Use of Meixner Functions in Estimation of Volterra Kernels of Nonlinear Systems With Delay

Musa H. Asyali*, *Member, IEEE,* and Mikko Juusola

Abstract—**Volterra series representation of nonlinear systems is a mathematical analysis tool that has been successfully applied in many areas of biological sciences, especially in the area of modeling of hemodynamic response. In this study, we explored the possibility of using discrete time Meixner basis functions (MBFs) in estimating Volterra kernels of nonlinear systems. The problem of estimation of Volterra kernels can be formulated as a multiple regression problem and solved using least squares estimation. By expanding system kernels with some suitable basis functions, it is possible to reduce the number of parameters to be estimated and obtain better kernel estimates. Thus far, Laguerre basis functions have been widely used in this framework. However, research in signal processing indicates that when the kernels have a slow initial onset or delay, Meixner functions, which can be made to have a slow start, are more suitable in terms of providing a more accurate approximation to the kernels. We, therefore, compared the performance of Meixner functions, in kernel estimation, to that of Laguerre functions in some test cases that we constructed and in a real experimental case where we studied photoreceptor responses of photoreceptor cells of adult fruitflies (Drosophila melanogaster). Our results indicate that when there is a slow initial onset or delay, MBF expansion provides better kernel estimates.**

Index Terms—**Laguerre functions, nonlinear system identification Meixner functions, Volterra series.**

I. INTRODUCTION

S IGNIFICANT applications of nonlinear system identification techniques to biological sciences have occurred for nearly a century, starting from the early work of Volterra [1]. In Volterra's approach, a nonlinear time-invariant dynamic system is represented by a series of multidimensional functions known as "Volterra Series" that represent the system's transformation action. Application areas of this mathematical tool in biomedical engineering as a method of system identification include modeling of respiratory response [2], [3], renal auto-regulation [4], and neural plasticity [5], [6].

The typical problem of system identification involves estimation of system functions or kernels from a pair of possibly noisy input and output data. When the system under study is not linear, which is generally the case for physiological systems, we can use Volterra's approach.

Volterra has shown that a nonlinear time-invariant system's output or response $y(t)$ to an input or stimulus $x(t)$ can be expressed by the following multiconvolution relation [7], [8]

$$y(t) = C_{dc} + \int_0^\infty x(t - \tau_1)K_1(\tau_1)d\tau_1$$
$$+ \int_0^\infty \int_0^\infty x(t - \tau_1)x(t - \tau_2)K_2(\tau_1, \tau_2)d\tau_1 d\tau_2$$
$$+ \cdots$$
$$= C_{dc} + \sum_{i=1}^\infty \int_0^\infty \cdots \int_0^\infty x(t - \tau_1)$$
$$\ldots x(t - \tau_i)K_i(\tau_1, \ldots, \tau_i)d\tau_1 \ldots d\tau_i. \quad (1)$$

The series/sum given by (1) is known as the Volterra series. Here, $K_i(\tau_1, \ldots, \tau_i)$ denotes the system's ith order Volterra kernel, associated with the system's ith order nonlinearity. The C_{dc} is a constant term that balances the means of the two sides. For a linear system C_{dc} is 0 and the remainder of the right-hand side of (1) reduces to the well-known convolution integral, where $K_i(\tau_i)$ is called the impulse response. As the complexity or nonlinearity of the system increases, more kernels are needed to accurately describe the functioning of the system. According to this description of nonlinear systems, knowing its Volterra kernels is sufficient to find the system's response to any input. However, in practice, due to the problems like short data length and ill-conditioned matrices, it is only possible to estimate accurately kernels up to a certain order.

The problem of estimation of Volterra kernels can be formulated as a multiple regression problem and solved using least squares estimation. To elucidate this approach, let us assume that we are estimating the first two kernels of a nonlinear system. Even if the system under study may have nonlinearities higher than second-order, we can still find out how well we can approximate the system's behavior with a second-order Volterra model. Extension of this technique to the estimation of higher order kernels is straightforward.

As we will be using numerical techniques for kernel estimation, we need to translate (1) into discrete-time. We do this by assuming a default sampling period of 1 s and input and output data lengths of N points. If we denote input and output sequences by $x[n]$ and $y[n]$, where $n = 0, 1, \ldots, N - 1$, after discretization of (1) we obtain

$$y[n] = C_{dc} + \sum_{m=0}^{N_1 - 1} x[n - m]K_1[m]$$
$$+ \sum_{m_1=0}^{N_2 - 1} \sum_{m_2=0}^{N_2 - 1} x[n - m_1]x[n - m_2]K_2[m_1, m_2] + e[n] \quad (2)$$

where, K_1 and K_2 denote the first- and second-order kernels and N_1 and N_2 their lengths. Notice that, during the conversion to discrete-time, we included kernels up to second-order and assumed that the contribution of the higher order kernels are embedded in the error sequence $e[n]$.

To keep the number of points in the kernels at a reasonable level, one can select the kernel lengths N_1 and N_2 in such a way that $N \geq N_1 \geq N_2$, as the length of the kernels cannot be greater than the data length N and the number of points in the kernels increases exponentially with the order. For instance, according to formulation in (2), we have N_1 and $N_2(N_2+1)/2$ distinct values or points in K_1 and K_2, respectively.

The orthogonal series expansion method [9], [10] that we will use assumes that the kernels can be expressed in terms of some suitable orthogonal basis functions B_q, q being the order, as

$$K_1[m] \cong \sum_{q=0}^{Q_1-1} C_q B_q[m]$$
$$m = 0, 1, \ldots, N_1 - 1 \qquad (3.a)$$
$$K_2[m_1, m_2] \cong \sum_{q_1=0}^{Q_2-1} \sum_{q_2=0}^{Q_2-1} C_{q_1,q_2} B_{q_1}[m_1] B_{q_2}[m_2],$$
$$m_{1,2} = 0, 1, \ldots, N_2 - 1. \qquad (3.b)$$

Here, C_q and C_{q_1,q_2} are the coefficients or weights, and Q_1 and Q_2 are the number of basis functions used in the expansion of K_1 and K_2, respectively. By substituting (3.a) and (3.b) in (2) and expressing the convolution of $x[n]$ with B_q as $v_q[n] = \sum_{m=0}^{N-1} x[n-m]B_q[m]$, we can express $y[n]$ as

$$y[n] = \sum_{q=0}^{Q_1-1} C_q v_q[n] + \sum_{q_1=0}^{Q_2-1} \sum_{q_2=0}^{Q_2-1} C_{q_1,q_2} v_{q_1}[n] v_{q_2}[n] + e[n].$$
$$(4)$$

