This is useful because, in experiments, delta functions cannot be realized, whereas approximate delta functions are nearly always easy to generate.

In a more general context, we aim at analyzing nonlinear differential equations of the form (1.1) where x belongs to a Banach space E, usually $\mathbb{R}^N$ for some finite integer N. We will study the unique nonanticipative solution operator $\varphi \mapsto x = A(\varphi)$. A solution operator is said to be nonanticipative, or causal, if for each t the value $A(\varphi)(t)$ depends on the restriction of φ to $(-\infty, t]$ only, i.e., on the past of the input.

In the case of a linear system with a linear operator $L : E \longrightarrow E$ appearing in the equation of motion

$$(1.2) \qquad\qquad x' = Lx + \varphi,$$

the nonanticipative solution operator is the convolution operator A defined by

$$(1.3) \qquad\qquad A(\varphi) = \kappa * \varphi,$$

where $\kappa(t) = Y(t)e^{tL}$ is the unique solution $\mathcal{T}$ of the equation

$$(1.4) \qquad\qquad \left(\frac{\mathrm{d}}{\mathrm{d}t} - L\right)\mathcal{T} = \delta\, I,$$

having its support in the interval $[0, +\infty)$; cf. Schwartz [36]. Here δ is the Dirac delta measure ("function"), the unit mass at zero; I is the identity operator on E; and Y is the Heaviside one-step function: 0 for $t < 0$ and 1 for $t \geq 0$. We note that in (1.3) φ may be any distribution whose support is limited to the left. In particular,

$$(1.5) \qquad\qquad \kappa = A(\delta),$$

which is the reason why κ is called the *impulse response*.

In the nonlinear case the Volterra kernels κ_n replace the kernel $\kappa = \kappa_1$ of the linear case. Since Volterra [39] researchers have been interested in a *series expansion* of the nonanticipative solution operator that represents the time evolution induced by (1.1). The Volterra series expansion for the scalar case ($N = 1$) amounts to

$$(1.6)$$
$$A(\varphi)(t) = \kappa_0 + \sum_{n \geq 1} \int_{-\infty}^{t} \mathrm{d}s_1 \,\ldots\, \int_{-\infty}^{t} \mathrm{d}s_n\, \kappa_n(t - s_1, \ldots, t - s_n)\, \varphi(s_1)\,\ldots\,\varphi(s_n).$$

The nth-order term in this expansion will be called $A_n(\varphi)$, A_n denoting the mapping $\varphi \mapsto A_n(\varphi)$. In that which follows we assume $f(0) = 0$ so that $\kappa_0 = 0$.

A few words on convergence of the Volterra series (1.6) are in order. The solution operators are not analytic, but "quotient-analytic" (Q-analytic), as set forth in [37]. The consequence of this for the present paper, in which we shall not go into the details of Q-analytic maps, is that a Volterra series converges uniformly for t in a compact interval I, for inputs φ in a neighborhood of zero, *depending on I*.

We can now precisely formulate the main purposes of the paper. They are three-fold. First is to obtain a recurrence relation giving the nth-order term A_n in the Taylor expansion (1.6) of the nonanticipative solution operator A as a function of the lower-order terms $A_1, \ldots, A_{n-1}$. This is of particular relevance to its (extremely) efficient handling by computer-algebraic software. The operators A_n are n-linear and nonanticipative. The corresponding responses will be called nth-order responses.

Second is to generalize the classical impulse-response idea to the case of nonlinear equations by a method which we call *differential sampling*. The method consists of determining the kernels κ_n as nth-order responses to appropriate combinations of delta measures ("functions"). We justify this by proving that even in the nonlinear case the inputs φ may be measures. Furthermore, it is shown that the kernels κ_n with $n \geq 2$ are continuous, in fact, smooth, functions. In practice, we obtain the nth-order responses directly from the solution operator A by a finite-difference operation. The practical implementation of this is explained in detail in section 7.

Third is to prove a continuity theorem stating that, when the input is a measure, the operators A_n are continuous with respect to the weak (vague) topology on measures with support in a fixed compact set. Hence Volterra kernels depend continuously on the initial conditions in the weak topology so that approximate delta functions give rise to good approximations to the kernel κ_n. In this way we justify experimental praxis where delta measures are hard to realize but generating good approximations is straightforward. For instance, in hearing research an approximate delta function is an utter triviality: a click, which is ideal in experiments and widely used—though not yet for probing input space and, hence, the output systematically.

In itself, the idea of determining the Volterra kernels through "impulse response" is not new but, to the best of our knowledge, the suggestions [2, 32, 34] so far have not been proven to be effective. We have shown elsewhere [20] that the numerics of the differential-sampling algorithm, which is explained here for the first time, is fast and easy to handle when being applied to, e.g., the Hodgkin–Huxley system [8, 16]; cf. the appendix. See Figure 1 for a plot of the corresponding second-order kernel κ_2.

In biological and physical practice [1, 3, 5, 7, 9, 10, 11, 12, 13, 17, 23, 24, 25, 26, 33], the response $x(t)$ is given experimentally. As we have already indicated, there is in general no hope of reconstructing (1.1), i.e., f, from its output $x(t)$. It is therefore preferable to compute the kernels κ_n since they determine the response to an external input φ; in reality, e.g., in the neighborhood of a fixed point, a few small n (say ≤ 2) often suffice [17, 20]. Rather than determining the *Wiener* kernels through sampling the input space by white noise [4, 21, 22, 28, 29, 30, 31, 40], which is somewhat hard to generate and handle, and through averaging, which means that one has to repeat the white-noise sampling many times so as to obtain an average, we will show that we may take the input φ to be a simple linear combination of Dirac delta functions (i.e., point masses). In so doing, we recuperate the *Volterra* kernels κ_n in (1.6) directly from the corresponding output $x(t)$ by differentiation.

Not only from an experimental but also from a theoretical point of view is studying the Volterra kernels worthwhile. An analytic expansion such as the Volterra series in (1.6) cannot be extended beyond a singularity, e.g., an action potential (spike)
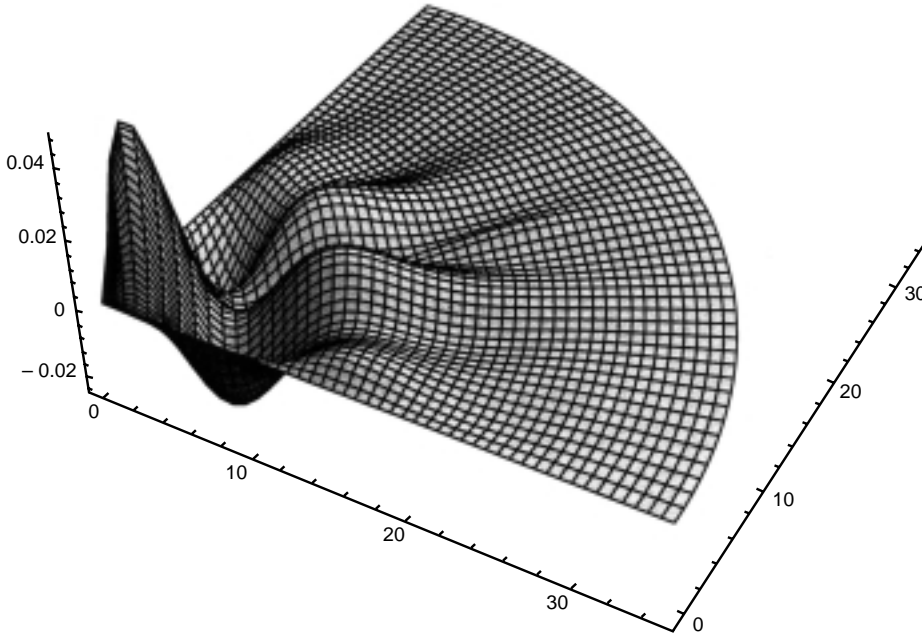
FIG. 1. *The second-order Volterra kernel κ_2 of the Hodgkin–Huxley system [8, 16]—see also the appendix—on its natural domain $\mathbb{R}_+^2$. This system was devised to describe action potentials in the giant axon of the squid. It consists of four nonlinear ordinary differential equations, of which only the first is accessible to experimentation and, thus, to physical input. The Volterra kernels directly give the system's response to the input; cf. (1.6).*

originating from the Hodgkin–Huxley equations when the input current I exceeds a critical value I_c; see the appendix. The important question is now: How can one modify the kernels so that they do take into account the last or, even better, all previous action potentials? It has been shown elsewhere [20] that for the Hodgkin–Huxley system the last spike in conjunction with a modified first-order kernel and a threshold suffices to give a faithful, though not perfect, description of the times at which the action potentials originating from the system occur. The corresponding kernels can be determined numerically, and efficiently, through differential sampling. Since the ensuing arguments are beyond the scope of the present paper, we refer to Kistler, Gerstner, and van Hemmen [20] for details and return to our main theme, determining the Volterra kernels.

