/2011

Annals of Biomedical Engineering, Vol. 19, pp. 383–399, 1991
Printed in the USA.

0090-6964/91 $3.00 + .00

Asymptotic Approach of Generalized Orthogonal Functional Expansions to Wiener Kernels

Jonathan D. Victor

Department of Neurology and Neuroscience
Cornell University Medical College
1300 York Avenue
New York, NY

and

Laboratory of Biophysics
The Rockefeller University
1230 York Avenue
New York, NY

(Received 5/4/90; Revised 3/21/91)

Wiener-like orthogonal functional expansions may be constructed with respect to test ensembles that are non-Gaussian, nonwhite, or both. Although the original Wiener expansion has particularly advantageous analytical properties, orthogonal expansions constructed with respect to other ensembles have practical advantages for laboratory implementation. We show how functional expansions based on two classes of input ensembles—white but non-Gaussian discrete noises and the sum of sinusoids—converge to the standard Wiener kernels. For discrete noises, the disparity between the standard and nonstandard kernels of a linear-static nonlinear transducer is proportional to the kurtosis of the input signal and inversely proportional to the ratio of the integration time of the linear filter to the time discretization. For the sum of sinusoids, the disparity is inversely proportional to the effective number of sinusoids passed by the initial linear stage.

Keywords—Central limit theorem; M-sequence; Nonlinear systems analysis; Orthogonal functional; Sum of sinusoids; Wiener kernel.

INTRODUCTION

Wiener's theory (1) of orthogonal functional expansions has provided a theoretical foundation for the experimental study of nonlinear transductions. In this approach, the input-output properties of a nonlinear transducer are represented as an infinite series of standard orthogonal functionals. Each of these standard functionals is described by an integral kernel. Qualitative properties of the kernel expansion provide clues to the internal structure of the transducer under study. A general introduction to physiological applications of the Wiener approach is provided by Marmarelis

---

Acknowledgments—This work was supported by NIH grant EY7977.

Address correspondence to Dr. Jonathan D. Victor, Department of Neurology and Neuroscience, Cornell University Medical College, 1300 York Avenue, New York, NY.

and Marmarelis (10); Sakai *et al.* (11) review applications of this approach to visual physiology.

Laboratory implementation of the Wiener theory is impossible without making certain approximations. This is because rigorous measurement of the Wiener kernels would require that the transducer be tested with Gaussian white noise. Ideal Gaussian white noise cannot be realized in the laboratory because of its infinite bandwidth and its infinite variance. Because of this limitation, a number of related approaches have been developed in which an alternative stimulus ensemble replaces Gaussian white noise as the test input. Such approaches include a multilevel random stimulus in discrete time (9), a binary pseudorandom stimulus (M-sequence) in discrete time (12), and a sum of multiple sinusoids (14).

Despite the impossibility of rigorous measurement of Wiener kernels, Wiener kernels play a central role in interpreting orthogonal functional expansions. Reasons for this include 1. the simple analytic form of Wiener kernels of particular transducers (6), 2. relationships between Wiener kernels determined with white noises of different means and powers (10), and 3. a maximum-entropy property of the Wiener expansion (13).

In this article, I analyze how the orthogonal functional expansions based on a variety of commonly used test inputs differ from the standard Wiener expansion, and the rapidity with which they converge to the standard Wiener expansion in a useful limiting case. With a knowledge of these differences, the investigator is in a better position to interpret experimentally determined orthogonal expansions and to select an efficient procedure for investigating a particular transducer.

RESULTS

Overview

The goal of this analysis is to compare orthogonal expansions based on a variety of different stimulus ensembles. For this purpose, I will choose a reference stimulus ensemble and a test set of nonlinear transducers. The reference stimulus ensemble is ideal Gaussian white noise. The transducers used to make the comparison will consist of a linear filter L followed by an instantaneous nonlinearity N. The nonlinearity will be taken to be a Hermite polynomial H_n associated with a Gaussian of variance P, where P is the power passed by the initial linear filter. This transducer will be denoted by LH_n. Its response $[(LH_n)s](t)$ to the stimulus $s(t)$ is given by

$$[(LH_n)s](t) = H_n\left(\int_0^\infty L(\tau)s(t-\tau)\,d\tau\right) \tag{1}$$

where $L(\tau)$ is the impulse response of the linear filter L.

The Wiener expansion of these transducers consists only of a single nth order term, as exhibited in Eq. 1. Thus, the degree to which the Wiener expansion differs from an alternative expansion can be gauged by the extent to which the expansion of the transducer LH_n has components of order j, for $j \neq n$. We will calculate the orthogonal expansions of the transducers LH_n with respect to two kinds of input ensembles: a discrete multilevel stimulus based on an even-symmetric distribution, and a sum of sinusoids.

For the discrete multilevel stimulus, we are interested in asymptotic behavior as the time discretization becomes short. For the sum of sinusoids, we are interested in asymptotic behavior as the number of sinusoids becomes large. Before the detailed analysis is undertaken, certain features of the result may be discerned. For any stimulus ensemble Ω, the transducer LH_n is exactly represented by an orthogonal expansion of order n. Therefore, terms of order $j > n$ do not enter into this expansion. The even-symmetric nature of the stimulus ensemble permits another simplification: The only terms of order j which enter into the expansion of LH_n are those for which the parity of j and n agree. Thus, the only terms which enter into the expansion of LH_n are those for which $j = n$, $n - 2$, $n - 4$, and so on.

This analysis reflects the point of view that the transducer LH_n is known, and its orthogonal expansion in the ensemble Ω is to be determined. In the laboratory, however, terms in the orthogonal expansion are measured, and the transducer is to be determined. For this reason, it is more useful to take a slightly different point of view: How does the transducer LH_n contribute to the jth term in an orthogonal expansion with respect to the stimulus ensemble Ω? The foregoing argument shows that the only transducers LH_n which contribute to the jth orthogonal term are those for which $n = j, j + 2, j + 4$, and so on. To assess the size of these contributions in a dimensionless fashion, we will calculate them for transducers LH_n rescaled to provide an output of variance unity.