We should note here that the error terms in (4) and (2) are slightly different. The error term in (4) includes not only missing/ignored contribution of higher order kernels to the output but also the error introduced due the approximate kernel expansions (3.a) and (3.b) substituted in (2). By further defining column vectors corresponding to the output, convolution, and error sequences, respectively, as $\mathbf{y} = [y[0] \ldots [N-1]]^T$, $\mathbf{v}_q = [v_q[0] \ldots v_q[N-1]]^T$, and $\mathbf{e} = [e[0] \ldots e[N-1]]^T$, we can put (4) into matrix form as

$$\mathbf{y} = \mathbf{V}\mathbf{C} + \mathbf{e} \qquad (5)$$

where, $\mathbf{V} = [[1 \ldots 1]^T | \mathbf{v}_0 \mathbf{v}_1 \ldots \mathbf{v}_{Q_1-1} | \mathbf{v}_{0,0} \mathbf{v}_{0,1} \ldots \mathbf{v}_{0,Q_2-1} \mathbf{v}_{1,1} \mathbf{v}_{1,2} \ldots \mathbf{v}_{1,Q_2-1} \ldots \mathbf{v}_{Q_2-1,Q_2-1}]$ is the $N \times M$ observation matrix formed by using v_q's and their element-wise multiplicative combinations $\mathbf{v}_{q_1,q_2} = \mathbf{v}_{q_1}\mathbf{v}_{q_2}$ and $\mathbf{C} = [C_{dc} | C_0 C_1 \ldots C_{Q_1-1} | C_{0,0} 2C_{0,1} \ldots 2C_{0,Q_2-1} C_{1,1} 2C_{1,2} | \ldots 2C_{1,Q_2-1} \ldots C_{Q_1-1,Q_2-1}]^T$ is the $M \times 1$ vector of coefficients. The column of 1's in $\mathbf{V}$ allows for the estimation of the constant term C_{dc}. Since $\mathbf{v}_{q_1,q_2} = \mathbf{v}_{q_2,q_1}$, we collected similar terms in the expansion of K_2 in (4) and doubled the corresponding coefficient in vector $\mathbf{C}$, hence $M = 1 + Q_1 + Q_2(Q_2+1)/2$.

The over-determined system of equations given in (5) can be solved conveniently for C using the least squares technique. Inclusion of the constant term in the regression assures that the error sequence will have zero mean. Once the coefficients are estimated, they are substituted in the expansions (3.a) and (3.b) and the kernels are constructed. Extension of this technique to the estimation of higher order kernels is straightforward. The important issue here is to utilize basis functions that are morphologically similar to the kernels of the system under study. This enables accurate representation of kernels with a relatively small number of basis functions or weights, which implies a reduction in the number parameters to be estimated. Estimating a smaller number of parameters may improve the numerical condition of the estimation problem and produce coefficient estimates with less variance. The series expansion utilizing discrete orthogonal Laguerre Basis Functions (LBFs) has been used widely. The kernels of physiological systems typically die away after some certain time, known as the memory of the system, and many studies have shown that LBF can efficiently represent such kernels. However, when the kernels have sluggish initial onset, using LBF to represent them may not be suitable. A physiological example where the LBF expansion does not do as well as other approaches can be seen in [11].

In this paper, we suggest an alternative set of basis functions to represent kernels that have a slow initial onset. The research in signal processing has shown that the "Meixner-like basis functions having rational z-transform" is a better option to represent such functions (see [12] and the references therein). We will simply refer to these functions as Meixner basis functions (MBF) for brevity. As we will see shortly, LBF are closely related to MBF, actually LBF is a special case of MBF. However, until now, MBF have not been used in physiological modeling studies. In an attempt to highlight the cases where MBF can be a better alternative to LBF, we applied kernel estimation using MBF and LBF expansions, on some simulated and experimental data. Then we analyzed and compared the performance of these two families of basis functions in kernel estimation. In Section II, we will first give a brief background on LBF and MBF, and discuss the details of kernel estimation, e.g., the issues of selecting the number of basis functions to be used in the expansions and the underlying parameters of the basis functions. We will then describe the data from hypothetical systems and real experiments from which we estimated Volterra kernels using LBF and MBF expansions. In Section III, we will present graphs and tables summarizing our kernel estimation results. In Section IV, we will discuss possible advantages of using MBF over LBF.

II. METHODS

A. The Discrete-Time Laguerre and Meixner Basis Functions

Since the generation of discrete-time LBF and MBF using their explicit formulas is not practical, we generated them using the filter structure shown in Fig. 1 (adopted from [12] with permission). A MATLAB™ (The MathWorks, Inc. Natick, MA) code for generating MBF is available upon request. For explicit formulas for these orthogonal basis functions, we refer the reader to [12]. LBF and MBF are denoted by $L_q[k]$ and $M_q^{(n)}[k]$, respectively, where the subscript q is the order of basis functions and k is the index. We note that the first block is a low-pass filter and all the remaining cascaded blocks are all-pass filters, i.e., their magnitude response is unity. For both LBF and MBF, the
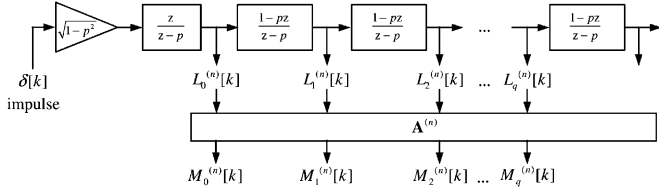
Fig. 1. The cascaded filter structure to generate LBF and MBF (Adopted from [12], courtesy of Prof. A. C. Den Brinker). $L_q[k]$ and $M_q[k]$, respectively, denote LBF and MBF; $q = 0, 1, 2, Q - 1$ is the order and $k = 0, 1, 2, N - 1$ is the index of basis functions; $\mathbf{A}^{(n)}$ is the $Q \times Q$ orthogonal matrix that transforms LBF to MBF.

pole parameter p $(0 < p < 1)$ determines how soon the basis functions will die away. As p increases, the functions become more oscillatory and prolonged. For each different value of p we have a different set of basis functions.

$\mathbf{A}^{(n)}$ is an orthogonal transformation matrix that operates on LBF to generate MBF (the Appendix gives details of the construction of this matrix). The parameter $n = 0, 1, 2 \ldots$ is the "order of the generalization" that determines how late the family of MBF will start to fluctuate. For each different value of n we have a different set of MBF. When $n = 0$, $\mathbf{A}^{(n)}$ becomes the identity matrix and hence MBF become identical to LBF.

Selection of p and n are of crucial importance for an accurate representation of kernels to be estimated. We demonstrate the effect of p and n on the family of basis functions in Fig. 2. We note that LBF and MBF become more oscillatory, i.e., die away more slowly, as the pole parameter p gets closer to 1 and that when the order of generalization n increases, the delay in the onset of the functions increases.