We illustrate differential sampling by considering a simple example. In one dimension, $N = 1$, we start by taking $\varphi = \lambda\delta_{t'}$ in (1.1), where $\delta_{t'}$ denotes a Dirac delta measure, the unit point mass at t'. Equation (1.6) then takes the form

$$(1.7) \qquad x_\lambda(t) = \sum_{n \geq 1} \lambda^n \kappa_n(t - t', \ldots, t - t'),$$

which, for t in a given compact interval, converges provided λ is sufficiently small [37]. Differentiating the series once with respect to λ at $\lambda = 0$ we obtain $\kappa_1(t - t')$. Similarly, substituting $\varphi = \lambda_1\delta_{t_1} + \lambda_2\delta_{t_2}$ in the series (1.6) and differentiating with respect to λ_1 and λ_2 at 0, we obtain $\kappa_2(t - t_1, t - t_2)$.

The paper is organized as follows. In the next section we establish recurrence

relations for the nth-order terms of the Volterra expansion, and derive the kernels which belong to them in section 3. The finite-dimensional case allows for a practical formulation in terms of tensors (section 4) and is illustrated by way of some examples (section 5). In section 6 we use the method of differential sampling to obtain the Volterra kernels and prove the continuity theorem. Finally, we present a numerical implementation of our procedure (section 7) and discuss our results (section 8).

2. Recurrence relations. Let E be a real Banach space and $f : E \to E$, $f(0) = 0$, be a function having a holomorphic extension to the strip $\{z \in E + iE : |\mathrm{Im}\, z| < \epsilon\}$, and satisfying the Lipschitz condition $|f(u) - f(v)| \leq L_\epsilon\, |u - v|$ for all u and v in the strip. We consider the differential equation

$$(2.1) \qquad x'(t) = f(x(t)) + \varphi(t)\,, \quad x \in C_+^1(\mathbb{R}, E)\,, \quad \varphi \in C_+(\mathbb{R}, E).$$

Here $C_+(\mathbb{R}, E)$ denotes a space of continuous functions from $\mathbb{R}$ to E that vanish on some interval $(-\infty, a]$ with $a \in \mathbb{R}$. The space $C_+^1(\mathbb{R}, E)$ of continuously differentiable functions is defined similarly.

An operator $A : C_+(\mathbb{R}, E) \longrightarrow C_+^1(\mathbb{R}, E)$ such that $A(0) = 0$ and such that $A(\varphi)$ is a solution of (1.1) for all $\varphi \in C_+(\mathbb{R}, E)$ is called a solution operator. The operator A is nonanticipative if for all $\tau \in \mathbb{R}$

$$\varphi_1(t) = \varphi_2(t) \quad \text{for all } t \leq \tau \implies A(\varphi_1)(\tau) = A(\varphi_2)(\tau),$$

or, equivalently,

$$\varphi_1(t) = \varphi_2(t) \quad \text{for all } t \leq \tau \implies A(\varphi_1)(t) = A(\varphi_2)(t) \quad \text{for all } t \leq \tau.$$

There exists a unique nonanticipative solution operator $A : C_+(\mathbb{R}, E) \to C_+^1(\mathbb{R}, E)$, and this operator is Q-analytic [37, Thm. 6.1]

The solution operator can be expanded into a series, viz.,

$$(2.2) \qquad A(\varphi) = \sum_{n=1}^{\infty} A_n(\varphi)\,, \quad A_n(\varphi) = \frac{1}{n!}\,(\mathrm{D}^n A)(0)(\varphi, \ldots, \varphi).$$

Here $(\mathrm{D}^n A)(0)(\varphi_1, \varphi_2, \ldots, \varphi_n)$ denotes the nth-order directional derivative at 0 in the directions $\varphi_1, \ldots, \varphi_n$, the dependence upon t being understood; see also (1.6). More precisely, the fact that A is Q-analytic implies that for every compact interval $I \subset \mathbb{R}$ there exists a neighborhood of zero $O = O(I) \subset C_+(\mathbb{R}, E)$ such that the series

$$(2.3) \qquad A(\varphi)(t) = \sum_{n=1}^{\infty} A_n(\varphi)(t)$$

is convergent for every $\varphi \in O$ and every $t \in I$; cf. [37, Remark 1.9]. Thus the series in (2.2), with a given φ, does not necessarily converge for all $t \in \mathbb{R}$.

The A_n are homogeneous polynomials of degree n, defined on $C_+(\mathbb{R}, E)$, with values in $C_+^1(\mathbb{R}, E)$. Since we have, for every $t \in \mathbb{R}$,

$$(2.4) \qquad A_n(\varphi)(t) = \frac{1}{n!} \left(\frac{d}{d\lambda} \right)^n A(\lambda \varphi)(t)_{|\lambda=0},$$

the operator $\varphi \mapsto A_n(\varphi)$ itself is nonanticipative.

2.1. Scalar case. We will first investigate the scalar case with $E = \mathbb{R}$. The generalization to Banach spaces will be discussed in the next subsection.

In order to derive a recurrence relation for the operators A_n we expand f in a Taylor series around 0,

$$(2.5) \qquad f(x) = \sum_{k=1}^{\infty} c_k x^k , \quad c_k = \frac{1}{k!} (D^k f)(0) \in \mathbb{R},$$

substitute the series expansions of f and A into (2.1), and obtain formally

$$(2.6) \qquad \sum_{n=1}^{\infty} A_n(\varphi)' = c_1 \left(\sum_{n=1}^{\infty} A_n(\varphi) \right) + c_2 \left(\sum_{n=1}^{\infty} A_n(\varphi) \right)^2 + \cdots + \varphi.$$

Next, we replace φ by $\lambda \varphi$ with $\lambda \in \mathbb{R}$,

$$(2.7) \qquad \sum_{n=1}^{\infty} \lambda^n A_n(\varphi)' = c_1 \left(\sum_{k=1}^{\infty} \lambda^k A_k(\varphi) \right) + c_2 \left(\sum_{k=1}^{\infty} \lambda^k A_k(\varphi) \right)^2 + \cdots + \varphi,$$

and identify the nth-order terms in λ.

The variable t is not mentioned explicitly in (2.6) and (2.7), so it must be borne in mind that the convergence is valid uniformly for t in a compact interval I, provided φ, or λ, belongs to a sufficiently small neighborhood of the origin depending on I. This is a consequence of the Q-analyticity of the solution operator [37]. Hence for any t we can, as usual, identify terms of the same degree in λ.

Obviously, in first order we obtain the linearized differential equation,

$$(2.8) \qquad A_1(\varphi)' = c_1 A_1(\varphi) + \varphi.$$

Let Y be the Heaviside one-step function, i.e., $Y(t) = 0$ for $t < 0$ and $Y(t) = 1$ for $t \geq 0$, and let $Y_{c_1}(t) = Y(t) e^{c_1 t}$. Then Y_{c_1} is the fundamental solution of (2.8) with support in $[0, \infty)$, i.e.,

$$(2.9) \qquad Y_{c_1}' - c_1 Y_{c_1} = \delta,$$

δ being the Dirac delta function. The nonanticipative solution operator for (2.8) is then given by

$$(2.10) \qquad A_1(\varphi) = Y_{c_1} * \varphi,$$

where "$*$" denotes a convolution, i.e., $(f * g)(t) = \int ds \, f(t - s) \, g(s)$.

To identify higher-order terms in the series expansion of the solution operator A, we note that a restriction of the sums in the right-hand side of (2.7) to $1 \leq k \leq n$ affects only terms of order at most n in the left-hand side. Thus using the multinomial equation,

$$(2.11) \qquad (x_1 + x_2 + \cdots + x_n)^l = \sum_{j_1 + \cdots + j_n = l} \binom{l}{j_1 \ldots j_n} x_1^{j_1} x_2^{j_2} \ldots x_n^{j_l},$$

we finally get

$$(2.12) \quad A_n(\varphi)' = \sum_{l=1}^{n} c_l \sum_{\substack{j_1 + \cdots + j_n = l \\ j_1 + 2j_2 + \cdots + n j_n = n}} \binom{l}{j_1 \ldots j_n} A_1(\varphi)^{j_1} A_2(\varphi)^{j_2} \ldots A_n(\varphi)^{j_n}, \quad n \geq 2.$$

If $l = 1$, the sum over $j_1, \ldots, j_n$ contains only a single term because $j_1 + j_2 + \cdots + j_n = 1$ implies that only a single index j_k equals one and all the others vanish. From

$j_1 + 2\,j_2 \cdots + n\,j_n = n$ we conclude that $j_n = 1$ and $j_1, \ldots, j_{n-1} = 0$. It is easy to see that $l = 1$ is also the only case with $j_n \neq 0$. So we put the term with $l = 1$ apart from the rest of the sum and obtain

(2.13)

$$A_n(\varphi)' = c_1\, A_n(\varphi) + \sum_{l=2}^{n} c_l \sum_{\substack{j_1 + \cdots + j_{n-1} = l \\ j_1 + 2j_2 + \cdots + (n-1)\,j_{n-1} = n}} \binom{l}{j_1 \ldots j_n} A_1(\varphi)^{j_1}\, A_2(\varphi)^{j_2}\, \ldots\, A_{n-1}(\varphi)^{j_{n-1}}.$$

This is a linear differential equation for $A_n(\varphi)$ and, the operator A_n being nonanticipative, its solution is given by a recurrence relation for the operator A_n, $n \geq 2$, in terms of the operators $A_1, \ldots, A_{n-1}$, i.e.,

(2.14)

$$A_n(\varphi) = Y_{c_1} * \sum_{l=2}^{n} c_l \sum_{\substack{j_1 + \cdots + j_{n-1} = l \\ j_1 + 2j_2 + \cdots + (n-1)\,j_{n-1} = n}} \binom{l}{j_1 \ldots j_n} A_1(\varphi)^{j_1}\, A_2(\varphi)^{j_2}\, \ldots\, A_{n-1}(\varphi)^{j_{n-1}}.$$

This will be justified more adequately in the next section.