In some sense, we are undertaking a generalization of the central limit theorem. The central limit theorem states that the distribution of a sum of a large number of independent random variables approaches that of a Gaussian, provided that no single variable dominates, and that the number Q of variables increases without bound (3). A measure of the departure of a composite even distribution from that of a Gaussian is its kurtosis. The kurtosis of the composite distribution of Q independent variables declines like $1/Q$.

In the current context, the distribution of a single variable is replaced by a probability space of stimuli, the ensemble Ω. For the sum-of-sinusoids stimulus, Q corresponds to the number of sinusoids which pass through the initial filter L. For the discrete multilevel stimulus, Q corresponds to the number of time intervals within the integration time of L. In either case, as Q grows, the cross-sectional distribution of the stimulus presented to the nonlinearity H_n approaches a Gaussian distribution. Thus, in analogy with the central limit theorem, one might expect that the kurtosis of the cross-sectional distribution of the input stimulus plays a critical role, and that kernels measured with a general ensemble Ω approach kernels measured with Gaussian white noise as $1/Q$ approaches zero.

Theoretical Background

The theoretical framework of Victor and Knight (14) will serve as the basis for the present calculations. We consider deterministic time-stationary transducers of finite memory. A transducer μ is specified by its responses to test stimuli $s(t)$. The response of the transducer μ to a test stimulus $s(t)$ is denoted by $[(\mu)s](t)$. Transducers $\mu, \nu, \ldots$ may be considered to be elements in a vector space $\hat{M}$. The addition of two transducers μ and ν results in a new transducer $\mu + \nu$ whose response to a stimulus $s(t)$ is

$$[(\mu + \nu)s](t) = [(\mu)s](t) + [(\nu)s](t) \ . \tag{2}$$

Multiplication of a transducer μ by a scalar α results in a new transducer $\alpha\mu$ whose response to a stimulus $s(t)$ is

$$[(\alpha\mu)s](t) = \alpha \cdot [(\mu)s](t) \ . \tag{3}$$

Once an ensemble Ω of test stimulus has been chosen, the vector space of transducers $\hat{M}$ has a natural quadratic form:

$$(\mu, v) = \langle[(\mu)s](t) \cdot [(v)s](t)\rangle_{\Omega,t} \tag{4}$$

where $\langle\ \rangle_{\Omega,t}$ denotes an average over the stimulus ensemble and time. Since we will consider only time-stationary transducers and stimulus ensembles, averages such as those in Eq. (4) may be taken over the stimulus ensemble alone, with t fixed at zero.

There is a natural notion of size $\|\mu\|$ associated with the quadratic form of Eq. 4:

$$\|\mu\|^2 = (\mu, \mu) \ . \tag{5}$$

Although Eq. 5 satisfies the triangle inequality, it may not represent a norm on $\hat{M}$. This is because some nonzero transducers μ_0 may satisfy $\|\mu_0\|^2 = 0$. Such transducers are those that produce no output for nearly all test stimuli in the ensemble Ω, and constitute a subspace $\hat{M}_0$. In other words, transducers μ_0 in $\hat{M}_0$ cannot be distinguished from the zero transducer by testing with stimuli drawn from Ω. Consequently, transducers v and v' whose difference $v - v'$ is in $\hat{M}_0$ cannot be distinguished by testing with stimuli drawn from Ω. To proceed with an orthogonal decomposition, we need to recognize that distinct transducers may not be distinguishable for particular choices of stimulus ensembles.

In formal terms, we replace the vector space $\hat{M}$ by the quotient vector space $M = \hat{M}/\hat{M}_0$ of distinguishable transducers. The symbols $\mu, v, \ldots$ now represent equivalence classes of indistinguishable transducers, rather than individual transducers themselves. With this understanding, Eq. 4 constitutes an inner product on M, Eq. 5 represents a quadratic norm on M, and M is a Hilbert space.

In this framework, one may generalize the orthogonalization of the Volterra expansion to arbitrary test ensembles Ω. (If Ω is taken to be Gaussian white noise, the standard Wiener orthogonal functional expansion will be obtained.) The Volterra series representation of a transducer μ is a infinite sum of transducers μ_j, each of which is a member of the homogeneous subspace M_j. The homogeneous subspaces are defined in a manner independent of the test ensemble Ω. By definition, members μ_j of M_j have an input-output relationship that may be written in the form

$$[(\mu_j)s](t) = \iint \ldots \int [\mu_j](\tau_1, \ldots, \tau_j)s(t - \tau_1) \cdot \ldots \cdot s(t - \tau_j)\, d\tau_1 \ldots d\tau_j \ , \tag{6}$$

for some kernel function $[\mu_j]$ of j prior times $\tau_1, \ldots, \tau_j$.

A convenient basis for the homogeneous subspace M_j is transducers whose kernel function is a product of delta functions. We use $D_{j,\mathbf{T}}$ to represent a transducer whose output is the product of its input at the j prior times $\mathbf{T} = (T_1, \ldots, T_j)$. The response of the transducer $D_{j,\mathbf{T}}$ to the stimulus $s(t)$ is given by

$$[(D_{j,\mathbf{T}})s](t) = s(t - T_1) \cdot \ldots \cdot s(t - T_j) \ . \tag{7}$$

A Volterra series for a transducer μ is simply a representation of μ as a sum of transducers, with each term in the sum drawn from the corresponding homogeneous subspace M_j:

$$\mu = \sum_{j=0}^{\infty} \mu_j \ . \tag{8}$$

Rearrangement of the Volterra series in Eq. 8 into an orthogonal functional series depends on the test ensemble Ω, because the inner product in Eq. (4) depends on Ω. The key step is reorganizing the homogeneous subspaces M_j into a set of orthogonal subspaces K_j which have the property that the linear span of $M_0, \ldots, M_n$ is equal to that of $K_0, \ldots, K_n$ (for any n). This reorganization amounts to a sequence of (linear) projection maps ϕ_j from M into K_j.