B. Simulation Study

1) Test Systems and Simulated Data: In order to compare the performance of LBF and MBF in kernel estimation, we constructed two hypothetical second-order nonlinear systems. The use of hypothetical systems is essential, as we need to know the exact kernels to compare kernel estimation approaches based on different basis function expansions. The first- and second-order kernels of the systems will be derived from the following template function that is comprised of two exponentials with time constants of 12.5 and 6.25 s

$$h(t) = 4(e^{-0.08t} - e^{-0.16t}).$$

We discretized this function with a sampling period of 1 s and over the interval $t = 0$ to 169 s, giving a total of 170 samples. Based on the morphology of kernels of physiological systems that have been identified in earlier studies [2]–[6], we can assume that kernels constructed using this function will be a good representative for kernels of real physiological systems.

The first test system has the following first- and second-order kernels, $h_1(\tau)$ and $h_2(\tau_1, \tau_2)$, respectively

$$h_1(\tau) = h(\tau), \quad h_2(\tau_1, \tau_2) = h(\tau_1)h(\tau_2)$$

The second test system has the same kernels except for a delay of 10 s along each dimension

$$h_2(\tau) = h(\tau - 10)$$
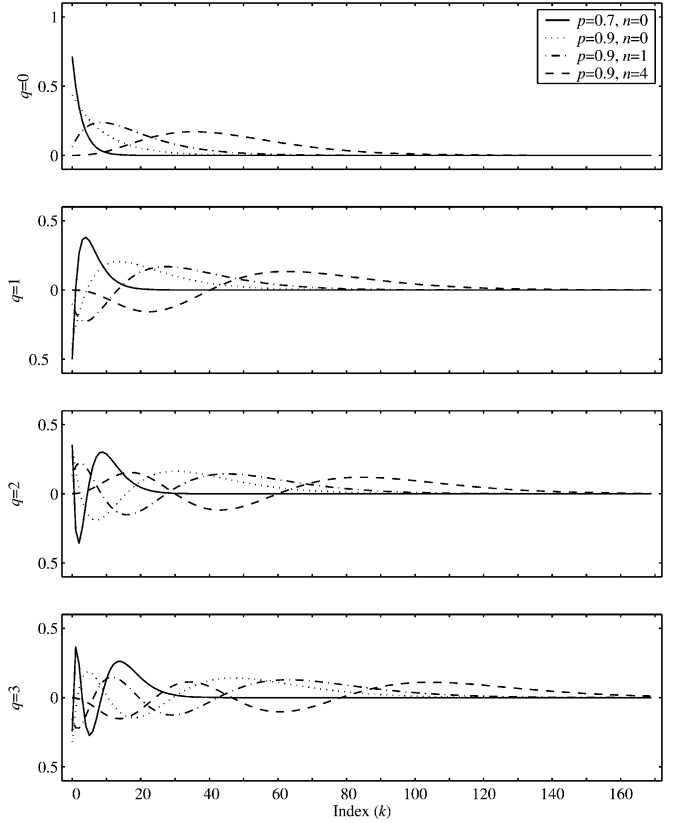$$h_2(\tau_1, \tau_2) = h(\tau_1 - 10)h(\tau_2 - 10).$$

Fig. 2. The discrete time MBF of order $q = 0, 1, 2, 3$ for different values of (p, n) as a function of index k. (Solid line: $p = 0.7$, $n = 0$. Dotted line: $p = 0.9$, $n = 0$. Dash-dot line: $p = 0.9$ and $n = 1$, Dashed line: $p = 0.9$ and $n = 4$).

We assumed that these test systems were driven by a white Gaussian noise (WGN) sequence of zero mean, unit variance and length 500 points. The WGN is an ideal choice for input in system identification studies, since it has an essentially flat power spectrum over a broad range of frequencies that allows stimulation a larger range of dynamic modes of the system under study. We also studied the effect of output noise in kernel estimation. To simulate measurement noise in the output, we added WGN at different power levels to the output as follows:

$$y_w = y + w\sigma_y \sqrt{\frac{P_w}{100}}$$

where y and σ_y refer to the output and its standard deviation respectively, P_w is the percentage of the noise power with respect to the output power (i.e., σ_y), w is again a zero mean, unit variance WGN and y_w is the noisy output. By stimulating the test systems with WGN inputs and collecting corresponding outputs, we obtained input and noise-free output ($P_w = 0$) and input and two noisy output ($P_w = 5, 10$) data pairs from which kernels of the system can be estimated using LBF and MBF expansions.

2) Kernel Estimation: In order to carry out kernel estimation, we need to decide about the number of basis functions to be used in (3.a) and (3.b). Q_1 and Q_2 can be selected depending on the desired level of accuracy for representation of the kernels. However, if too many basis functions are used, we may be attempting to explain even the inherent noise in the data. This

will diminish our model's ability to generalize and the numerical condition of the estimation problem will deteriorate, i.e., we may end up having coefficient estimates with large variances. On the contrary, if too few basis functions are used, we may have a poor model with a large estimation error. The number of coefficients to be estimated exponentially increases with the kernel order, therefore, selecting $Q_1 \geq Q_2$ is a good practice. Given M, this study aims at comparing the LBF and MBF in kernel estimation. Thus, after some trial and error, we decided to use $(Q_1 = 8, Q_2 = 4)$ and $(Q_1 = 12, Q_2 = 6)$, corresponding to two different resolution levels. (With these choices for the number of basis functions, we had, respectively, $M = 19$ and 34 model parameters or coefficients to estimate.)

However, the selection Q_1 and Q_2 (or equivalently M) is a nontrivial problem in general. This issue has been dealt with extensively in system identification literature and various criteria based on information theoretical results have been proposed. The most commonly used of such criteria are Akaike's information criterion (AIC), final prediction error (FPE) [13]–[15], and minimum descriptive length (MDL) [16], [17]. All of these criteria involve two terms acting in opposite sense, the variance of the estimation error that monotonically decreases as M increases and an expression that increases with M.