2.2. Vector case. If we generalize the above calculation to arbitrary Banach spaces, the coefficients in the series expansion (2.5) of f are no longer real numbers but elements of the space $\mathcal{L}_n(E; E)$ of bounded multilinear operators from E^n to E. Instead of (2.5) we now find

(2.15) $$f(x) = \sum_{n=1}^{\infty} C_n(\underbrace{x, \ldots, x}_{n \text{ times}}), \quad C_n = \frac{1}{n!}\,(\mathrm{D}^n f)(0) \in \mathcal{L}_n(E; E).$$

Since the operators C_n are symmetric, the application of the multinomial formula remains valid and the calculation of the recurrence relation goes through almost unchanged as compared to the scalar case. We only have to be more careful with our notation.

A combination of (2.15) and (2.2) yields

(2.16) $$\sum_{n=1}^{\infty} A_n(\varphi)' = C_1\left(\sum_{n=1}^{\infty} A_n(\varphi)\right) + C_2\left(\sum_{n=1}^{\infty} A_n(\varphi), \sum_{n=1}^{\infty} A_n(\varphi)\right) + \cdots + \varphi.$$

As before, the first-order term corresponds to the linearized differential equation,

(2.17) $$A_1(\varphi)' = C_1\, A_1(\varphi) + \varphi,$$

with the solution

(2.18) $$A_1(\varphi) = Y_{C_1} * \varphi,$$

where Y_{C_1} is the fundamental solution of the corresponding homogeneous differential equation, with support in $[0, \infty)$,

(2.19) $$Y_{C_1}(t) = Y(t)\, \exp(t\,C_1),$$

C_1 now being a bounded linear operator in E. The terms of order $n \geq 2$ are identified as above, and we obtain the differential equation

(2.20)
$$A_n(\varphi)' = C_1 \, A_n(\varphi)$$
$$+ \sum_{l=2}^{n} \sum_{\substack{j_1+\cdots+j_{n-1}=l \\ j_1+2j_2+\cdots+(n-1)\,j_{n-1}=n}} \binom{l}{j_1 \ldots j_n} C_l\big(A_1(\varphi)^{j_1}, A_2(\varphi)^{j_2}, \ldots, A_{n-1}(\varphi)^{j_{n-1}}\big),$$

where we have introduced the abbreviation

(2.21)
$$C_l\big(A_1(\varphi)^{j_1}, \ldots, A_{n-1}(\varphi)^{j_{n-1}}\big) = C_l\big(\underbrace{A_1(\varphi), \ldots, A_1(\varphi)}_{j_1 \text{ times}}, \ldots, \underbrace{A_{n-1}(\varphi), \ldots, A_{n-1}(\varphi)}_{j_{n-1} \text{ times}}\big).$$

The recurrence relation in the vector case is

(2.22)
$$A_n(\varphi) = Y_{C_1} * \sum_{l=2}^{n} \sum_{\substack{j_1+\cdots+j_{n-1}=l \\ j_1+2j_2+\cdots+(n-1)\,j_{n-1}=n}} \binom{l}{j_1 \ldots j_n} C_l\big(A_1(\varphi)^{j_1}, A_2(\varphi)^{j_2}, \ldots, A_{n-1}(\varphi)^{j_{n-1}}\big).$$

This is a consequence of the following lemma.

LEMMA 1. *Let $F : \mathbb{R} \longrightarrow E$ be a locally integrable function, or a distribution, with support bounded from below. Let C be a continuous linear operator in E. Then the differential equation*

(2.23)
$$x' = Cx + F$$

has a unique solution with support bounded from below, viz.,

(2.24)
$$x = Y e^{tC} * F.$$

Proof. The solution (2.24) has its support contained in the support of F, limited to the left. The most general solution has the form $Y e^{tC} * F + e^{tC} q$ for some $q \in E$, and this is different from zero for all t to the left of the support of F, unless $q = 0$.

Since the operator A_n is nonanticipative, and φ has support limited to the left, the function $A_n(\varphi)$ has support limited to the left as well. Thus, according to the lemma, (2.20) implies (2.22).

3. Derivation of the kernels. We are going to show, by induction on n, that the operators A_n are integral operators with kernel κ_n, viz.,

(3.1) $$A_n(\varphi)(t) = \int ds_1 \, \ldots \int ds_n \, \kappa_n(t - s_1, \ldots, t - s_n) \, \varphi(s_1) \, \ldots \, \varphi(s_n).$$

In the scalar case κ_n is a function of n variables. If E is a Banach space, then $\kappa(s_1, \ldots, s_n)$ is no longer a real number but a bounded multilinear operator from $\mathcal{L}_n(E; E)$.[1] So we use the notation

(3.2) $$A_n(\varphi)(t) = \int ds_1 \, \ldots \int ds_n \, \kappa_n(t - s_1, \ldots, t - s_n)\big(\varphi(s_1), \ldots, \varphi(s_n)\big).$$

---

[1] κ_n is thus a family of operators parameterized by n real parameters.

We have seen that the solution to the linearized differential equation (2.17) corresponds to a convolution with the fundamental solution Y_{C_1}. The first-order kernel is therefore readily determined to be

$$(3.3) \qquad \kappa_1 = Y_{C_1}.$$

Higher-order kernels follow from the recurrence relation (2.22). We go on to derive explicit expressions for κ_2 and κ_3. For $n = 2$, (2.22) reads

$$A_2(\varphi) = Y_{C_1} * C_2\big(A_1(\varphi)^2\big)$$
$$(3.4) \qquad\qquad = \kappa_1 * C_2(\kappa_1 * \varphi, \kappa_1 * \varphi).$$

Evaluating this at t we obtain

(3.5)

$$A_2(\varphi)(t) = \int \mathrm{d}u\ \kappa_1(t-u)\, C_2\bigg[\int \mathrm{d}s_1\ \kappa_1(u-s_1)\,\varphi(s_1), \int \mathrm{d}s_2\ \kappa_1(u-s_2)\,\varphi(s_2)\bigg]$$

$$= \iint \mathrm{d}s_1\, \mathrm{d}s_2\ \bigg\{\int \mathrm{d}u\ \kappa_1(u)\, C_2\big(\kappa_1(t-s_1-u)(\cdot), \kappa_1(t-s_2-u)(\cdot)\big)\bigg\}\, (\varphi(s_1), \varphi(s_2))$$

$$=: \iint \mathrm{d}s_1\, \mathrm{d}s_2\ \kappa_2(t-s_1, t-s_2)\big(\varphi(s_1), \varphi(s_2)\big).$$

Here we have exploited the fact that the C_n are linear in each argument so that we may interchange the linear operation with C_2 and the (Bochner) integration over s_1 and s_2. The dots indicate the positions where the arguments $\varphi(s_1)$ and $\varphi(s_2)$ are to be plugged in. The braced subexpression corresponds to the second-order kernel $\kappa_2(t-s_1, t-s_2) \in \mathcal{L}_2(E; E)$. Comparing the left two expressions in (3.5) we readily see that

$$(3.6) \qquad \kappa_2(s_1, s_2) = \int_0^{\min(s_1, s_2)} \mathrm{d}u\ \ \kappa_1(u)\, C_2\big(\kappa_1(s_1-u)(\cdot), \kappa_1(s_2-u)(\cdot)\big),$$

which is symmetric in s_1 and s_2. The integration limits are due to the fact that $\kappa_1(t) = 0$ for $t < 0$. Also $\kappa_2(s_1, s_2) = 0$ if $s_1 < 0$ or $s_2 < 0$.

Finally, for $n = 3$, (2.22) reads as

$$A_3(\varphi) = Y_{C_1} * \sum_{l=2}^{3} \sum_{\substack{j_1+j_2=l \\ j_1+2j_2=3}} \binom{l}{j_1 j_2} C_l\big(A_1(\varphi)^{j_1}, A_2(\varphi)^{j_2}\big)$$

$$(3.7) \qquad\qquad = Y_{C_1} * \big\{2\, C_2\,(A_1(\varphi), A_2(\varphi)) + C_3\,\big(A_1(\varphi)^3\big)\big\}.$$

In order to evaluate this expression at time t we handle the two contributing terms separately. First, we have

$$(3.8) \quad 2\, \{Y_{C_1} * C_2(A_1(\varphi), A_2(\varphi))\}\, (t)$$

$$= 2\, \iiint \mathrm{d}s_1\, \mathrm{d}s_2\, \mathrm{d}s_3 \int \mathrm{d}u\ \kappa_1(t-u)\, C_2\big(\kappa_1(u-s_1)\,\varphi(s_1), \kappa_2(u-s_2, u-s_3)\,\varphi(s_2)\,\varphi(s_3)\big).$$

The second term boils down to

$$(3.9) \quad \big\{Y_{C_1} * C_3\big(A_1(\varphi)^3\big)\big\}\, (t)$$

$$= \iiint \mathrm{d}s_1\, \mathrm{d}s_2\, \mathrm{d}s_3 \int \mathrm{d}u\, \kappa_1(t-u)\, C_3\big(\kappa_1(u-s_1)\varphi(s_1), \kappa_1(u-s_2)\varphi(s_2), \kappa_1(u-s_3)\varphi(s_3)\big).$$

From (3.8) and (3.9) we read off the third-order kernel κ_3,

$$\kappa_3(s_1, s_2, s_3) = \int_0^{\min(s_1,s_2,s_3)} du \ \kappa_1(u) \left\{ 2\, C_2\big(\kappa_1(s_1 - u)(\cdot), \kappa_2(s_2 - u, s_3 - u)(\cdot, \cdot)\big) \right.$$

$$(3.10) \qquad\qquad\qquad \left. + C_3\big(\kappa_1(s_1 - u)(\cdot), \kappa_1(s_2 - u)(\cdot), \kappa_1(s_3 - u)(\cdot)\big) \right\}.$$

This expression is apparently *not* symmetric in s_1, s_2, and s_3. The reason is that we have derived the polar form of A_3, i.e., the three-linear operator with three different arguments $A_3(\varphi_1, \varphi_2, \varphi_3)$, from an equation where A_3 appears only in its polynomial form $A_3(\varphi)$. There is, however, nothing wrong in using the nonsymmetric form of κ_3 because the sole context in which this kernel appears is in connection with three *identical* arguments $\varphi_1 = \varphi_2 = \varphi_3$. Of course, we could do a cosmetic symmetrization and replace κ_3 by

$$(3.11) \qquad\qquad \tilde{\kappa}_3(s_1, s_2, s_3) = \frac{1}{3!} \sum_{\sigma \in S_3} \kappa(s_{\sigma(1)}, s_{\sigma(2)}, s_{\sigma(3)}),$$

where S_3 is the set of all permutations of $\{1, 2, 3\}$.

Once we have seen how the procedure works in the cases $n = 2$ and $n = 3$, we can easily provide the general form of the nth-order kernel κ_n. Equation (2.22) reappears as

$$(3.12) \quad \kappa_n(s_1, \ldots, s_n) = \int_0^{\min(s_1,\ldots,s_n)} du \ \kappa_1(u) \sum_{l=2}^{n} \sum_{\substack{j_1 + \cdots + j_{n-1} = l \\ j_1 + 2j_2 + \cdots + (n-1)\, j_{n-1} = n}} \binom{l}{j_1 \ldots j_n}$$

$$\times\, C_l \Big(\underbrace{\kappa_1(s_1 - u), \ldots, \kappa_1(s_{j_1} - u)}_{j_1 \text{ times}},$$

$$\underbrace{\kappa_2(s_{j_1+1} - u, s_{j_1+2} - u), \ldots, \kappa_2(s_{j_1+2j_2-1} - u, s_{j_1+2j_2} - u)}_{j_2 \text{ times}}, \ldots, \underbrace{\kappa_{n-1}(\ldots)}_{\ldots} \Big).$$

In order to avoid blowing up this expression even more, we have omitted the dots indicating the arguments of the operators $\kappa_1, \ldots, \kappa_{n-1}$. It is plain that the operator κ_1 takes one argument, the operator κ_2 two, and so on. The present expression can be read as an operator equation if we define the operation of C_l on the operators κ_k in the natural way, i.e., $C_1(\kappa_1)(\varphi) = C_1(\kappa_1(\varphi))$, and so on. A consequence of (3.12) is the following.

THEOREM 1. *For $n \geq 2$, the kernels κ_n are* continuous *functions on $\mathbb{R}^n$ and vanish on the complement of the hyperquadrant $[0, \infty)^n$. For all $n \geq 1$, we have*

$$(3.13) \qquad\qquad\qquad \operatorname{supp}(\kappa_n) \subset [0, \infty)^n.$$

Moreover, the kernels κ_n are real-analytic (polynomial-exponential) functions on $(0, \infty)^n$, outside the "diagonal skeleton" $s_i = s_j$ with $i \neq j$.

Detailed expressions illustrating the theorem can be found in section 5.

4. Tensor notation. If E is a finite-dimensional vector space, e.g., $E = \mathbb{R}^N$, tensor calculus may provide a convenient framework for the calculations that have been presented above in a coordinate-free notation.

In the following we will use lower indices to denote vectors and upper indices for elements from the dual space E'. We will also employ the Einstein summation rule: every pair of identical upper and lower indices has to be summed over from 1 to N. In order to avoid confusion with indices denoting the order of the operators κ_n and c_n and those denoting specific components, we omit the indices that give the order of the operators and identify these objects by the number of their arguments or indices.

The differential equation in tensor notation is

$$(4.1) \quad x_i(t)' = f_i(\boldsymbol{x}(t)) + \varphi_i(t),$$
$$\boldsymbol{x} = (x_i)_{1 \leq i \leq N} \in C_+^1(\mathbb{R}, E), \quad \boldsymbol{\varphi} = (\varphi_i)_{1 \leq i \leq N} \in C_+(\mathbb{R}, E),$$

and the Taylor series for $\boldsymbol{f}$ reads

$$(4.2) \qquad f_i(x) = \sum_{k=1}^{\infty} c_i^{j_1 \, j_2 \ldots j_k} \, x_{j_1} \, x_{j_2} \ldots x_{j_k},$$

with

$$(4.3)$$
$$c_i^{j_1 \ldots j_k} = \frac{1}{k!} \, \frac{\partial^k}{\partial x_{j_1} \ldots \partial x_{j_k}} f_i(x) \bigg|_{x=0}, \quad \boldsymbol{c}_k = \left(c_i^{j_1 \ldots j_k} \right)_{1 \leq i \leq N}^{1 \leq j_1 \ldots j_k \leq N} \in E \otimes \underbrace{E' \otimes \cdots \otimes E'}_{k \text{ times}}.$$

The first-order kernel is the fundamental solution of the linearized equation. This can be translated into tensor notation as

$$(4.4) \qquad \kappa_i^j(s) = Y(s) \left[\exp(\boldsymbol{c}_1) \right]_i^j.$$

The expression (3.6) for the second-order kernel is

$$(4.5) \qquad \kappa_i^{j_1 j_2}(s_1, s_2) = \int_0^{\min(s_1, s_2)} \mathrm{d}u \; C_m^{k_1 k_2} \, \kappa_i^m(u) \, \kappa_{k_1}^{j_1}(s_1 - u) \, \kappa_{k_2}^{j_2}(s_2 - u),$$

and the third-order kernel, as given by (3.10), is

$$(4.6)$$
$$\kappa_i^{j_1 j_2 j_3}(s_1, s_2, s_3) = \int_0^{\min(s_1, s_2, s_3)} \mathrm{d}u \; \kappa_i^m(u) \Big\{ 2\, C_m^{k_1 k_2} \, \kappa_{k_1}^{j_1}(s_1 - u) \, \kappa_{k_2}^{j_1 j_2}(s_2 - u, s_3 - u)$$
$$+ C_m^{k_1 k_2 k_3} \, \kappa_{k_1}^{j_1}(s_1 - u) \, \kappa_{k_2}^{j_2}(s_2 - u) \, \kappa_{k_3}^{j_3}(s_3 - u) \Big\}.$$

In the next section we take advantage of the above formalism and calculate the first few kernels for a couple of examples.

5. Examples. In this section we present some examples illustrating our procedure. First we treat the scalar case ($\mathbb{R}$) and then turn to the vector case ($\mathbb{R}^N$), where the first derivative of f at $x = 0$, a matrix, is (or is not) diagonable.

5.1. Scalar case. Let us calculate the first few kernels for a one-dimensional differential equation explicitly. The first-order kernel is the fundamental solution (or the Green's function located at zero) so that we obtain

$$(5.1) \qquad \kappa_1(s_1) = Y(s_1) \, e^{c_1 \, s_1}.$$

The second-order kernel is given in (4.5), i.e.,

$$(5.2) \qquad \kappa_2(s_1, s_2) = \begin{cases} Y(s_1, s_2) \frac{c_2}{c_1} e^{c_1 (s_1+s_2)} \left[1 - e^{-c_1 \min(s_1,s_2)} \right], & c_1 \neq 0, \\ Y(s_1, s_2) c_2 \min(s_1, s_2), & c_1 = 0. \end{cases}$$

If $c_1 \neq 0$, the third-order kernel turns out to be

$$(5.3) \quad \kappa_3(s_1, s_2, s_3) = Y(s_1, s_2, s_3) \left(\frac{c_2}{c_1^2} + \frac{c_3}{2c_1} \right) e^{c_1 (s_1+s_2+s_3)} \left[1 - e^{2c_1 \min(s_1,s_2,s_3)} \right]$$
$$- Y(s_1, s_2, s_3) \frac{2c_2^2}{c_1^2} e^{-c_1 \min(s_2,s_3)} \left[1 - e^{-c_1 \min(s_1,s_2,s_3)} \right].$$

If $c_1 = 0$, we have

$$(5.4)$$
$$\kappa_3(s_1, s_2, s_3) = Y(s_1, s_2, s_3) \min(s_1, s_2, s_3) \left[c_3 - 2c_2^2 \min(s_1, s_2, s_3) \min(s_2, s_3) \right].$$

5.2. Vector case with diagonal Jacobi matrix. We start with an N-dimensional system of differential equations and assume the Jacobi matrix $(\partial f_j / \partial x_k)$ at $x = 0$ to be diagonable. If we denote the eigenvalues of the Jacobi matrix by λ_j, $1 \leq j \leq N$, we can put the differential equation in the form

$$(5.5) \qquad x_j(t)' = f_j(x(t)) + \varphi_j(t), \quad \text{with} \quad c_j^k \equiv \frac{\partial f_j}{\partial x_k}(0) = \lambda_j \, \delta_j^k.$$

The fundamental solution to the linearized equation is

$$(5.6) \qquad \kappa_j^k(s) = \delta_j^k \, Y(s) \, e^{\lambda_j s}.$$

Using (4.5) we can readily calculate the second-order kernel,

$$\kappa_i^{j_1 j_2}(s_1, s_2) = C_i^{j_1 j_2} \int_0^{\min(s_1,s_2)} \mathrm{d}u \; Y(u) \exp \left[\lambda_i u + \lambda_{j_1} (s_1 - u) + \lambda_{j_2} (s_2 - u) \right]$$
$$= \begin{cases} C_i^{j_1 j_2} Y(s_1, s_2) \exp (\lambda_{j_1} s_1 + \lambda_{j_2} s_2) \frac{\exp[(\lambda_i - \lambda_{j_1} - \lambda_{j_2}) \min(s_1,s_2)] - 1}{\lambda_i - \lambda_{j_1} - \lambda_{j_2}} & \text{if } \lambda_i \neq \lambda_{j_1} + \lambda_{j_2}, \\ C_i^{j_1 j_2} Y(s_1, s_2) \exp (\lambda_{j_1} s_1 + \lambda_{j_2} s_2) \min(s_1, s_2) & \text{if } \lambda_i = \lambda_{j_1} + \lambda_{j_2}. \end{cases}$$

In general, this kernel is not diagonal and the calculation of the nonsymmetric part of the third-order kernel in (4.6) becomes rather tedious. From the general form (3.12), however, we can deduce that all kernels κ_n are linear combinations of pure exponentials if the eigenvalues $\lambda_1, \ldots, \lambda_N$ are *incommensurable*.

5.3. Vector case with nondiagonal Jacobi matrix. If some of the eigenvalues of the linearized equation are degenerate, the Jacobi matrix $(\partial f_j / \partial x_k)$ at $x = 0$ need not be diagonable. In order to illustrate this situation, we investigate the case with $N = 2$ and the Jacobi matrix being in Jordan normal form, i.e.,

$$(5.7) \qquad \left(\frac{\partial f_j}{\partial x_k}(0) \right)_{1 \leq j,k \leq 2} = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

The first-order kernel is the fundamental solution of the linearized equation,

$$(5.8) \qquad \left(\kappa_j^k(s) \right)_{1 \leq j,k \leq 2} = \begin{pmatrix} e^{\lambda s} & s \, e^{\lambda s} \\ 0 & e^{\lambda s} \end{pmatrix} Y(s).$$

Even in this simple case the expressions for the second-order kernel are rather lengthy. Here we show only a few exemplary components of $\kappa_j^{k_1 k_2}$:

$$\kappa_1^{11}(s_1, s_2) = Y(s_1, s_2) \frac{e^{\lambda (s_1 + s_2)}}{\lambda^2} \left\{ \lambda c_1^{11} + c_2^{11} \right\}$$

$$(5.9) \qquad - Y(s_1, s_2) \frac{e^{\lambda \max(s_1, s_2)}}{\lambda^2} \left\{ \lambda c_1^{11} + c_2^{11} \left[1 + \lambda \min(s_1, s_2) \right] \right\},$$

$$\kappa_1^{22}(s_1, s_2) = Y(s_1, s_2) \frac{e^{\lambda(s_1 + s_2)}}{\lambda^3} \left\{ c_1^{22} \lambda^2 + c_1^{12} \left[-2\lambda + \lambda^2 (s_1 + s_2) \right] \right.$$

$$\left. + c_2^{12} \left[-4 + \lambda (s_1 + s_2) \right] + c_2^{22} \lambda \right\}$$

$$- Y(s_1, s_2) \frac{e^{\lambda \max(s_1, s_2)}}{\lambda^3} \left\{ c_1^{22} \lambda^2 + c_1^{12} \left[-2\lambda + \lambda^2 |s_1 - s_2| \right] \right.$$

$$+ c_2^{12} \left[-4 + \lambda (s_1 + s_2 - 4\min(s_1, s_2)) + \lambda^2 \min(s_1, s_2) |s_1 - s_2| \right]$$

$$(5.10) \qquad \left. + c_2^{22} \lambda \left[1 + \lambda \min(s_1, s_2) \right] \right\},$$

$$(5.11)$$

$$\kappa_2^{11}(s_1, s_2) = Y(s_1, s_2) \frac{c_2^{11} e^{\lambda(s_1 + s_2)}}{\lambda} \left\{ 1 - e^{-\lambda \min(s_1, s_2)} \right\},$$

$$\kappa_2^{22}(s_1, s_2) = Y(s_1, s_2) \frac{e^{\lambda(s_1 + s_2)}}{\lambda^2} \left\{ c_2^{12} \left[-2 + \lambda (s_1 + s_2) \right] + c_2^{22} \lambda \right\}$$

$$(5.12) \qquad - Y(s_1, s_2) \frac{e^{\lambda \max(s_1, s_2)}}{\lambda^2} \left\{ c_2^{12} \left[-2 + \lambda \max(s_1, s_2) \right] + c_2^{22} \lambda \right\}.$$

In contrast to the nondegenerate case, the kernels now contain combinations of exponentials *and* polynomials in the s_i.

6. Sampling input space by delta functions. By the results of [37, sect. 7] it follows that the differential equation (1.1) has a solution, if φ is a *measure* with compact support, and that the solution, in the sense of distributions, exists in the space $\mathcal{R}(\mathbb{R}, E)$ of right-continuous functions having left limits, with values in E. The unique nonanticipative solution operator

$$A : \mathcal{M}_c(E) \longrightarrow \mathcal{R}(\mathbb{R}, E)$$

is, moreover, Q-analytic; cf. [37, Thm. 7.1] for a more precise result.

Q-analyticity ensures uniform convergence, for t in a given compact interval I, of the series

$$(6.1) \qquad A(\varphi)(t) = \sum_{n=1}^{\infty} A_n(\varphi)(t),$$

for φ in a sufficiently small neighborhood, $O = O(I)$, of zero in $\mathcal{M}_c(E)$. Moreover, since for $t \in I$ and $\varphi \in O$ (6.1) is a Taylor series, it follows that the series may be differentiated termwise with respect to φ as often as desired, a fact which we will use below when differentiating with respect to the parameters λ_i.

Let us consider an "input" φ that is made up of a sum of δ pulses at times $t_1, \ldots, t_n$, multiplied by real prefactors $\lambda_1, \ldots \lambda_n$,

$$(6.2) \qquad \varphi = \sum_{i=1}^{n} \lambda_i \delta_{t_i}, \quad \text{with} \quad \delta_{t_i}(t) = \delta(t - t_i).$$

Then the series (6.1) converges uniformly on the interval I, provided the numbers λ_i are sufficiently small. The corresponding "output" $x(t) = (A\,\varphi)(t)$ is of the form

$$(6.3) \qquad x(t) = \sum_{i=1}^{n} \lambda_i\,\kappa_1(t - t_i) + \sum_{i,j=1}^{n} \lambda_i\,\lambda_j\,\kappa_2(t - t_i, t - t_j) + \cdots.$$

We can single out the terms of order k by taking a kth derivative with respect to the parameters λ_i at $\lambda_1 = \cdots = \lambda_n = 0$. In doing so, we obtain

$$(6.4) \qquad \left.\frac{\partial x(t)}{\partial \lambda_i}\right|_{\lambda_{1\ldots n}=0} = \kappa_1(t - t_i)\,,$$

$$(6.5) \qquad \left.\frac{\partial^2 x(t)}{\partial \lambda_i\,\partial \lambda_j}\right|_{\lambda_{1\ldots n}=0} = \kappa_2(t - t_i, t - t_j) + \kappa_2(t - t_j, t - t_i)\,,$$

$$\vdots$$

$$(6.6) \qquad \left.\frac{\partial^n x(t)}{\partial \lambda_1\,\ldots\,\partial \lambda_n}\right|_{\lambda_{1\ldots n}=0} = \sum_{\sigma \in S_n} \kappa_n(t - t_{\sigma(1)}, \ldots, t - t_{\sigma(n)}).$$

Here S_n is the permutation group of order n and $\lambda_{1,\ldots,n}$ is an abbreviation of the vector $(\lambda_1, \ldots, \lambda_n)$.

Since only the symmetric components of the kernels κ are relevant, we may assume that the kernel κ_n *is* symmetric. So all terms in the sum over all possible permutations of the indices $\{1, \ldots, n\}$ in (6.6) are equal and we obtain an explicit expression for the nth-order kernel κ_n,

$$(6.7) \qquad \kappa_n(t - t_1, \ldots, t - t_n) = \frac{1}{n!}\,\left.\frac{\partial^n x(t)}{\partial \lambda_1\,\ldots\,\partial \lambda_n}\right|_{\lambda_{1\ldots n}=0}.$$

If we introduce the abbreviation

$$(6.8) \qquad \kappa_{t_1,\ldots,t_n}(t) = \kappa_n(t - t_1, \ldots, t - t_n)$$

and resubstitute the definition of x,

$$\kappa_{t_1,\ldots,t_n}(t) = \frac{1}{n!}\,\left.\frac{\partial^n A\big[\sum_{i=1}^n \lambda_i\,\delta(t - t_i)\big]}{\partial \lambda_1\,\ldots\,\partial \lambda_n}\right|_{\lambda_{1\ldots n}=0}$$

$$(6.9) \qquad \equiv \frac{1}{n!}\,\mathrm{D}^n A(0;\,\delta_{t_1}, \ldots, \delta_{t_n})(t),$$

we see that $\kappa_{t_1,\ldots,t_n}$ is simply the nth derivative of the solution operator A taken at $\varphi = 0$ in the directions $\delta_{t_1}, \ldots, \delta_{t_n}$. With a convenient abbreviation this amounts to

$$(6.10) \qquad \kappa_{t_1,\ldots,t_n} = A_n(\delta_{t_1}, \ldots, \delta_{t_n}),$$

where A_n is n-linear symmetric and such that, as previously, $A_n(\varphi) = A_n(\varphi, \ldots, \varphi)$.

In [37] it was noted that the Q-analyticity on the space of measures with the strong topology was not satisfactory because, in this topology, the subspace of absolutely continuous measures is closed, and δ_h does not converge to $\delta_0 \equiv \delta$ as $h \to 0$. The recurrence relations for the operators A_n allow us to prove their continuity with respect to a more convenient, weaker topology.

We consider, on the space of measures with compact support, the so-called *weak topology* for which

(6.11) $$\varphi_i \to \varphi \iff \langle \varphi_i, f \rangle \to \langle \varphi, f \rangle$$

for all real-valued continuous functions f. Here $\langle \varphi, f \rangle = \int \varphi(ds) f(s)$. If φ is vector-valued, the quantity $\langle \varphi, f \rangle$ still makes sense as a vector in E. For the sake of simplicity we assume E to be finite-dimensional. We note that standard approximations of the Dirac delta function, with support in a fixed compact set, converge in the weak topology.

THEOREM 2 (continuity theorem). *Let E be finite-dimensional. Let the measures $\varphi_i \in \mathcal{M}_c(E)$ with $i \in \mathbb{N}$ converge weakly to the measure φ, keeping their support in a fixed compact set. Then the following results hold:*

(i) $A_1(\varphi_i) = \kappa_1 * \varphi_i$ *converges to $A_1(\varphi)$ in the space $L^1_{\text{loc}}(\mathbb{R}, E)$.*

(ii) *If $n \geq 2$, then $A_n(\varphi_i)(t)$ converges to $A_n(\varphi)(t)$ uniformly for t in a compact subset of $\mathbb{R}$.*

(iii) *If for $1 \leq j \leq n$ the measures $\varphi_i^{(j)}$, with supports in a fixed compact set, converge weakly to $\varphi^{(j)}$ as $i \to \infty$, then we have*

(6.12) $$A_n(\varphi^{(1)}, \dots, \varphi^{(n)})(t) = \lim_{i \to \infty} A_n(\varphi_i^{(1)}, \dots, \varphi_i^{(n)})(t),$$

uniformly for t in a compact set.

Proof. (i) We first consider the case $E = \mathbb{R}$. It is well known that, if $\varphi_i \geq 0$ and f is a locally bounded Borel function, we have $\langle \varphi_i, f \rangle \to \langle \varphi, f \rangle$ provided f is continuous at the atoms of φ. Here, κ_1 being discontinuous at 0, this means that $\int \varphi_i(ds) \kappa_1(t-s)$ converges to $\int \varphi(ds) \kappa_1(t-s)$ for all t with $\varphi(\{t\}) = 0$, in particular, for all but countably many t. By the uniform boundedness principle the total masses of the φ_i remain bounded. Thus the functions $\kappa_1 * \varphi_i$ remain uniformly bounded on compact intervals and so, by Lebesgue's theorem, the convergence takes place in L^1_{loc}. If the measures φ_i are not nonnegative, we can extract a subsequence so that the positive and negative parts converge weakly, and for this subsequence the same conclusion holds. But since any subsequence of φ_i similarly contains a subsequence converging in L^1_{loc} to $\kappa_1 * \varphi$, and L^1_{loc} is metrizable, it follows by a standard argument that the sequence itself converges.

In the case $E = \mathbb{R}^N$ one can apply this argument to each component of the vector measure φ_i.

(ii) We have seen that, at least for some subsequence, $A_1(\varphi_i)$ converges to $A_1(\varphi)$ at almost every t, while remaining uniformly bounded on compact sets. Then it follows from Lebesgue's theorem that $A_2(\varphi_i)(t) = \kappa_1 * (C_2(A_1(\varphi_i), A_1(\varphi_i)))(t)$ converges uniformly on compact sets. In fact let $f_i(t) = C_2(\kappa_1 * \varphi_i(t), \kappa_1 * \varphi_i(t))$ and let $f(t)$ be similarly associated with φ, then, if φ_i and φ have support in $[a, b]$, it follows that f_i and f vanish for $t < a$ and that, for $t \geq a$,

$$A_2(\varphi_i)(t) - A_2(\varphi)(t) = \int_a^t \kappa_1(t-s)[f_i(s) - f(s)]ds.$$

If $|\kappa_1(t-s)| \leq M_{a,b}$ for $a \leq s \leq t \leq b$, then we have

$$|A_2(\varphi_i)(t) - A_2(\varphi)(t)| \leq M_{a,b} \int_a^b |f_i(s) - f(s)|ds,$$

which, by part (i), converges to zero, uniformly with respect to $t \in [a, b]$.

More generally, the induction formula (2.21) shows that $A_n(\varphi_i)(t)$ converges to $A_n(\varphi)(t)$ uniformly on compact sets. Since the space of continuous functions on $\mathbb{R}$ with the topology of uniform convergence on compact sets is metrizable, we see again that the conclusion holds for any sequence φ_i converging weakly to φ and with support contained in a fixed compact set.

(iii) In the case of A_2 we have the expression

$$A_2(\varphi^{(1)}, \varphi^{(2)}) = \frac{1}{2}[A_2(\varphi^{(1)} + \varphi^{(2)}) - A_2(\varphi^{(1)}) - A_2(\varphi^{(2)})].$$

For $n > 2$ there are similar, but more complicated, expressions of the symmetric multilinear form in terms of its restriction to the main diagonal. In fact, we have the following general lemma, in which we make use of the notation

$$(\Delta_h v)(x) = v(x + h) - v(x).$$

LEMMA 2 (polarization formula [38]). *Let X and Y be linear spaces over $\mathbb{R}$. Let $u : X^n \longrightarrow Y$ be a symmetric n-linear map. Let $v(x) = u(x, \dots, x)$ be the restriction to the diagonal. Then*

$$(6.13) \qquad u(x_1, \dots, x_n) = \frac{1}{n!}\big(\Delta_{x_n} \dots \Delta_{x_1} v\big)(0).$$

This implies, in particular, that if X and Y are topological vector spaces the multilinear map u is continuous if and only if its restriction to the main diagonal is continuous. Applying this with X, the space of measures with support in a fixed compact interval, equipped with the weak topology, we obtain the desired result.

As for the proof of the lemma, it suffices to prove it in the case of a finite-dimensional (n-dimensional) space X. So we may assume that X and Y are finite-dimensional, and equip them with their natural topologies. We have

$$(6.14) \qquad u(x_1, \dots, x_n) = \frac{1}{n!} \frac{\partial^n}{\partial t_1, \dots, \partial t_n} v(t_1 x_1 + \dots + t_n x_n),$$

whatever $t_1, \dots, t_n$. Integration of (6.14), n times in succession, between 0 and 1, now yields (6.13); cf. [27, Eq. 3], [34, Eq. 5.4–10] for a related result.

7. Numerical procedure: Differential sampling. Equation (6.7) provides a means of calculating the kernels κ_n numerically. If we put $t = 0$ and $s_i = -t_i$, $i = 1, \dots, n$, we get

$$(7.1) \qquad \kappa_n(s_1, \dots, s_n) = \frac{1}{n!} \left.\frac{\partial^n x(0)}{\partial\lambda_1 \dots \partial\lambda_n}\right|_{\lambda_{1\dots n}=0}, \quad \text{with } x = A\left[\sum_{i=1}^{n} \lambda_i \, \delta_{-s_i}\right].$$

The derivatives on the left-hand side of (7.1) can be expressed by a differential quotient, e.g.,

$$(7.2) \quad \kappa_n(s_1, \dots, s_n) = \frac{1}{n!} \lim_{\epsilon \to 0} \frac{1}{(2\epsilon)^n} \sum_{\{q_i = \pm 1; \, 1 \le i \le n\}} \left(\prod_{i=1}^{n} q_i\right) A\left[\epsilon \sum_{i=1}^{n} q_i \, \delta_{-s_i}\right](0).$$

In order to evaluate $\kappa_n(s_1, \dots, s_n)$ we have to determine the response of the system described by the operator A to 2^n different input functions φ, namely to $\varphi(t) =$

$\epsilon \sum_{i=1}^{n} [\pm \hat{\delta}(t + s_i)]$, with all possible combinations of positive and negative signs, $\hat{\delta}$ being an approximate δ function, and ϵ a "small" constant.

By the continuity theorem of section 6, the approximate kernels

$$(7.3) \qquad \hat{\kappa}_n(t - t_1, \ldots, t - t_n) = A_n(\hat{\delta}_{t_1}, \ldots, \hat{\delta}_{t_n})(t)$$

converge to the ideal kernel

$$(7.4) \qquad \kappa_n(t - t_1, \ldots, t - t_n) = A_n(\delta_{t_1}, \ldots, \delta_{t_n})(t).$$

Generalization to vector-valued systems is straightforward, the λ_k now being interpreted as vectors in E. In the finite-dimensional case the scalar sample input φ is replaced by

$$(7.5) \qquad \varphi_i = \sum_{k=1}^{n} \lambda_k \, \delta_i^{j_k} \, \delta_{t_k}.$$

Here, $\delta_i^{j_k}$ is the Kronecker δ and $\delta_{t_k} = \delta(t - t_k)$ the Dirac delta function as before. The upper indices $j_1, \ldots, j_n$ are kept fixed and used to select a specific component of the kernel we are interested in; cf. (4.5) and (4.6).

If we substitute this ansatz into the Volterra series, we obtain

$$(7.6) \qquad x_i(t) = \sum_{k=1}^{n} \lambda_k \, \kappa_i^{j_k}(t - t_k) + \sum_{k,l=1}^{n} \lambda_k \, \lambda_l \, \kappa_i^{j_k j_l}(t - t_k, t - t_l) + \cdots.$$

As before, we differentiate with respect to the parameters λ_k and obtain

$$(7.7) \qquad \left. \frac{\partial^n x_i(t)}{\partial \lambda_1 \ldots \partial \lambda_n} \right|_{\lambda_{1 \ldots n}=0} = \sum_{\sigma \in S_n} \kappa_i^{j_{\sigma(1)} \cdots j_{\sigma(n)}}(t - t_{\sigma(1)}, \ldots, t - t_{\sigma(n)}).$$

Analogous to the scalar case, we exploit the symmetry of the kernels which now involves a permutation of the arguments *and* of the indices of the kernel, i.e.,

$$(7.8) \qquad \kappa_i^{j_{\sigma(1)} \cdots j_{\sigma(n)}}(t - t_{\sigma(1)}, \ldots, t - t_{\sigma(n)}) = \kappa_i^{j_1 \cdots j_n}(t - t_1, \ldots, t - t_n).$$

The result is a generalization of (7.1),

$$(7.9) \qquad \kappa_i^{j_1 \ldots j_n}(s_1, \ldots, s_n) = \frac{1}{n!} \left. \frac{\partial^n x_i(0)}{\partial \lambda_1 \ldots \partial \lambda_n} \right|_{\lambda_{1 \ldots n}=0},$$

with

$$(7.10) \qquad x_i = A \left[\sum_{k=1}^{n} \lambda_k \, \delta_i^{j_k} \, \delta_{-s_k} \right].$$

This completes the description of the algorithm to compute the Volterra kernels from the output $x(t)$. It is fair to call it "differential sampling" as the Volterra kernels are obtained by sampling the input space through (approximate) delta functions of strength λ_i and differentiating the output with respect to the λ_i at zero so as to obtain the kernels themselves. As such it is a differential method. It is to be contrasted with integral methods such as that of Wiener [40], which need to perform an average over a stochastic process, an integral based on some probability space.

For practical purposes, it is not necessary to determine the Volterra kernels for all values of their arguments. It is a consequence of Theorem 1 that a finite ε net suffices. Here ε is to be determined by the numerics.

8. Discussion. We have studied the nonlinear differential equation (1.1) describing some process in a Banach space E, viz., $x' = f(x) + \varphi$ with φ representing a time-dependent input, here a "perturbation" of the equilibrium state $x = 0$. We have seen that differential sampling is a direct method of determining the Volterra kernels with respect to a given equilibrium state. With the benefit of hindsight, a natural question is now, What is the relation between differential sampling, taking a linear combination of delta functions as input, and the Wiener approach [40], starting with white noise?

If for the purpose of this discussion we assume E to be a finite-dimensional Euclidean space, we can show that our solution operator A can indeed have white noise, starting for instance at $t = 0$, as input. In fact, the integral equation

$$(8.1) \qquad x(t) = \int_{-\infty}^{t} f(x(s))ds + \Phi(t)$$

has a unique solution $x \in C_+(\mathbb{R}, E)$ for every $\Phi \in C_+(\mathbb{R}, E)$. In particular, if we take Φ at random in $C_0([0, +\infty), E) = \{\Phi \in C([0, +\infty), E) : \Phi(0) = 0\}$ according to Brownian motion, the resulting random output x may be viewed as $x = A(\varphi_{\mathrm{wn}})$ where φ_{wn} is white noise. Moreover, one can show that for each $t > 0$ the map $\varphi \mapsto x(t) = A(\varphi)(t)$ is in L^2 with respect to white noise and, hence, susceptible to be analyzed by Wiener's method, i.e., expanded in Fourier–Hermite functions [6]. This is explained extensively in the literature [40, 6, 31, 34].

In contrast to the Wiener approach, the present method need not generate white noise and, thus, avoids a notoriously hard task. Furthermore, instead of a Fourier–Hermite expansion we directly obtain the Volterra kernels themselves, a key advantage. The latter characterize the response of the system under consideration, whereas the former is intrinsic to white noise. Finally, we have implemented differential sampling by a straightforward numerical algorithm as described in section 7. The averaging over many white-noise realizations is now replaced by a sampling through *approximate* delta functions so as to numerically perform partial differentiation and obtain the Volterra kernels; cf. (7.9). It is here that our continuity theorem (section 6) pays off. Whereas a delta function poses no problem during the numerical integration of a differential equation, approximate delta functions are more appropriate in experiments.

From our point of view, the key idea behind the Wiener approach is that Gaussian white noise allows determination of the kernels because its variance is a *delta* function (concentrated at 0): $\langle \varphi_{\mathrm{wn}}(t)\varphi_{\mathrm{wn}}(t + s)\rangle = \delta(s)$, the angular brackets denoting an expectation value obtained through averaging. In the linear case where $x(t) = \int ds\, \kappa(t - s)\varphi(s)$ one finds

$$(8.2) \qquad \kappa_1(t - t') = \langle x(t)\varphi_{\mathrm{wn}}(t')\rangle.$$

A formal calculation shows the idea behind (8.2) most clearly:

$$\begin{aligned}
\langle x(t)\varphi_{\mathrm{wn}}(t')\rangle &= \int_0^{\infty} ds\, \kappa_1(t - s)\langle \varphi_{\mathrm{wn}}(s)\varphi_{\mathrm{wn}}(t')\rangle \\
&= \int_0^{\infty} ds\, \kappa_1(t - s)\delta(s - t') \\
(8.3) \qquad &= \kappa_1(t - t').
\end{aligned}$$

In the general case, the Wiener expansion, in contrast to its Volterra counterpart, uses an orthogonalization à la Gram–Schmidt which we do not discuss; see [14, 15, 18, 19, 6, 40].

In a sense, we have suggested returning ad fontes and taking advantage of Wiener's idea directly. That is to say, instead of generating delta functions through white noise, we directly focus on delta functions as input.

Appendix. The Hodgkin–Huxley equations [8, 16] constitute a four-dimensional, highly nonlinear system of ordinary differential equations. The first equation describes the time evolution of the membrane voltage V under the influence of a given input current I,

$$C\frac{\mathrm{d}V}{\mathrm{d}t} = I - [g_{\mathrm{Na}}(V - V_{\mathrm{Na}}) + g_{\mathrm{K}}(V - V_{\mathrm{K}}) + g_{\mathrm{L}}(V - V_{\mathrm{L}})],$$

where C is a capacitance per unit area, the terms on the right between square brackets describe ionic currents (Na^+, K^+, and Cl^-), and

$$g_{\mathrm{Na}} = \bar{g}_{\mathrm{Na}}\, m^3 h\ , \quad g_{\mathrm{K}} = \bar{g}_{\mathrm{K}}\, n^4$$

are time-dependent conductances with $\bar{g}_{\mathrm{Na}}$, $\bar{g}_{\mathrm{K}}$, and g_{L} being constants while $x = m$, n, and h are three auxiliary variables satisfying a differential equation of the form

$$\tau_x(V)\frac{\mathrm{d}x}{\mathrm{d}t} = x_\infty(V) - x.$$

If V were constant, e.g., under voltage clamp, then x would approach its limit $x_\infty(V)$ at an exponential rate determined by the relaxation time $\tau_x(V)$. The nonlinearity is here in the functions $\tau_x = 1/[\alpha_x + \beta_x]$ and $x_\infty(V) = \alpha_x/[\alpha_x + \beta_x]$, where α_x and β_x have been determined by Hodgkin and Huxley as a function of V through an extensive numerical fit. The terms V_{Na}, V_{K}, and V_{L} are reversal potentials, known constants. The Hodgkin–Huxley system is difficult to analyze [8, section 2.5]. The only variable that is accessible to experimentation is the membrane potential V. If the input φ is a current I, then the system we are left with is one-dimensional ($N = 1$). The Volterra kernel κ_2 of the corresponding Volterra series has been plotted in Figure 1.

REFERENCES

[1] A.M. AERTSEN AND P.I.M. JOHANNESMA, *The spectro-temporal receptive field. A functional characteristic of auditory neurons*, Biol. Cybernet., 42 (1981), pp. 133–143.

[2] E. BEDROSIAN AND S.O. RICE, *The output properties of Volterra systems (nonlinear systems with memory) driven by harmonic and Gaussian input*, Proc. IEEE, 59 (1971), pp. 1688–1707.

[3] E. DE BOER, *Correlation studies applied to the frequency resolution of the cochlea*, J. Aud. Res., 7 (1957), pp. 209–217.

[4] R.W. BROCKETT, *Volterra series and geometric control theory*, Automatica J. IFAC, 12 (1976), pp. 167–176.

[5] H.L. BRYANT AND J.P. SEGUNDO, *Spike initiation by transmembrane current: A white-noise analysis*, J. Physiol. (Lond.), 260 (1976), pp. 279–314.

[6] R.H. CAMERON AND W.T. MARTIN, *The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals*, Ann. Math., 48 (1947), pp. 385–392.

[7] K.H. CHON, T.J. MULLEN, AND R.J. COHEN, *A dual-input nonlinear system analysis of autonomic modulation of heart rate*, IEEE Trans. Biomed. Eng., 43 (1996), pp. 530–544.

[8] J. CRONIN, *Mathematical Aspects of Hodgkin-Huxley Theory*, Cambridge University Press, Cambridge, UK, 1987.

[9] P. VAN DIJK, P. WIT, AND J.M. SEGENHOUT, *Wiener kernel analysis of inner ear function in the American bullfrog*, J. Acoust. Soc. Am., 95 (1994), pp. 904–919.

[10] J.J. EGGERMONT, *Wiener and Volterra analyses applied to the auditory system*, Hear. Res., 66 (1993), pp. 177–201.

[11] A.S. FRENCH AND M.J. KORENBERG, *A nonlinear cascade model for action potential encoding in an insect sensory neuron*, Biophys. J., 55 (1989), pp. 655–661.

[12] A.S. FRENCH AND M.J. KORENBERG, *Dissection of a nonlinear cascade model for sensory encoding*, Ann. Biomed. Eng., 19 (1991), pp. 473–484.

[13] K.J. FRISTON, O. JOSEPHS, G. REES, AND R. TURNER, *Nonlinear event-related responses in fMRI*, Magn. Reson. Med., 39 (1998), pp. 41–52.

[14] T. HIDA, *Brownian Motion*, Springer-Verlag, Berlin, 1980. See in particular section 4.2 and Appendix A.4.

[15] T. HIDA, H.H KUO, J. POTTHOFF, AND L. STREIT, *White Noise*, Kluwer, Dordrecht, 1993.

[16] A.L. HODGKIN AND A.F. HUXLEY, *A quantitative description of ion currents and its application to conduction and excitation in nerve membranes*, J. Physiol. (Lond.), 117 (1952), pp. 500–544.

[17] S. JOEKEN, H. SCHWEGLER, AND C.P. RICHTER, *Modeling stochastic spike train responses of neurons: An extended Wiener series analysis of pigeon auditory nerve fibers*, Biol. Cybernet., 76 (1997), pp. 153–162.

[18] S. KAKUTANI, *Determination of the spectrum of the flow of Brownian motion*, Proc. Natl. Acad. Sci. USA, 36 (1950), pp. 319–323.

[19] S. KAKUTANI, *Spectral analysis of stationary Gaussian processes*, in Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 2, 1961, pp. 239–247.

[20] W.M. KISTLER, W. GERSTNER, AND J.L. VAN HEMMEN, *Reduction of the Hodgkin-Huxley equations to a single-variable threshold model*, Neural Comput., 9 (1997), pp. 1015–1045.

[21] M.J. KORENBERG, *A robust orthogonal algorithm for system identification and time-series analysis*, Biol. Cybernet., 60 (1989), pp. 267-276, and references quoted therein.

[22] Y.W. LEE AND M. SCHETZEN, *Measurements of the Wiener kernels of nonlinear systems by crosscorrelation*, Int. J. Control, 2 (1965), pp. 237–254.

[23] V.Z. MARMARELIS, *Identification of nonlinear biological systems using Laguerre expansions of kernels*, Ann. Biomed. Eng., 21 (1993), pp. 573–589.

[24] 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.

[25] V.Z. MARMARELIS, M.C. CITRON, AND C.P. VIVO, *Minimum-order Wiener modelling of spike-output systems*, Biol. Cybernet., 54 (1986), pp. 115–123.

[26] W.J. MELSSEN AND W.J. EPPING, *Selectivity for temporal characteristics of sound and interaural time difference of auditory midbrain neurons in the grassfrog: A system-theoretical approach*, Hear. Res., 60 (1992), pp. 178–198.

[27] E. NELSON, *Probability theory and Euclidean field theory*, in Constructive Quantum Field Theory, G. Velo and A.S. Wightman, eds., Lecture Notes in Phys. 25, Springer-Verlag, New York, 1973.

[28] G. PALM AND T. POGGIO, *The Volterra representation and the Wiener expansion: Validity and pitfalls*, SIAM J. Appl. Math., 33 (1977), pp. 195–216.

[29] G. PALM AND T. POGGIO, *Stochastic identification methods for nonlinear systems: An extension of the Wiener theory*, SIAM J. Appl. Math., 34 (1978), pp. 524–534.

[30] G. PALM, *On representation and approximation of nonlinear systems*, Biol. Cybernet., 31 (1978), pp. 119–124; *On representation and approximation of nonlinear systems: II. Discrete time*, Biol. Cybernet., 34 (1979), pp. 49–52.

[31] G. PALM AND B. PÖPEL, *Volterra representation and Wiener-like identification of nonlinear systems: scope and limitations*, Quart. Rev. Biophys., 18 (1985), pp. 135–164.

[32] W.J. RUGH, *Nonlinear System Theory—The Volterra/Wiener Approach*, The Johns Hopkins University Press, Baltimore, MD, 1981.

[33] M. SAKURANAGA, S. SATO, E. HIDA, AND K. NAKA, *Nonlinear analysis: Mathematical theory and biological applications*, Crit. Rev. Biomed. Eng., 14 (1986), pp. 127–184, and references quoted therein.

[34] M. SCHETZEN, *The Volterra and Wiener Theories on Nonlinear Systems*, Wiley, New York, 1980.

[35] L. SCHWARTZ, *Théorie des Distributions*, Hermann, Paris, 1966.

[36] L. SCHWARTZ, *Mathematics for the Physical Sciences*, Addison-Wesley, Reading, MA, 1966.

[37] E. G. F. Thomas, *Q-Analytic Solution Operators for Nonlinear Differential Equations*, Univ. of Groningen, preprint, 1997.

[38] E. G. F. Thomas, *A Polarization Identity for Multilinear Maps*, Univ. of Groningen, preprint, 1997.

[39] V. Volterra, *Theory of Functionals and of Integral and Integrodifferential Equations*, Blackie, London, 1930; reprinted by Dover, New York, 1959, Chap. I, pp. 25–30.

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