With the projections ϕ_j in hand, the Volterra series in Eq. 8 can be rewritten

$$\mu = \sum_{j=0}^{\infty} \kappa_j \ , \tag{9}$$

with each $\kappa_j = \phi_j(\mu)$ in the orthogonal subspace K_j.

The foregoing orthogonalization may be carried out for any test ensemble Ω. Further details and a discussion of how it relates to the usual white-noise procedures (8,10) may be found in Victor and Knight (14). Here, we focus on two cases, two kinds of stimulus ensemble: 1. a discrete multilevel stimulus, and 2. a sum of sinusoids.

The Discrete Multilevel Stimulus

An ensemble Ω of discrete multilevel stimuli is specified by a probability distribution $w(x)$ of amplitude levels and a discretization ΔT for time. Stimuli $s(t)$ in Ω have a constant amplitude for intervals of length ΔT. After each such interval, a new value for $s(t)$ is chosen according to the probability distribution $w(x)$. More formally, $s(t) = s_i/(\Delta T)^{1/2}$ for $-i\Delta T < t \leq -(i-1)\Delta T$, with s_i distributed according to $w(x)$. The factor $(\Delta T)^{1/2}$ is required to ensure that the test signal has finite power per unit bandwidth, in the limit $\Delta T \to 0$. Since $w(x)$ defines a probability distribution, it satisfies

$$\int_{-\infty}^{\infty} w(x)\, dx = 1 \ . \tag{10}$$

We will require that $w(x)$ has finite variance, which we will set equal to 1 for convenience:

$$\int_{-\infty}^{\infty} x^2 w(x)\, dx = 1 \ . \tag{11}$$

Finally, $w(x)$ is assumed to be an even-symmetric distribution: $w(x) = w(-x)$. As a consequence, the mean value $\langle s(t)\rangle_{\Omega,t}$ and all higher odd moments of $s(t)$ are zero.

The weight function $w(x)$ is associated with a series of orthogonal polynomials $B_m(x)$:

$$B_m(x) = x^m + b_{m,m-2}x^{m-2} + \dots \quad (12)$$

That is, we have chosen a normalization with the leading term $b_{m,m} = 1$. The polynomials $B_m(x)$ satisfy orthogonality relations

$$\int_{-\infty}^{\infty} B_m(x)B_{m'}(x)w(x)\,dx = \delta_{m,m'}b_m \ . \quad (13)$$

The even-symmetric nature of $w(x)$ and our normalization conditions in Eqs. 10, 11, and 12 constrain the first few orthogonal polynomials to be

$$B_0(x) = 1 \ , \quad (14a)$$

$$B_1(x) = x \ , \quad (14b)$$

$$B_2(x) = x^2 - \sigma_2 \ , \quad (14c)$$

$$B_3(x) = x^3 - \frac{\sigma_4}{\sigma_2} x \ , \quad (14d)$$

and

$$B_4(x) = x^4 - \frac{\sigma_6 - \sigma_4\sigma_2}{\sigma_4 - \sigma_2^2} x^2 + \frac{\sigma_6\sigma_2 - \sigma_4^2}{\sigma_4 - \sigma_2^2} \ . \quad (14e)$$

where $\sigma_k = \langle x^k \rangle$, the kth moment of the distribution $w(x)$, and σ_2 is necessarily 1.

We now consider the projection of the transducers $D_{j,\mathrm{T}}$ from M_j into the orthogonal subspace K_j. If all of the times $T_1, \dots, T_j$ are distinct, then the transducer $D_{j,\mathrm{T}}$ is automatically in the orthogonal subspace K_j. This is readily seen as follows: The response of $D_{j,\mathrm{T}}$ is the product of the stimulus at j prior times. Transducers in any of the lower-order homogeneous subspaces M_i $(i < j)$ can depend on stimulus values only at fewer times. Since the stimulus is assumed to be white (i.e., values at distinct prior times are independent), then the response of $D_{j,\mathrm{T}}$ must be independent of all transducers in lower-order homogeneous subspaces. Thus, $D_{j,\mathrm{T}}$ is orthogonal to M_i, for $i < j$.

If times $T_1, \dots, T_j$ are not all distinct, the orthogonal polynomials in Eq. 12 are needed to define the projection of $D_{j,\mathrm{T}}$ into the orthogonal subspace K_j. We use the integer vector $\mathrm{m} = (m_1, \dots, m_k, \dots)$ to count how many times each time lag is used: m_1 is the number of T_d's for which $T_d = \Delta T$, m_2 is the number of T_d's for which $T_d = 2\Delta T, \dots$. For transducers $D_{j,\mathrm{T}}$ in the jth homogeneous subspace M_j, $\sum m_k = j$.

With this notation, the response of the projection $\phi_j D_{j,\mathrm{T}}$ of the transducer $D_{j,\mathrm{T}}$ to the stimulus $s(t)$ is given by

$$[(\phi_j D_{j,\mathrm{T}})s](t) = \prod_{k=1}^{\infty} B_{m_k}(s(t - k\Delta T)) \ . \quad (15)$$

The formal infinite range of the product presents no difficulties, since only a finite number of the m_j's are nonzero, and $B_0(s) = 1$. Note that the projection ϕ_j leaves $D_{j,\mathrm{T}}$ unchanged if all times T_i are distinct (all $m_i = 0$ or 1), since $B_1(s) = s$. Furthermore, because of the orthogonality properties of the polynomials B_{m_i}, The projections $\phi_j D_{j,\mathrm{T}}$ constitute an orthogonal basis for K_j. The projection of an arbitrary transducer μ into K_j is defined by its projection along each basis vector $\phi_j D_{j,\mathrm{T}}$:

$$\phi_j \mu = \sum_{\mathrm{T}} C_{\mathrm{m,T}}(\mu) \phi_j D_{j,\mathrm{T}} \quad , \tag{16}$$

where

$$C_{\mathrm{m,T}}(\mu) = \frac{(\mu, \phi_j D_{j,\mathrm{T}})}{(\phi_j D_{j,\mathrm{T}}, \phi_j D_{j,\mathrm{T}})} \quad . \tag{17}$$

These inner products are defined by averages (see Eq. 4) with respect to the ensemble Ω.