Besides Q_1 and Q_2, we also need to select the parameters for the basis functions. For both LBF and MBF, we need to select a suitable value for p that determines how soon the basis functions will die away. For MBF, we also need to select a suitable value for n that determines how late the functions will start to fluctuate. In order to determine the optimal basis function parameters, we let n vary from 0 to 20 and for each value of n made an efficient search on p to locate the value that minimizes the Euclidean norm of the estimation error, $\|\mathbf{e}\|$. Using Nelder–Mead simplex (direct search) method [18], [19] and a starting/initial p value of 0.5, we established the optimal p with a resolution of 10^{-4}. Since the error is zero mean, there is the following simple relationship between σ_e, the standard deviation of the error, and $\|\mathbf{e}\|$

$$\sigma_e^2 = \frac{1}{N-1} \sum_{n=0}^{N-1} e[n]^2 = \frac{1}{N-1} \|\mathbf{e}\|^2.$$

Therefore, using σ_e or $\|\mathbf{e}\|$ as the cost function in the minimization produces the same optimal p. There was one minor complication that stemmed from the fact that when n and/or p is large, the corresponding basis functions may not be fully orthogonal over the interval where the kernels are defined, e.g., [0 169] s for our test systems. We overcame this difficulty by modifying the cost function, i.e., $\|\mathbf{e}\|$, to incorporate a penalty for such cases. Specifically, when the basis functions are not orthogonal, $\|\mathbf{e}\|$ is assigned to a large positive number, which ensures that the search avoids such (n, p) pairs.

To check the orthogonality of the basis functions, we placed them in a matrix and multiplied the matrix with its transpose. Ideally, the resultant matrix should be the identity matrix of size Q_1. Therefore, if any of the elements of the resultant matrix deviated from the corresponding elements in the identity matrix by more than 10^{-6}, we identified the set of basis functions as not orthogonal. This rather firm test on the orthogonality improves the numerical condition of the estimation problem, i.e.,
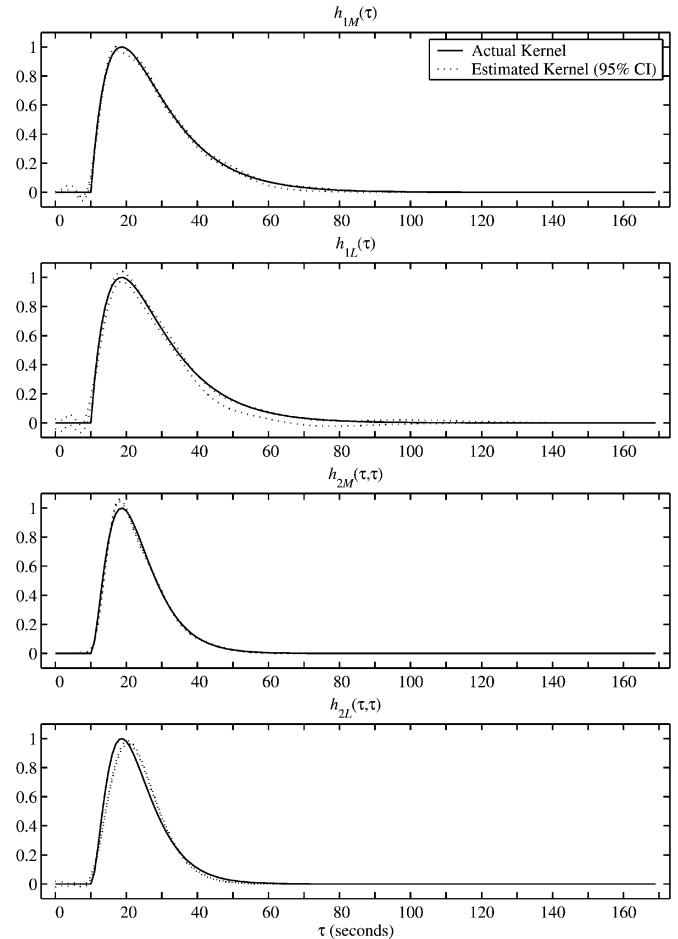

Fig. 3. Estimated first- and second-order kernel of the second test system that has delays, obtained after 100 trials using $Q_1 = 12$, $Q_2 = 6$. Upper two panels: $h_{1M}(\tau)$ and $h_{1L}(\tau)$, first-order kernels estimated using MBF and LBF. Lower two panels: $h_{2M}(\tau, \tau)$ and $h_{2L}(\tau, \tau)$, main diagonals of second-order kernels estimated using MBF and LBF. The solid and dotted lines show the actual kernels and the 95% confidence intervals (CI) for the estimated kernels, respectively.

increases the chances that the matrix $\mathbf{V}$ introduced in (5) is full-rank, which in turn helps produce coefficient estimates with less variance.

3) Simulation Results: In Fig. 3, we show a sample kernel estimation result obtained for the second test system using $Q_1 = 12$ and $Q_2 = 6$, under 0% output noise condition. After 100 optimal kernel estimation trials, for MBF the median of the optimal n and p parameters were 15 and 0.5864, respectively, whereas for LBF the median of optimal p parameter was 0.8191. (Each optimal kernel estimation trial or run takes about 0.4 s on a PC with 1.6 GHz Intel Pentium IV™ processor and 384 MB of memory, running under Microsoft Windows 2000™.) The estimated kernels are shown using bands of 95% confidence intervals (CIs). The upper two panels show the actual and estimated first-order kernels using MBF and LBF. The lower two panels show the diagonal of actual and estimated second-order kernels using MBF and LBF. We clearly see that the kernel estimation using MBF provides better results when the kernels have delays.

We repeated kernel estimation trials 100 times for both systems at different resolutions and output noise conditions and recorded $\|\mathbf{e}_M\|$ and $\|\mathbf{e}_L\|$, where the subscripts M and L refer

TABLE I
RESULTS OF THE SIMULATION STUDY

Resolution	Output Noise Level		Test System 1 (No Delay)		Test System 2 (Delay)	
			$\|\mathbf{e}_M\|$	$\|\mathbf{e}_L\|$	$\|\mathbf{e}_M\|$	$\|\mathbf{e}_L\|$
$Q_1 = 8$ $Q_2 = 4$	0%	Mean±SEM	9.1217±0.2678	9.1217±0.2678	52.3209±0.9955	125.2401±11.2612
		P-value	1.000000		<1e−6	
	5%	Mean±SEM	101.7061±3.3549	101.7773±3.3608	110.2192±3.2974	171.7964±10.6746
		P-value	0.988053		<1e−6	
	10%	Mean±SEM	143.4101±4.7351	143.5862±4.7479	146.7594±4.6614	208.0317±10.9927
		P-value	0.979066		0.000001	
$Q_1 = 12$ $Q_2 = 6$	0%	Mean±SEM	7.4834±0.2474	7.4834±0.2474	37.1981±0.6299	71.5167±1.3329
		P-value	1.000000		<1e−6	
	5%	Mean±SEM	101.2406±3.9610	101.4153±3.9717	105.7069±3.2794	124.8709±3.8056
		P-value	0.975181		0.000182	
	10%	Mean±SEM	142.8731±5.6011	143.1865±5.6146	144.4474±4.6446	159.7437±4.8867
		P-value	0.968519		0.024355	

to the results obtained using MBF and LBF, respectively. In Table I, we present the mean, the standard error of the mean (SEM), and the P-value of paired t-test for these two variables in each case.

We note that, when there is no delay in the kernels of the system, e.g., the first test system, MBF do not provide a statistically significant performance improvement over LBF under different resolutions and output noise conditions. However, when there is a delay in the kernels, e.g., the second test system, MBF is superior to LBF.

C. Experimental Data

To compare the performance of MBF and LBF in kernel estimation we analyzed photoreceptor responses of photoreceptor cells of adult fruitflies (*Drosophila melanogaster*) to WGN light. The dynamics of insect photo-transduction reactions, which convert light input into voltage output with an absolute delay, dead-time [20], has been studied at length [21]–[23]. Therefore, the photoreceptor responses provide the ideal testing data for evaluating the predictions of the two methods. Furthermore, the detailed knowledge about the workings of this preparation, the locations and properties of the reaction cascades and ion channels, gives one the confidence to relate the kernels to actual biophysical processes.

Intracellular current clamp recordings were made from green sensitive R1–6 photoreceptor cells at 25 °C using filamented quartz microelectrodes of resistance 60–150 MΩ. Voltage responses were sampled together with light stimuli at 1 KHz and filtered at 500 Hz using SEC-10L amplifiers (NPI Electronic GMBH, Tamm, Germany) and custom written MATLAB™ software (BIOSYST©, M. Juusola, 1997-2002) with an interface package for National Instrument boards (MATDAQ©, H.P.C. Robinson, 1997-2001).

Photoreceptors were stimulated with light from a small field (5° as seen by the fly) generated by a high-intensity green LED with peak wavelength 525 nm (Marl Optosource, U.K.) that was driven by a custom-built driver. The maximum light intensity level, or background, BG0 during the Gaussian white-noise stimulation was extrapolated to be at least 5×10^6 absorbed photons/second. This was done by placing neutral density filters, each reducing the light by a log-unit, between the light source and the eye until we could count identifiable voltage responses

to single photons [24]. The signaling characteristics of photoreceptors were studied at five backgrounds: BG0, BG-1, BG-2, BG-3, and BG-4 that covered a light intensity range of four log-units. In the experiment the photoreceptor was first studied at the lowest background (BG-4, ≈ 500 photons/s) before systematically proceeding to brighter BGs. Prior to recording, photoreceptors were adapted for 30 s to a chosen light background. However, to minimize the effects of noise and nonstationary adaptation the first 10-s-long response to the repeated WGN pattern was omitted and the data was averaged 10 times. For analysis, the light intensity of WGN stimulus was converted to contrast units $c = \Delta I / I$, where ΔI is the change in light intensity in respect to the mean I. The characteristic contrast of the WGN pattern was 0.32, defined by dividing the standard deviation of the contrast values by the unit mean [24]. The details of the set-up, recording criteria, light intensity calibration and data acquisition are further explained in [24] and [25].

III. RESULTS

We analyzed the responses of *Drosophila* photoreceptors at five light intensity levels (BGs), each one log-unit apart, to an identical WGN light contrast stimulus. Table II presents the results of second-order Volterra models for the five responses, from BG-4 (≈ 500 photons/s) to BG0 ($\approx 5 \times 10^6$ photons/s), using $Q_1 = 16$ and $Q_2 = 8$. The kernel estimation was done "optimally" as described in Section II-B2. The number of parameters to estimate/construct the first- and second-order kernels was 16 and $Q_2(Q_2 + 1)/2 = 36$, respectively. The last two columns in the table show the ratio of the number of significant parameters. The ratios presented in the table indicate that the given selection of Q_1 and Q_2 is very reasonable. Along with the norm of estimation error, the percentage of the error norm with respect to the output/response norm is also reported in the table. This figure indicates the percentage of signal power that was not explained by the model. We observe that from BG-3 to BG0, both MBF and LBF models perform acceptably well in terms of producing a low error power with respect to the output power. However, at BG-4, as the rates of photons, both emitted and absorbed, are low, the voltage responses are noisy. This together with the low number of averaged traces ($n = 10$) makes both models perform poorly. Nevertheless, in each case MBF produces a model with a lower estimation error norm.

TABLE II
COMPARISON OF SECOND-ORDER VOLTERRA MODELS
OBTAINED USING MBF AND LBF

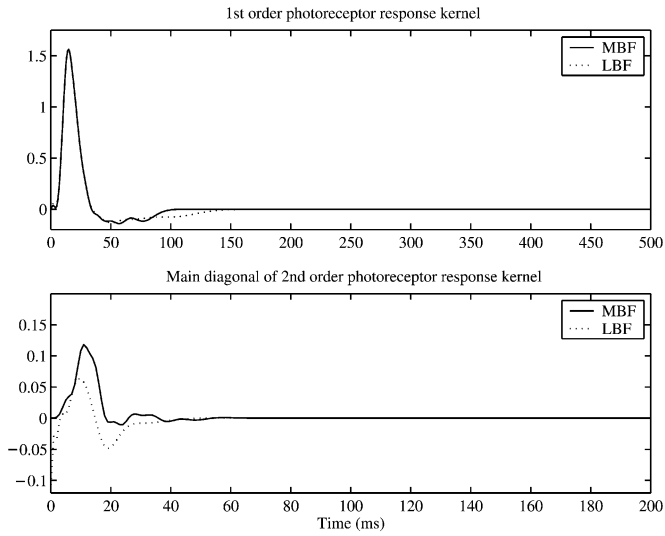| Photoreceptor Response	Basis	Optimal Parameters		$\|\mathbf{e}\|$ (% of $\|\mathbf{y}\|$)	Ratio of Significant Coefficients	
		n	p		1st ord kernel	2nd ord kernel
BG-4	MBF	15	0.5594	102.8226 (34.47%)	16/16	33/36
	LBF	0	0.7911	105.9469 (35.52%)	15/16	29/36
BG-3	MBF	19	0.4751	88.7502 (28.64%)	16/16	33/36
	LBF	0	0.7698	93.4042 (30.14%)	15/16	26/36
BG-2	MBF	17	0.3789	82.4540 (29.82%)	14/16	26/36
	LBF	0	0.5687	89.8707 (32.50%)	13/16	15/36
BG-1	MBF	2	0.8482	34.9956 (30.84%)	16/16	25/36
	LBF	0	0.8718	35.1624 (30.98%)	16/16	25/36
BG0	MBF	6	0.9463	28.7769 (71.20%)	15/16	30/36
	LBF	0	0.9581	29.4966 (72.99%)	12/16	24/36

Fig. 5.　Volterra kernels for the photoreceptor response collected at BG-1. Upper panel: first-order kernel, Lower panel: main diagonal of the second-order kernel, obtained using MBF (solid) and LBF (dotted).

Fig. 4.　Volterra kernels for the photoreceptor response collected at BG0. Upper panel: first-order kernel, Lower panel: main diagonal of the second-order kernel, obtained using MBF (solid) and LBF (dotted).

Fig. 6.　Volterra kernels for the photoreceptor response collected at BG-2. Upper panel: first-order kernel, Lower panel: main diagonal of the second-order kernel, obtained using MBF (solid) and LBF (dotted).

Figs. 4–8 show the estimated first-order (upper panels) and second–order (lower panels) kernels for the five different BG cases, using MBF (solid lines) and LBF (dotted lines). (Only the main diagonals are shown for the second-order kernels).

We also obtained third-order Volterra models, in an attempt to see how much performance improvement we can realize by incorporating the third-order kernels. The results are shown in Table III, where $Q_1 = 16, Q_2 = 8$, and $Q_3 = 4$. In this case, we had 16, 36, and $Q_3(Q_3+1)(Q_3+2)/6 = 20$ parameters corresponding to the first-, second-, and third-order kernels, respectively (for details of third-order Volterra kernel estimation see [26]). We observe that the inclusion of the third-order kernels introduce only a slight decrease $(< 3\%)$ in the norm of the estimation error. Therefore, one may argue that using a third-order model is not justified. We also note that almost 50% of the estimated coefficients for the third-order kernels are not significantly different than 0. However, as in the case of second-order Volterra models, for each different BG case, the use of MBF results in a lower estimation error norm.
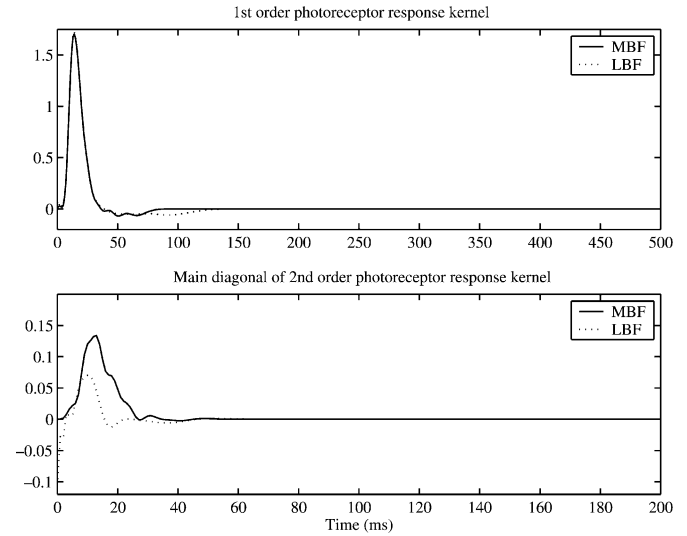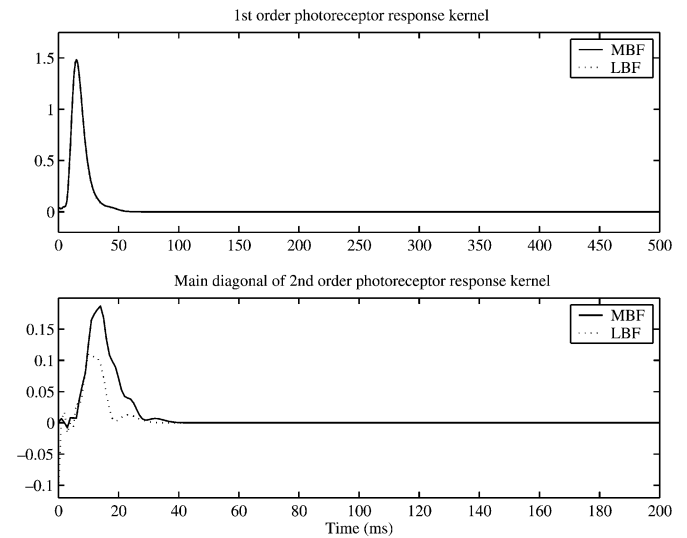
The voltage responses of *Drosophila* photoreceptors are superimposed on a light-induced mean potential, or dc-potential. When the mean light increases so does the dc-potential. In our exemplary recordings this was 2, 8, 13, 17, 20 mV for the five light backgrounds, respectively. *Drosophila* photoreceptors express voltage-sensitive potassium channels (Shaker-channels, Sh^{14}), whose time-dependent activation and inactivation in adult flies can be correlated to the shape of the second-order Volterra kernel at different dc-potentials [21]. This early positive nonlinear component amplifies and accelerates the photoreceptor voltage responses progressively as the dc-potential increases in brighter light [21]. Here, the second-order Volterra kernels obtained using MBFs behaved more consistently with the proposed activation-inactivation model of the shaker channel dynamics than the second-order

TABLE III
COMPARISON OF THIRD-ORDER VOLTERRA MODELS OBTAINED USING MBF AND LBF

| Photoreceptor Response	Basis	Optimal Parameters		$\|e\|$ (% of $\|y\|$)	Ratio of Significant Coefficients		
		n	p		1st ord kernel	2nd ord kernel	3rd ord kernel
BG-4	MBF	16	0.5374	99.6706 (33.42%)	16/16	34/36	12/20
	LBF	0	0.8670	104.4875 (35.03%)	14/16	22/36	13/20
BG-3	MBF	19	0.4751	86.1548 (27.80%)	16/16	35/36	14/20
	LBF	0	0.8679	92.3680 (29.80%)	15/16	28/36	12/20
BG-2	MBF	17	0.3983	80.4143 (29.08%)	12/16	28/36	16/20
	LBF	0	0.5740	89.6553 (32.43%)	13/16	16/36	4/20
BG-1	MBF	3	0.8304	34.5429 (30.44%)	14/16	26/36	11/20
	LBF	0	0.8705	34.9586 (30.80%)	16/16	26/36	8/20
BG0	MBF	6	0.9462	27.9864 (69.25%)	15/16	31/36	16/20
	LBF	0	0.9581	28.7712 (71.19%)	11/16	23/36	15/20
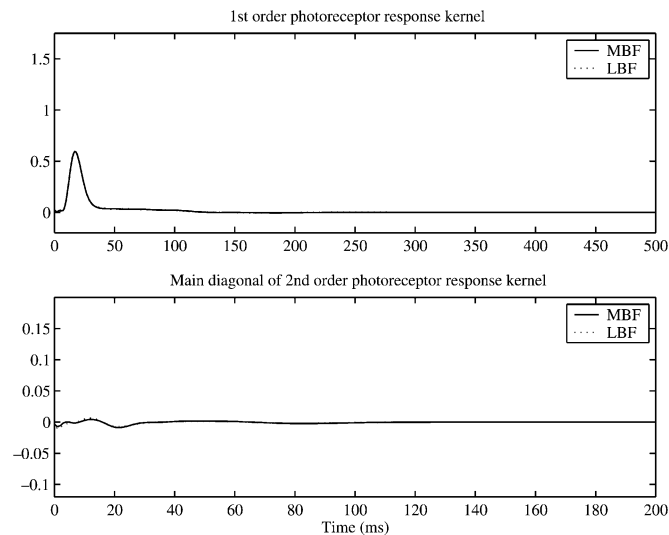
Fig. 7. Volterra kernels for the photoreceptor response collected at BG-3. Upper panel: first-order kernel, Lower panel: main diagonal of the second-order kernel, obtained using MBF (solid) and LBF (dotted).
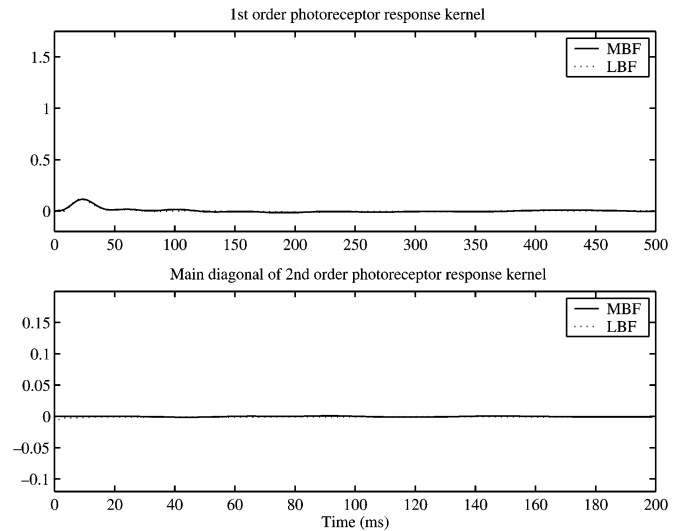

Fig. 8. Volterra kernels for the photoreceptor response collected at BG-4. Upper panel: first-order kernel, Lower panel: main diagonal of the second-order kernel, obtained using MBF (solid) and LBF (dotted).

Volterra kernels of Laguerre kind that showed unexpectedly prominent bipolarity at high BGs (Figs. 4–6).

For dim illumination, the general strategy of photoreceptors is to maximize photon capture and signal amplification [24], [27]. The filtering is very much low-passed as the slow and small amplitude kernels appear to enhance input redundancies (Figs. 7 and 8). On the other hand at high backgrounds, when bombarded by millions of photons each second, *Drosophila* photoreceptors must desensitize to prevent saturation. The kernels depict now a system that by adapting rapidly to the incoming photon rate is able to utilize the limited amplitude and frequency range of its voltage responses for the transfer of fast and localized signals (Figs. 4–6).

IV. DISCUSSION AND CONCLUSION

In this study, we explored the possibility of using MBFs, instead of widely known/used LBFs, in estimation of Volterra kernels of nonlinear and/or physiological systems using the least squares technique. The results obtained in the simulation study clearly indicate that using MBFs is advantageous over LBFs when there is a delay in the kernels. Our experimental results support the findings of the simulated data. This is judged by 1) their controlled, virtually oscillation-free onset, 2) universally

lower norm of estimation error at different noise conditions, and 3) more meaningful behavior as correlated to the known biophysical factors.

As delays are always possible/expected in physiological systems, we suggest that it may be more advantageous to use MBF expansion rather than LBF expansion in the least squares Volterra kernel estimation. This result is not surprising as MBFs are a superset of LBFs. However, what is surprising is that MBFs have not been used in the context of physiological system identification thus far.

In this study, we have also devised an optimization method through which the "optimal" basis function parameters can be selected. Given the number of basis functions, using basis functions generated with those optimal parameters produces a model with the minimum possible estimation error norm. (A Windows™ application named *MeixKerEst* that estimates Volterra kernels, up to third-order, from single input–output data using MBFs can be downloaded from the following URL: http://rc.kfshrc.edu.sa/bssc/staff/MusaAsyali/ Downloads.asp)

An underlying assumption of our modeling scheme was that the noise is assumed to exist only on the output, in which case the use of the ordinary least squares technique is justified. However, in experimental situations, there will be measurement noise on the input signal as well. It is known that the input

noise will bias the coefficient estimates [28]. The methods that address this issue include instrumental variables [29], modified least-squares algorithms that optimize the coefficients and estimate the noise covariance simultaneously [30], and other more recent approaches such as stochastic error whitening [31].

Another limitation of our modeling approach stemmed from the assumption that the higher order, i.e., higher dimensional, kernels are separable in their variables. This simplification, while serving to reduce the number of adaptable weights, severely limits the modeling capabilities of the Volterra series, not to mention the extreme truncations practiced (at order 2 or 3). The separability condition can be lifted by the use of multi-dimensional Laguerre or MBFs to approximate the higher order kernels. This way, the number of parameters can be still kept at a reasonable level while the restrictive separability assumption is removed from the model. However, in this case, we expect that the optimization procedure will be more complex, as it will involve generation of the higher dimensional basis functions and their test of orthogonality at each step of the search.

APPENDIX

Following the treatment given in [12], $\mathbf{A}^{(n)}$ (the $Q \times Q$ orthogonal matrix that transforms Laguerre functions to Meixner functions) can be expressed as

$$\mathbf{A}^{(n)} = \mathbf{X}^{(n)}\mathbf{Y}^n \quad (A.1)$$

where $\mathbf{Y}$ is an upper band matrix given as follows:

$$\mathbf{Y} = \begin{bmatrix} 1 & p & 0 & \dots & 0 \\ 0 & 1 & p & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}_{Q \times Q}$$

and $\mathbf{X}^{(n)}$ is a lower triangular matrix that we need to find out. (Q is the number of basis functions to be generated.) Since $\mathbf{A}^{(n)}$ is an orthogonal matrix, we have

$$\mathbf{A}^{(n)}\left\{\mathbf{A}^{(n)}\right\}^T = \mathbf{X}^{(n)}\mathbf{Y}^n \left\{\mathbf{Y}^n\right\}^T \left\{\mathbf{X}^{(n)}\right\}^T = \mathbf{I}. \quad (A.2)$$

The matrix $\mathbf{Y}^n \left\{\mathbf{Y}^n\right\}^T$ is a positive definite band matrix with equal elements along its diagonals. (The width of the band is $2n + 1$.) From (A.2) we conclude that $X^{(n)}$ can be determined by an inversion of the Cholesky factorization of $\mathbf{Y}^n \left\{\mathbf{Y}^n\right\}^T$. Next, $\mathbf{A}^{(n)}$ can be obtained by (A.1). Having calculated $\mathbf{A}^{(n)}$ in this way, a simple numerical check on the accuracy of the result can be obtained by comparing $\mathbf{A}^{(n)}\left\{\mathbf{A}^{(n)}\right\}^T$ with the identity matrix $\mathbf{I}$.

ACKNOWLEDGMENT

REFERENCES

[1] V. Volterra, *Theory of Functionals and of Integro and Integro-Differential Equations*. New York: Dover, 1959.
[2] M. H. Asyali, M. C. K. Khoo, and R. B. Berry, "Assessment of closed-loop ventilatory stability in obstructive sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 3, pp. 206–216, Mar. 2002.
[3] M. C. K. Khoo, "Understanding the dynamics of state-respiratory interaction during sleep: A model-based approach," in *Bioengineering Approaches to Pulmonary Physiology and Medicine*, M. C. K. Khoo, Ed. New York: Plenum, 1996, pp. 1–23.
[4] V. Z. Marmarelis *et al.*, "Nonlinear analysis of renal autoregulation in rats using principal dynamic modes," *Ann. Biomed. Eng.*, vol. 27, pp. 23–31, 1999.
[5] M. Iatrou, T. W. Berger, and V. Z. Marmarelis, "Application of a novel modeling method to the nonstationary properties of potentiation in the rabbit hippocampus," *Ann. Biomed. Eng.*, vol. 27, pp. 581–591, 1999.
[6] M. T. Chian, V. Z. Marmarelis, and T. W. Berger, "Decomposition of neural systems with nonlinear feedback using stimulus-response data," *Neurocomputing*, vol. 26–27, pp. 641–654, 1999.
[7] V. Z. Marmarelis, "New approaches in physiological system modeling," in *Advanced Methods of Physiological System Modeling*, V. Z. Marmarelis, Ed. New York: Plenum, 1987, vol. I, pp. 8–27.
[8] ——, "Volterra and Wiener series modeling of physiological systems," in *Advanced Methods of Physiological System Modeling*, V. Z. Marmarelis, Ed. New York: Plenum, 1989, vol. II, pp. 1–18.
[9] D. T. Westwick and R. E. Kearney, "Nonparametric identification of nonlinear biomedical systems, part I: Theory," *Crit. Rev. Biomed. Eng.*, vol. 26, pp. 153–226, 1998.
[10] M. J. Korenberg and I. W. Hunter, "The identification of nonlinear biological systems: Volterra kernel approaches," *Ann. Biomed. Eng.*, vol. 24, pp. 250–268, 1996.
[11] A. S. French and V. Z. Marmarelis, "Nonlinear analysis of neuronal systems," in *Modern Techniques in Neuroscience Research*, U. Windhorst and H. Johansson, Eds. New York: Springer-Verlag, 1999, pp. 627–640.
[12] A. C. D. Brinker, "Meixner-like functions having a rational z-transform," *Int. J. Circuit Theory Appl.*, vol. 23, pp. 237–246, 1995.
[13] P. M. Clarkson, *Optimal and Adaptive Signal Processing*. Boca Raton, FL: CRC, 1993.
[14] H. Akaike, "Power spectrum estimation through autoregressive model fitting," *Ann. Inst. Stat. Meth.*, vol. 21, pp. 407–419, 1969.
[15] ——, "A new look at statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, 1974.
[16] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2743–2760, Oct. 1998.
[17] J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Trans. Inf. Theory*, vol. IT-30, pp. 629–636, 1984.
[18] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Meads simplex method in low dimensions," *SIAM J. Optimiz.*, vol. 9, 1998, pp. 112–147.
[19] *Optimization Toolbox User's guide for Use With MATLAB*, The MathWorks Inc., Natick, MA, 2000.
[20] A. S. French, "Phototransduction in the fly compound eye exhibits temporal resonances and a pure time delay," *Nature*, vol. 283, pp. 200–202, 1980.
[21] M. Juusola, J. E. Niven, and A. S. French, "Shaker k^+ channels contribute an early nonlinear amplification to the light response in Drosophila photoreceptors," *J. Neurophysiol.*, vol. 90, pp. 2014–2021, 2003.
[22] J. H. van Hateren and H. P. Snippe, "Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells," *Vis. Res.*, vol. 41, pp. 1851–1865, 2001.
[23] R. C. Hardie and P. Raghu, "Visual transduction in Drosophila," *Nature*, vol. 41, pp. 186–193, 2001.
[24] M. Juusola and R. C. Hardie, "Light-Adaptation in Drosophila photoreceptors: I. Response dynamics and signalling efficiency at 25°C," *J. Gen. Physiol.*, vol. 117, pp. 3–25, 2001.
[25] M. Juusola and G. G. de Polavieja, "The rate of information transfer of naturalistic stimulation by graded potentials," *J. Gen. Physiol.*, vol. 122, pp. 191–206, 2003.
[26] M. H. Asyali, "Estimation of Volterra Kernels of Physiological System Using Meixner Basis Functions," BESC Dept., King Faisal Hospital and Research Center, Riyadh, Saudi Arabia, 2002.

[27] J. H. van Hateren, "Theoretical predictions of spatiotemporal receptive-fields of fly LMC's, and experimental validation," *J. Comp. Physiol. A*, vol. 171, pp. 157–170, 1992.

[28] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, 2nd ed. London, U.K.: Chapman & Hall, 1989.

[29] T. Söderström and P. Stoica, *Instrumental Variable Methods for System Identification*. Berlin, Germany: Springer-Verlag, 1983.

[30] S. Van Huffle, Ed., *Recent Advances in total Least Squares Techniques and Errors-in-Variables Modeling*. Philadelphia, PA: SIAM, 1997.

[31] Y. N. Rao, D. Erdogmus, G. Y. Rao, and J. C. Principe, "Stochastic error whitening algorithm for linear filter estimation with noisy data," *Neural Netw.*, vol. 16, no. 5–6, pp. 873–880, 2003.

Musa H. Asyali (M'99) received the B.S. degree in electrical and electronics engineering in 1990 from Bilkent University, Ankara, Turkey. He received the Ph.D. degree in biomedical engineering from the University of Southern California, Los Angeles, in 1998.

He joined the Electrical and Electronics Engineering Department, Ege University, Izmir, Turkey, as an Assistant Professor and became an Associate Professor in August 2001. Since October 2001, he has been working as a research scientist at the Biostatistics, Epidemiology and Scientific Computing Department of the King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia. His research interests include biomedical signal processing, physiological modeling, and bioinformatics.

Dr. Asyali is a member of IEEE Engineering in Medicine and Biology society.

Mikko Juusola was born in 1966. He received the M.D. and Ph.D. (neurophysiology) degrees from the University of Oulu, Oulu, