Sparse-stimulation and Wiener Kernels

Stanley A. Klein
School of Optometry, UC Berkeley
Berkeley, CA 94720
klein@garnet.berkeley.edu

Abstract

Wiener and Volterra kernel analysis is useful for analyzing nonlinear neural systems. In order to obtain kernels with good signal to noise properties one must optimize the stimulus characteristics. This optimization has been hampered by the lack of simple expressions for kernel variance. We not only have found simple expressions for the kernel variance but also propose a new technique that we call "sparse-stimulation" that greatly enhances the signal-to-noise ratio of higher order kernel estimates. The technique involves using a slower sampling rate for the stimulus than for the response.

Wiener kernels and white noise stimuli provide a powerful approach for analyzing nonlinear input-output characteristics of neurons (Marmarelis & Naka, 1972; Emerson et al., 1985; Marmarelis & Marmarelis, 1978) and other biological systems (Stark, 1968). Although there have been many interesting and important advances using this approach, it hasn't yet received wide acceptance. Part of the reason is that this approach is mathematically forbidding to most researchers, but I believe a more important reason is that the white-noise machinery has not yet been suitably packaged to provide the biologist what he or she needs.

Five recent advances are enabling a change in this situation:
1) Erich Sutter's development of the Fast M Transform (Sutter, 1990), that
together with complete M-sequences enables one to calculate hundreds of
kernels (higher-order and/or multiple input) in the same time that only a single
kernel could be calculated with previous methods.
2) General formulas have been developed that show the interrelationships
among kernels obtained using different stimuli (Klein & Yasui, 1979; Klein
1989; Victor, 1990). These formulas connect the Wiener kernels obtained with
Gaussian stimuli to the more efficient kernels from binary, ternary and flash
(Poisson) stimuli. In addition they connect kernels obtained at different
luminances and power densities (the connection between Volterra and Wiener
kernels are special cases). These latter relationships show how the second and
third order kernels specify the system's luminance and contrast gain control
respectively.
3) A new formalism has been developed for ternary stimuli (Emerson et al,
1987) that allows all the important information up through fourth order to be
displayed as second order kernels. This circumvents the problem of how to
display information above second order. There are many advantages of ternary
kernels for systems based on separate on and off channels such as is found in
the nervous system.
4) Simplified formulas for kernel variance have been derived that are based on
using complete M-sequences as the stimuli. The use of these stimuli
eliminates the variance due to the nonorthogonality of the stimuli.
5) By using sparse-stimulation one can dramatically improve the signal-to-
noise ratio, especially for higher order kernels.

My poster at the Neural Analysis and Modeling meeting was concerned
with items 4 and 5, which is material that has not been published previously.
The present paper provides further information on both the new formula for
kernel variance and on sparse-stimulation. Before getting to this new material,
however, it may be useful to present a quick review of the standard white-noise
formalism.

Wiener expansion for white noise stimulation. A white stimulus is
one whose second order correlation function is a delta function. That is:
\[ E[x(t) x(t-\tau)] = x^2 \text{ for } \tau = 0 \] (1)
and \[ E[x(t) x(t-\tau)] = \frac{x^2}{\tau} \text{ for } \tau \neq 0 \]
The expectation value in Eq. 1 is defined by:
\[ E[f(x(t))] = f(t) \]
\[ = \frac{1}{T} \sum_{t=0}^{T} f(x(t)) \Delta \] (2)
where \(\Delta\) is the time between samples. The present analysis is based on the
time indices \(t\) and \(\tau\) being discrete rather than continuous variables. This
assumption does not severely limit the analysis since most stimuli and responses are sampled at discrete intervals.

It is also assumed that stimuli at different times are independent, so that the expectation value of a product of stimuli with different time delays equals the product of the expectation values:

$$E\left( \prod_{i=1}^{n} x(t - \tau_i) \right) = \overline{x}^n$$  \hspace{1cm} (3)

Before getting to the mathematical Wiener expansion it is useful to put it into words. The idea is to express the response $y(t)$ as a polynomial expansion of the stimulus, similar to a Taylor’s expansion for a general function. The zeroth order kernel, $h_0$, is the average value of the response. The first order kernel is the linear impulse response. For a white noise stimulus, it can be obtained by time-locking the response to the stimulus. Eq. 8 will show that an alternative, but equivalent, definition of the first order kernel is given by the cross-correlation between the stimulus and the response. The second order kernel is obtained by time-locking the response to pairs of stimuli after having subtracted off the linear response.