Comparison of Orthogonal Kernels Measured with Different Stimuli

To compare the orthogonal decomposition corresponding to Gaussian white-noise inputs with the orthogonal decomposition corresponding to a discrete multilevel ensemble Ω, we will use Eq. 17 to calculate the coefficients in the expansion in Eq. 16 of $\phi_j(LH_n)$. The denominators of the terms on the right-hand side of Eq. 17 depend only on the orthogonal polynomials in Eq. 12 and their normalizations in Eq. 13:

$$(\phi_j D_{j,\mathrm{T}}, \phi_j D_{j,\mathrm{T}}) = \prod_{k=1}^{\infty} b_{m_k} \quad . \tag{18}$$

The numerators of the terms in the right-hand side of Eq. 17 contain the dependence on the transducer $\mu = LH_n$. We denote the numerators by $N_{\mathrm{m,T}}(LH_n)$, and begin the calculation by substituting Eq. 1 for LH_n and Eq. 7 for $\phi_j D_{j,\mathrm{T}}$:

$$N_{\mathrm{m,T}}(LH_n) = \left\langle H_n\left(\int_0^{\infty} L(\tau)s(t-\tau)\,d\tau\right) \prod_{k=1}^{\infty} B_{m_k}(s(t-k\Delta T))\right\rangle_{\Omega,t} \quad . \tag{19}$$

Since the signals $s(t - k\Delta T)$ are independently distributed according to $w(x)$, the average in Eq. 19 may be replaced by an ensemble-average at $t = 0$:

$$N_{\mathrm{m,T}}(LH_n) = \left\langle H_n\left(\int_0^{\infty} L(\tau)s(-\tau)\,d\tau\right) \prod_{k=1}^{\infty} B_{m_k}(s(-k\Delta T))\right\rangle_{\Omega} \quad . \tag{20}$$

Next, we exploit the fact that the ensemble Ω consists of signals that change only at discrete times $i\Delta T$. Thus, the stimulus can be considered as a discrete sequence of values, with each $s(-k\Delta T)$ given by $s_k/(\Delta T)^{1/2}$. The integral expression defining the linear transduction L reduces to a sum

$$\int_0^{\infty} L(\tau)s(-\tau)\,d\tau = \sum_{k=1}^{\infty} A_k s_k \quad . \tag{21}$$

where

$$A_k = (\Delta T)^{1/2} L(k\Delta T) \ . \tag{22}$$

With these reductions, the numerator $N_{m,T}(LH_n)$ may be written

$$N_{m,T}(LH_n) = \left\langle H_n\left(\sum_{k=1}^{\infty} A_k s_k\right) \prod_{k=1}^{\infty} B_{m_k}(s_k)\right\rangle \ . \tag{23}$$

where the average $\langle\ \rangle$ is calculated for each s_i distributed independently according to $w(x)$.

The average in Eq. 23 may be factored through the use of generating functions. We define the generating function $B(x;z)$ by

$$B(x;z) = \sum_{m=0}^{\infty} \frac{z^m}{m!} B_m(x) \ . \tag{24}$$

For the Hermite polynomials H_n constructed with respect to a Gaussian distribution of variance P, the generating function is

$$\begin{aligned} H(x;z) &= \sum_{n=0}^{\infty} \frac{z^n}{n!} H_n(x) \\ &= e^{zx - Pz^2/2} \ . \end{aligned} \tag{25}$$

We now calculate a generating function for $N_{m,T}(LH_n)$:

$$\begin{aligned} N_T(LH_n;y,z) &= \sum_{m,n} \prod_{k=1}^{\infty} \frac{y_k^{m_k}}{m_k!} \frac{z^n}{n!} N_{m,T}(LH_n) \\ &= \left\langle \prod_{k=1}^{\infty} B(s_k;y_k) \cdot H\left(\sum_{k=1}^{\infty} A_k s_k;z\right)\right\rangle \\ &= \left\langle \left(\prod_{k=1}^{\infty} B(s_k;y_k)\right) \cdot e^{z\Sigma A_k s_k}\right\rangle e^{-Pz^2/2} \ . \end{aligned} \tag{26}$$

The foregoing relationship holds for transducers LH_n constructed with respect to the variance P for the Hermite polynomial weight function. We have not yet exploited the specific properties postulated for $w(x)$ (i.e., zero-mean and unit variance), so Eq. 26 holds for more general $w(x)$ as well. Indeed, this expression can be viewed as a compact way of relating kernels obtained with a Gaussian noise to kernels obtained with a more general while noise and is similar to the result of Klein (4).

We now exploit the special features postulated for the ensemble Ω. To compare the Gaussian-based Wiener kernels with the kernels constructed with respect to Ω, the variance P is taken to be the variance of the response of the transducer L to the multilevel random signal. P therefore depends on the discretized impulse response for the transducer L. Since each value s_k is independently drawn from the distribution $w(x)$ whose variance is unity, we calculate

$$P = \left\langle \left(\sum_{1}^{\infty} A_k s_k \right)^2 \right\rangle$$
$$= \sum_{1}^{\infty} A_k^2 \tag{27}$$
$$= \frac{1}{\Delta T} \sum_{k=1}^{\infty} L(k\Delta T)^2 \ .$$

We use this relationship and the independence of each value s_k to obtain a factorization of the generating function in Eq. 26:

$$N_{\mathrm{T}}(LH_n;\mathbf{y},z) = \prod_{k=1}^{\infty} \langle\langle B(s_k;y_k)e^{zA_k s_k - A_k^2 z^2/2} \rangle\rangle \ . \tag{28}$$

Now, we focus on how the transducers LH_n project onto a particular basis vector $\phi_j D_{j,\mathrm{T}}$ in K_j characterized by m. We collect terms in Eq. 28 containing $\Pi(y_k^{m_k})$, and find

$$\sum_{n=0}^{\infty} \frac{z^n}{n!} N_{\mathbf{m},\mathrm{T}}(LH_n) = \prod_{k=1}^{\infty} F(m_k, A_k z) \ , \tag{29}$$

where

$$F(m,z) = \int B_m(s) e^{zs - z^2/2} w(s)\, ds \ . \tag{30}$$

The quantities $F(m,z)$ summarize how a transducer LH_n whose only nonzero Wiener kernel is of order n projects onto the typical basis vector of K_j. Considered as a power series in z, the first nonzero term of $F(m,z)$ is of order z^m. This is because lower powers of z, necessarily associated with powers of s less than m in Eq. 30, are orthogonal to $B_m(s)$. We re-express powers of s in terms of the orthogonal polynomials $B_m(s)$ to calculate a power series for $F(m,z)$ in terms of the coefficients $b_{m,k}$ and the normalization factors b_m of the orthogonal polynomials. The power series begins

$$F(m,z) = \frac{b_m}{m!} z^m \Bigg[1 - \left(\frac{1}{2} + \frac{b_{m+2,m}}{(m+2)(m+1)} \right) z^2 + \left(\frac{1}{8} + \frac{b_{m+2,m}}{2(m+2)(m+1)} + \frac{b_{m+4,m+2}\, b_{m+2,m} - b_{m+4,m}}{(m+4)(m+3)(m+2)(m+1)} \right) z^4 \ldots \Bigg] \ . \tag{31}$$

For the typical (nondiagonal) basis vector $\phi_j D_{j,\mathrm{T}}$, all components of m are 0 or 1. The first two terms of Eq. 31 for $F(0,z)$ and $F(1,z)$ depend only on the kurtosis

$\gamma = \sigma_4/\sigma_2^2 - 3$ of $w(x)$. Note that for $m = 0$, the z^2 term vanishes since $b_{2,0} = 1$ (Eq. 14c). Thus,

$$F(0,z) = 1 - \frac{\gamma}{24} z^4 \ldots \tag{32a}$$

and

$$F(1,z) = z + \frac{\gamma}{6} z^3 \ldots \quad . \tag{32b}$$

We now combine Eqs. 18, 29, and 32 to calculate the leading terms in the coordinates $C_{m,T}(LH_j)$ of the projection of LH_n into the jth orthogonal subspace. For $n < j$, this projection is zero, since the smallest power of z that can occur in $F(m_k, A_k z)$ is z^{m_k}, and $j = \sum m_k$. When n and j are of unequal parity, the projection is also zero, since $F(m, z)$ contains only alternate powers of z. For $n = j$, we find

$$C_{m,T}(LH_j) = j! \prod_{k=1}^{\infty} \frac{A_k^{m_k}}{m_k!} \quad . \tag{33}$$

For $n = j + 2$, there are contributions for all $m_k \neq 0$:

$$C_{m,T}(LH_{j+2}) = -(j+2)! \prod_{k=1}^{\infty} \frac{A_k^{m_k}}{m_k!} \sum_{m_k \neq 0} \left(\frac{1}{2} + \frac{b_{m_k+2, m_k}}{(m_k+2)(m_k+1)} \right) A_k^2 \quad . \tag{34}$$

For $n = j + 4$, there are two kinds of contributions. Pairs of factors in Eq. 29, with both $m_k \neq 0$ and $m_{k'} \neq 0$, lead to terms involving $A_k^2 A_{k'}^2$. Individual factors with $m_k = 0$ lead to terms involving A_k^4. We are interested in the limiting behavior as the discretization interval ΔT shrinks. Since the A_k are proportional to $(\Delta T)^{1/2}$ but the number of time lags is proportional to $1/\Delta T$, the contributions from factors with $m_k = 0$ dominate:

$$C_{m,T}(LH_{j+4}) = -(j+4)! \prod_{k=1}^{\infty} \frac{A_k^{m_k}}{m_k!} \left(\sum_{m_k=0} \frac{\gamma}{24} A_k^4 + 0((\Delta T)^2) \right) \quad . \tag{35}$$

For higher values of n (i.e., $n = j + 6, j + 8, \ldots$), the leading term in $C_{m,T}(LH_n)$ is $0((\Delta T)^2)$ or smaller.

The quantities in Eqs. 34 and 35 describe the projection of the transducers LH_{j+2} and LH_{j+4} (contained in the $(j + 2)$nd and $(j + 4)$th orthogonal subspaces of a standard Wiener expansion) into K_j, the jth orthogonal subspace with respect to Ω. For Gaussian inputs of unit variance, the mean-squared output of LH_n is $n!P^n$, where P (Eq. 27) is the sum of the squares of the A_k. To compare the projections of LH_{j+2} and LH_{j+4} with the projection of transducers LH_j, it is appropriate to normalize them by their magnitudes in the Gaussian norm in Eq. 5. We denote the ratios of these projections by $\rho_{j+2}(j)$ and $\rho_{j+4}(j)$:

$$\rho_n(j) = \left| \frac{C_{m,T}(LH_n)}{C_{m,T}(LH_j)} \right| \left(\frac{j!P^j}{n!P^n} \right)^{1/2} \quad . \tag{36}$$

For the typical nondiagonal term in which exactly j of the m_k are 1, and the rest are zero, we find

$$\rho_{j+2}(j) = \frac{|\gamma|}{6} \frac{[(j+2)(j+1)]^{1/2}}{P} \sum_{m_k=1} A_k^2 \tag{37}$$

and

$$\rho_{j+4}(j) = \frac{|\gamma|}{24} \frac{[(j+4)(j+3)(j+2)(j+1)]^{1/2}}{P^2} \left(\sum_{m_k=0} A_k^4 + 0((\Delta T)^2) \right) . \tag{38}$$

We now consider the limiting behavior of these expressions as ΔT approaches zero. At this stage, the shape of the impulse response $L(\tau)$ becomes important. In the continuum limit (as seen from Eqs. 22 and 27),

$$P \simeq \int_0^\infty L(\tau)^2 \, d\tau \tag{39}$$

and

$$\sum_{k=1}^{\infty} A_k^4 = \Delta T \int_0^\infty L(\tau)^4 \, d\tau . \tag{40}$$

Thus, ρ_{j+2} and ρ_{j+4} both depend on measures of the uniformity of the impulse response $L(\tau)$. For nondiagonal terms ($m_k \le 1$) we have

$$\rho_{j+2}(j) \le \Delta T \frac{|\gamma|}{6} j[(j+2)(j+1)]^{1/2} \frac{\max\{L(\tau)^2\}}{\int L(\tau)^2 \, d\tau} \tag{41}$$

and

$$\rho_{j+4}(j) \le \Delta T \frac{|\gamma|}{24} [(j+4)(j+3)(j+2)(j+1)]^{1/2} \frac{\int L(\tau)^4 \, d\tau}{\left[\int L(\tau)^2 \, d\tau \right]^2} (1 + 0(\Delta T)) . \tag{42}$$

For on-diagonal terms, the asymptotic dependence is similar, but the combinatorial coefficients are different, as a result of Eq. 30.

The Sum of Sinusoids

For the sum of sinusoids, much of the corresponding calculation has been carried out in detail previously (14). Here we highlight the major steps to emphasize the close analogy with the discrete multilevel input. We consider only a sum of incommensurate sinusoids. Extensions of the result to commensurate sinusoids are discussed in Victor and Knight (14).

The input ensemble Ω is specified by a choice of Q frequencies $\alpha = (\alpha_1, \ldots, \alpha_Q)$ and their corresponding amplitudes $a = (a_1, \ldots, a_Q)$. A test signal $S_\psi(t)$ is parameterized by a set $\psi = (\psi_1, \ldots, \psi_Q)$ of relative phases:

$$s_\psi(t) = \sum_{k=1}^{Q} a_k \cos(\alpha_k t + \psi_k) \quad . \tag{43}$$

We are interested in asymptotic behavior as Q increases. We assume that the transducer under study has a limited bandwidth, with no response above the frequency α_{max}. The input sinusoids are distributed approximately evenly from 0 to α_{max}, so $\alpha_k = k\alpha_{max}/Q$. The amplitudes of the sinusoids are chosen to ensure that the input power remains fixed at 1, that is, $a_k = (2/Q)^{1/2}$. (These assumptions guarantee that as the number of sinusoids increase, the power spectrum of the input signal approaches that of a band-limited but otherwise white Gaussian noise in an integral sense. Our result extends to other schemes of increasing the number of sinusoids, provided that the input power spectrum approaches a stable limit. In this case, the resulting kernels will converge to the Wiener kernels obtained with a Gaussian input with the corresponding colored spectrum.)

The response of a transducer μ to the sum-of-sinusoids signal in Eq. 43 may contain Fourier components at any frequency equal to an integer combination of the stimulus frequencies α. Such frequencies are called lattice frequencies, and they correspond to integer vectors **m**. Note that **m** may contain negative as well as positive numbers – for example, **m** = (2,0,−1,0, . . .) corresponds to the frequency $2\alpha_1 - \alpha_3$. With appropriate analytical fine print (14), a transducer μ may be represented by its Fourier resolution

$$\tilde{\mu}(\mathbf{m}) = \langle [(\mu)s_\psi](0) e^{-i\Sigma m_k \psi_k} \rangle_\Omega \quad . \tag{44}$$

This formula exploits the fact that for incommensurate sinusoids, an average over time may be replaced by an average over relative phases.

The orthogonal subspace K_j is readily described in terms of the Fourier resolution in Eq. 44. It consists of transducers whose Fourier resolution is nonzero only at lattice frequencies **m** for which $\sum |m_k| = j$. Consequently, the projection ϕ_j of a given transducer μ is a new transducer $\phi_j(\mu)$ whose Fourier resolution matches μ on jth order lattice points, and is zero elsewhere.

We have previously shown that in the limit that the number of sinusoids Q is large, the Fourier resolution on jth order lattice points corresponds to the Fourier transform of the jth order Wiener kernel (14). To examine the asymptotic approach of the Fourier resolution to the Fourier transform of the Wiener kernel, we calculate Fourier resolutions of the transducers LH_n.

Since the initial stage L of LH_n is a linear filter, its response to a sum-of-sinusoids input is also a sum of sinusoids. In analogy with Eq. 27, the power presented to the nonlinearity H_n is thus given by

$$\begin{aligned} P &= \langle [(L)s_\psi](0)^2 \rangle_\Omega \\ &= \sum_{1}^{Q} |A_k|^2 \quad , \end{aligned} \tag{45}$$

where

$$A_k = \frac{1}{Q^{1/2}} \tilde{L}(\alpha_k) \tag{46}$$

and $\tilde{L}(\alpha) = \int L(\tau)e^{-i\alpha\tau}\,d\tau$, the Fourier transform of the impulse response $L(\tau)$.

To exploit the generating function of the Hermite polynomials, we construct a formal transducer

$$G = \sum_0^\infty \frac{z^n}{n!} (LH_n) \ . \tag{47}$$

The Fourier resolution $\tilde{G}$ at the lattice point m is given by

$$\begin{aligned}\tilde{G}(\mathrm{m}) &= \left\langle \prod_{k=1}^{Q} e^{za_k Re[\tilde{L}(\alpha_k)e^{i\psi_k}]} e^{-im_k\psi_k} \right\rangle e^{-Pz^2/2} \\ &= \prod_{k=1}^{Q} e^{im_k \arg(\tilde{L}(\alpha_k))} F_{\mathrm{sos}}(m_k, |A_k|z) \ ,\end{aligned} \tag{48}$$

where

$$F_{\mathrm{sos}}(m,z) = \frac{1}{2\pi}\int_0^{2\pi} \cos m\psi\, e^{z\sqrt{2}\cos\psi - z^2/2}\, d\psi \ . \tag{49}$$

In analogy with the time-domain quantity $F(m,z)$ (Eq. 30), the quantities $F_{\mathrm{sos}}(m,z)$ summarize how a Wiener kernel of order n projects onto the orthogonal subspaces for the sum of sinusoids signal. To make the correspondence explicit, we transform Eq. 49 by $s = \sqrt{2}\cos\psi$. As ψ uniformly samples the phase circle, s is distributed according to the weight

$$w_{\mathrm{sos}}(s) = \begin{cases} \frac{1}{\pi}(2 - s^2)^{-1/2}, & -\sqrt{2} < s < \sqrt{2}, \\ 0, & \text{otherwise} \ . \end{cases} \tag{50}$$

The quantities $\cos m\psi$ may be re-expressed as polynomials $B_{\mathrm{sos},m}(s)$ in $s = \sqrt{2}\cos\psi$. Other than normalization factors, the polynomials $B_{\mathrm{sos},m}(s)$ are the Chebyshev polynomials (1). These polynomials are orthogonal with respect to the weight $w_{\mathrm{sos}}(s)$ by virtue of the orthogonality properties of $\cos m\psi$. Thus, $F_{\mathrm{sos}}(m,z)$ assumes the same form as $F(m,z)$ (Eq. 30):

$$F_{\mathrm{sos}}(m,z) = \int B_{m,\mathrm{sos}}(s) e^{zs - z^2/2} w_{\mathrm{sos}}(s)\, ds \ . \tag{51}$$

The rest of the analysis for the sum of sinusoids proceeds in parallel with the preceding analysis for multilevel signals. We focus on the lattice frequencies m with

$|m_k| \leq 1$; these are the analogs of the nondiagonal terms of the time-domain expansion. The size of the projection of a transducer from the nth Wiener subspace to the jth orthogonal subspace in the sum-of-sinusoids expansion is typified by the ratio of Fourier resolutions of normalized transducers. For $n = j + 2$ and $n = j + 4$ we find

$$\rho_{j+2}(j) \leq \frac{1}{Q}\frac{|\gamma|}{6} j[(j+2)(j+1)]^{1/2} \frac{\alpha_{max}\max\{|\tilde{L}(\alpha)|^2\}}{\int_0^\infty |\tilde{L}(\alpha)|^2 \, d\alpha} \tag{52}$$

and

$$\rho_{j+4}(j) \leq \frac{1}{Q}\frac{|\gamma|}{24} [(j+4)(j+3)(j+2)(j+1)]^{1/2} \frac{\alpha_{max}\int_0^\infty |\tilde{L}(\alpha)|^4 \, d\alpha}{\left[\int_0^\infty |\tilde{L}(\alpha)|^2 \, d\alpha\right]^2} \times \left(1 + 0\left(\frac{1}{Q}\right)\right). \tag{53}$$

In these expressions, γ denotes the kurtosis of the distribution $w_{sos}(s)$ (Eq. 50), which is equal to $-3/2$.

DISCUSSION

Summary of Results

We have analyzed how orthogonal functional expansions based on two classes of input ensembles differ from the standard Wiener kernels. As a measure of the disparity, we examined the contribution of transducers LH_n (whose standard Wiener expansion contained only an nth order term) to the jth order kernel in alternative orthogonal expansions.

The results for multilevel discrete inputs and for a sum of sinusoids are quite parallel. The largest contribution is from the transducer LH_j. This term corresponds directly to the standard Wiener kernel (for white but non-Gaussian discrete noises) or to its Fourier transform (for the sum of sinusoids). However, contributions from transducers $LH_{j+2}, LH_{j+4}, \ldots$ are also present. All of these contaminating terms approach zero as the sampling rate of the discrete noise, $1/\Delta T$, or the number of sinusoids, Q, increase. The contributions which decrease most slowly (in an asymptotic sense) are those from LH_{j+2} and LH_{j+4}; other contributions decrease as $0((\Delta T)^2)$ or $0(Q^{-2})$.

For LH_{j+2}, the relative asymptotic size of the contribution (Eqs. 41 and 52) depends on two dimensionless terms. The first term is an intrinsic property of the noise: $|\gamma|/6$, where γ represents the kurtosis of the multilevel discrete noise (for time-domain analysis) or the kurtosis of the cross-sectional distribution of a sinusoid (for frequency-domain analysis). The second term is a measure of how many independent components of the input signal are effectively "seen" by the nonlinear element. In the time domain, this is the ratio of the maximum of the square of the impulse response to the integral of the square of the impulse response, normalized by the time discretization. In the frequency domain, this term is an analogous quantity derived from the transfer function. For LH_{j+4}, two similar quantities again combine to determine the relative asymptotic size of the contribution to the jth order kernel (Eqs. 42 and 53). The kurtosis of the noise again enters, as $|\gamma|/24$. The second term is again a measure of the number of independent components of the signal. In both time-domain and frequency-domain techniques, the rapidity of convergence depends on j, the order of the kernel. This dependence is approximately inversely proportional to j^2.

Although we have considered only the linear-nonlinear transducers LH_n, the results we obtained have practical importance for more general transducers. This is because for each order n, the transducers LH_n form a basis of the nth orthogonal subspace of the standard Wiener expansion. However, it is important to recognize that the convergence of nonstandard Wiener kernels to Wiener kernels is *not* uniform in n, even for a fixed choice of the initial filter L. In other words, kernels measured with a particular ensemble Ω may be good approximations to the corresponding Wiener kernels of low-order but not high-order systems.

The present result extends the idea that the sum-of-sinusoids technique can be considered as a kind of Fourier transform of time-domain methods of nonlinear analysis. Convergence to the standard Wiener kernels is similar in rate and character. For the simple linear-nonlinear transducers LH_n, time-domain approaches provide close approximations to the standard Wiener kernels when L has a long integration time. Analogously, frequency-domain approaches provide close approximations to the standard Wiener kernels when L has a broad passband.

Klein and Yasui (5) and Klein (4) considered the relationship of orthogonal expansions with respect to non-Gaussian white noises to the Wiener kernels in detail. Equation 26 may be viewed as a compact formulation of their expressions which relate orthogonal expansions with respect to white noise of arbitrary cross-sectional distribution to the Wiener kernels. By analyzing the asymptotics of this expansion, we have shown how representations obtained with nonwhite cross-sectional distributions converge to the Wiener kernels. This forms a theoretical basis for the use of binary techniques such as Sutter's (12).

One consequence of the present result is that a particularly close approximation to the standard Wiener kernels is to be expected for input signals whose kurtosis is zero. This cannot be achieved for binary input signals (whose kurtosis is -2), but can be achieved for a ternary input signal, provided that the three input levels are presented with probabilities in the ratio 1:4:1. The more rapid convergence of the unequally weighted ternary signal is achieved at the expense of a reduction in power in the test signal, and may therefore not necessarily result in a net advantage in all applications.

Diagonal Values

We have shown that typical (off-diagonal) kernel values indeed approach estimates of Wiener kernels in a manner which depends in a stereotyped way on the kurtosis of the underlying distribution $w(x)$. On the diagonal, the asymptotic approach to Wiener kernels includes a numerical factor which depends on the multiplicity of equalities among the kernel arguments. Our analysis applies both to distributions $w(x)$ concentrated on a finite number of points (e.g., binary or ternary input) and to distributions $w(x)$ that are continuous. There is an important qualitative difference between these two cases for the diagonal values of the kernels. For distributions concentrated on N points, the corresponding set of orthogonal polynomials cannot be continued after the polynomial of order $N-1$. Thus, for such distributions, kernel values on diagonals of order N or higher (i.e., those with any time lag repeated N or more times) are undefined.

Whether such pathological behavior is of practical importance depends on the transducer under study and the test signal used. If the transducer has a front end with limited bandwidth, then a short enough time discretization will allow measurement of kernel values sufficiently near the diagonal to characterize the transducer. But, if the initial transduction is an instantaneous nonlinearity (at the time-scale of the time discretization), then the diagonal values are not well approximated by near-diagonal values, and important information will be lost unless the number of values in the test input exceeds the order of the nonlinearity.

Other Considerations in Choice of Input Signal

In comparing several test ensembles, we have emphasized the criterion of how well the measured kernels approximate standard Wiener kernels. This is important for applications in which one hopes to learn about the internal structure of the transducer from the qualitative features of the kernels. It is clearly less important in applications in which the kernels are used simply as a summary of responses of the transducer to a particular set of signals.

In any laboratory application, other criteria also must be considered. For example, with truly stochastic test ensembles, there will always be deviations of the actual statistical properties of the test input from those of the ensemble from which it is drawn. If there is an excess of time available for testing and intrinsic transducer noise dominates, this is a minor consideration. However, when intrinsic transducer noise is small or time available for testing is limited, the "noise in the noise" may be a dominant factor. Under such circumstances, one may eliminate the "noise in the noise" either by 1. exact orthogonalization with respect to the particular test signal drawn at random from a stochastic ensemble (7), or 2. employing a balanced stimulus, in which the key statistical properties of the ensemble are achieved exactly or nearly so. The M-sequence technique (12) and the sum-of-sinusoids technique (14) are examples of the second approach. In general, the computational labor involved in exact orthogonalization schemes is greater than that involved in the M-sequence technique (which requires a single rapid deconvolution) and the sum-of-sinusoids (which requires a single fast Fourier transform).

REFERENCES

1. Abramowitz, M.; Stegun, I.A. eds. Handbook of mathematical functions. New York: Dover; 1964.
2. Emerson, R.C.; Citron, M.; Vaughn, W.J.; Klein, S. Nonlinear directionally selective subunits in complex cells of cat striate cortex. J. Neurophysiol. 58:33–65; 1987.
3. Feller, W. An introduction to probability theory and statistics. 3rd ed. New York: J. Wiley & Sons; 1968.
4. Klein, S. Relationships between kernels measured with different stimuli. In: Marmarelis, V.Z., ed. Advanced methods of physiological system modeling. Vol. 1. Los Angeles, CA: University of Southern California; 1987.
5. Klein, S.; Yasui, S. Nonlinear systems analysis with non-Gaussian white stimuli: General basis functions and kernels. IEEE Trans. Inf. Theory. Vol. IT-25; 495–500; 1979.

6. Korenberg, M.J. Cross-correlation analysis of neural cascades. Proc. 10th Ann. Rocky Mountain Bioeng. Symp. 47–51; 1973.
7. Korenberg, M.J. Exact orthogonal kernel estimation from finite data records: Extending Wiener's identification of nonlinear systems. Ann. Biomed. Eng. 16:123–142; 1988.
8. Lee, Y.N.; Schetzen, M. Measurement of the kernels of a nonlinear system by cross-correlation. Int. J. Cont. 2:237–254; 1965.
9. Marmarelis, V.Z. Identification of nonlinear systems through multi-level random signals. In: Marmarelis, V.Z., ed. Proc. of 1st Symp. Testing and identification of nonlinear systems. Pasadena, CA: California Institute of Technology; 1975.
10. Marmarelis, P.Z.; Marmarelis, V.Z. Analysis of physiological systems: The white-noise approach. New York: Plenum; 1978.
11. Sakai, H.; Naka, K.-I.; Korenberg, M.J. White noise analysis in visual neuroscience. Visual Neuroscience 1:287–296; 1988.
12. Sutter, E. A practical nonstochastic approach to nonlinear time-domain analysis. In: Marmarelis, V.Z., ed. Advanced methods of physiological system modeling. Los Angeles: Biomedical Simulations Resource; 1987.
13. Victor, J.D.; Johanessma, P.I.M. Maximum-entropy approximations of stochastic nonlinear transductions: An extension of the Wiener theory. Biological Cybernetics. 54:289–300; 1986.
14. Victor, J.D.; Knight, B.W. Nonlinear analysis with an arbitrary stimulus ensemble. Q. Appl. Math. 37:113–136; 1979.
15. Volterra, V. Theory of functionals and of integral and integro-differential equations. London, Glasgow: Blackie; 1932.
16. Wiener, N. Nonlinear problems in random theory. New York: Wiley; 1958.