In mathematical language, the Wiener orthogonal expansion is given by:

$$y(t) = h_0 + \sum_{t' < t} h_1(t, t') X_1(t') \Delta + \sum_{t' < t, \tau_1 > 0} h_2(t, t', \tau_1) X_1(t') X_1(t - \tau_1) \Delta^2 + \sum_{t' < t, \tau_1 > 0} h_2(t, t', 0) X_2(t') \Delta^2 + \sum_{t' < t, \tau_1 > 0} h_2(t, t', \tau_2) X_1(t') X_1(t - \tau_2) \Delta^3 + \sum_{t' > t, \tau_1 > 0} h_3(t, t', \tau_1, \tau_2) X_2(t') X_1(t - \tau_1) \Delta^3 + \sum_{t' < t, \tau_1 > 0} h_3(t, t', \tau_1) X_1(t') X_2(t - \tau_1) \Delta^3 + \left( \sum_{t' < t} h_3(t, t', 0, 0) X_3(t') \Delta^3 \right) \text{ plus fourth and higher order terms.} \hspace{1cm} (4)$$

where the orthogonal polynomials are given by ratios of simple determinants (Klein, 1979, 1987). The first two polynomials are:

$$X_1(t) = \begin{vmatrix} \overline{x} & 1 \\ 1 & x(t') \end{vmatrix} \begin{vmatrix} 1 \end{vmatrix}$$ \hspace{1cm} (5)

$$X_2(t) = \begin{vmatrix} \overline{x} & \overline{x^2} & \overline{x^3} \\ \overline{x^2} & \overline{x^3} & x(t') x^2(t') \end{vmatrix} \begin{vmatrix} 1 \frac{\overline{x}}{\overline{x^2}} \end{vmatrix}$$ \hspace{1cm} (6)
where a matrix between vertical lines means the determinant of the matrix. The higher order polynomials are constructed in a similar manner. The constraints on the time indices in Eq. 4 are worth noting:

1. The first argument of the kernel \((t-t')\) represents the time after the last stimulus presentation. The time from the last stimulus, \(t'\), to the response, \(t\), must be a positive number.
2. The later arguments, \(\tau_1\), of the second and higher order kernels represent the time between a prior stimulus (at time \(t'-\tau_1\)) and the last stimulus (at time \(t'\)).
3. A time ordered convention has been adopted whereby \(\tau_1 > 0\) and \(\tau_1 > \tau_{i-1}\). This convention avoids double counting in the Wiener expansion of Eq. 4.

The purpose for introducing the orthogonal polynomials, Eq. 5 and 6, is to be able to calculate the kernels by cross-correlation:

\[
h_0 = E\{ y(t) X_0(t') \} / E\{ X_0^2(t) \} \quad \text{where} \quad X_0 = 1
\]

\[
h_1(t-t') = E\{ y(t) X_1(t') \} / E\{ X_1^2(t) \}
\]

\[
h_2(t-t',\tau_1) = E\{ y(t) X_1(t')X_1(t'-\tau_1) \} / E\{ X_1^2(t) \}^2
\]

\[
h_2(t-t',0) = E\{ y(t) X_2(t') \} / E\{ X_2^2(t) \}
\]

\[
h_3(t-t',\tau_1,\tau_2) = E\{ y(t) X_1(t')X_1(t'-\tau_1)X_1(t'-\tau_2) \} / E\{ X_1^2(t) \}^3
\]

\[
h_3(t-t',\tau_1,\tau_1) = E\{ y(t) X_1(t')X_2(t'-\tau_1) \} / E\{ X_1^2(t) \} E\{ X_2^2(t) \}
\]

\[
h_3(t-t',0,\tau_1) = E\{ y(t) X_2(t')X_1(t'-\tau_1) \} / E\{ X_1^2(t) \} E\{ X_2^2(t) \}
\]

where the expectation value is defined in Eq. 1. In the conventional formalism there is a factor of \(1/2!\) in Eq. 9 and a factor of \(1/3!\) in Eq. 11. These factors do not appear in the present formalism because the time-ordering discussed above avoids the double-counting that would otherwise have led to these factorial factors. An alternative method for constructing a fully general set of basis functions that are also properly normalized was presented elsewhere (Klein, 1979, 1987).

A simplified formula for kernel variance. One of the main obstacles limiting immediate acceptance of the white noise approach is that often the signal-to-noise ratio of the kernels is poor. I maintain that an important reason that one often obtains noisy kernels is that the contributions to the kernel variance are poorly understood by most users of the white-noise approach. Here we summarize some of our findings on a simplified expressions for the kernel variance:

\[
\text{var}(h_n) = \bar{n}^2 / [T \text{P}(P_d)^n] = \bar{n}^2 / [R (P_d)^n]
\]

\(h_n\) is the nth order kernel (similar to nth Taylor coefficient), \(R = T/d_s\) is
number of stimulus samples, \( T \) is the run duration, \( \overline{n^2} \) is the power of the noise, \( P \) is the power density of the stimulus, and \( d_s \) is the interval between stimulus samples.

The critical quantity to understand in order to appreciate the size of the kernel variance is \( P \), the power density, which is the average value of \( x^2 \) times the sampling time. For systems that adapt, the system responds to contrast rather than to absolute stimulus units, so the stimulus \( x \), should be measured in \% contrast. Furthermore if time is measured in units of the characteristic adaptation time of the system, then the value of the power density becomes meaningful because it is a unitless quantity. If \( P \) is much less than 1, very little power is being put into kernels above the first order and one should worry that the 2nd and higher order kernels will have an unacceptable signal-to-noise ratio.

It is useful to illustrate the power density for flashes on a dark background:

\[
P = \frac{d_s}{p} = D \quad (x \text{ is } \% \text{ contrast})
\]

(15)

where \( p \) is flash probability and \( D \) is average interflash time, so the kernel variance (Eq. 14) becomes:

\[
\text{var}(h_n) = \frac{\overline{n^2}}{R \cdot (Dd_s)^2}
\]

(16)

Formulas 14 and 16 for variance are much simpler than formulas given in the past (see Marmarelis and Marmarelis, 1978). Previous formulas had a much more complicated numerator due to the non-orthogonality of the stimulus for finite run lengths. Our formula is simple, with simply \( \overline{n^2} \) in the numerator because we assume that m-sequence noise was used with very clean autocorrelation properties. It should be pointed out that the m-sequence noise can lead to localized anomalies that must be avoided by clever cancellation techniques (not yet fully worked out—contact Erich Sutter, at Smith-Kettlewell, or Stan Klein for details).

The kernel variance isn't of great interest because it is typically measured in obscure units similar to the kernels (the rational units described above are never used in the usual Wiener approach i). A more appropriate quantity is the unitless \%error:

\[
\frac{\text{var}(h_n)}{h_n^2} = \frac{1}{R \cdot (d_s)} \overline{n^2} \frac{n}{G_n^2}
\]

(17)

where \( M \) is the system memory, \( \overline{h_n^2} \) is the average kernel, and \( G_n \) is the contribution to the response of the \( n \)th order kernels (\( G_n^2 \) is proportional to \( \overline{P^2} \)). For example, if \( M/d_s = 5, R = 104, \overline{n^2}/G_n^2 = 4 \), then for \( n = 2 \)nd order:

\%error of the kernel is \%error = \( 10^{-2} \cdot 5 \cdot 2 = 10\% \).

From Eqs. 14, 16 or 17 it is clear that in order to reduce the kernel variance
it would be advantageous to increase the stimulus sampling time $d_s$. The introduction of sparse-stimulation allows this without losing temporal resolution in the main time variable $(t - t')$.

**Sparse-stimulation.** Sparse-stimulation is the name given to the situation when the stimulus is sampled less often than the response. The same concept had been previously called superresolution (Klein, 1979) since the response is sampled more frequently than the white-noise stimulus. The use of sparse-stimulation overcomes the main disadvantage of white noise stimuli: namely the poor signal-to-noise (S/N) ratio of higher kernels. We believe that the second and third order kernels are of critical importance for understanding the dynamics of nonlinear systems and their estimation must be improved. This belief is the driving force behind our development of sparse-stimulation.

The following example is provided to show how sparse-stimulation can improve the S/N of kernel estimates. Suppose, for example, that the memory of the system is $N$ then there are $N$, $N^2$, and $N^3$ first, second and third order kernel elements (we assume the memory is non-zero for $t$ and $\Delta t \leq N$). For $N = 100$ this amounts to 1,010,100 elements. If the initial S/N is 1 and it is desired that the kernels should have $S/N = 0.5$ then the number of response samples should be about $25 \times 1,010,100$. At a rate of 2 msec/sample (as would be appropriate for the VEP and ERG) it would require 14 hours to obtain sufficient data. This is an unreasonable length of time.

Suppose, instead, that while the response is still sampled every 2 msec, the stimulus is sampled every 20 msec. Thus, although the memory of the system is still $N_f = 100$ samples (long 200 msec), the time between flashes has a memory of only $N_s = 10$ samples. There are $N_f N_s$, $N_f N_s N_s$, and $N_f N_s N_s^2$ elements for the first, second and third order kernels producing a total of 11,100 kernel elements. To achieve a 25-fold improvement in the S/N ratio would require only 9 min. This testing time is a factor of almost 100 less than the white-noise case.

The problem with sparse-stimulation is that at the response sampling time scale, the stimulus no longer has clean autocorrelation properties. Consider, first, an asymmetric binary stimulus that has a 50% chance of flashing every 20 msec. If the sampling is every 2 msec then the first order autocorrelation function would be unity at $\Delta t = 0$, it would be 0.5 whenever $\Delta t$ is a multiple of 20 msec, and it would be zero elsewhere. The comb function spikes every 20 msec could cause trouble for naive applications of the Lee-Schetzen (1965) algorithm. This problem with the first order autocorrelation can be eliminated by using symmetric stimuli (ternary or Gaussian) but then a similar problem occurs for higher order correlations such as $E(x^2(t)x^2(t+\Delta t))$ (Klein, 1979). Luckily, it turns out that by using a straightforward algorithm it is possible to avoid these problems associated with the non-whiteness of sparse-stimulation.

The only change that is needed in Eq. 4 is to put a time index on $h_0$, so it becomes $h_0(t)$. The zeroth order kernel $h_0(t)$ is obtained from Eq. 7 where $X_0(t)$ is the comb function which is unity at multiples of 20 msec and is zero
otherwise. Thus $h_0$ is a repetitive function of time, $t$, with a cycle time given by the stimulus sampling time. All the other kernels are obtained as before except one must be careful that in taking expectation values such as in Eqs. 5 and 6 one must only sample at the stimulus sampling times, not the response sampling times. In Eqs. 4 through 13 we have tried to be careful with our indices: the time index, $t$, of the response is sampled at the fast 2 msec rate, the time, $t'$, associated with the stimulus is sampled at the slower rate, as are all the delay indices $t$. With these provisos, it should be possible to implement the sparse-stimulation formalism which allows the first time index of each kernel to be sampled finely and the other indices to be sampled sparsely.

This method for obtaining kernels has the advantage that the signal-to-noise is greatly improved, especially for higher order kernels. The alternative method for improving the signal-to-noise ratio is to pool data from neighboring bins, but this method degrades the temporal resolution of the kernels.

I would be eager to discuss these issues with anyone interested. My E-mail address is klein@garnet.berkeley.edu, and my phone number is 415-643-8670.

Acknowledgement—Supported by grant AFOSR-89-0238 from the Air Force Office of Scientific Research.

References


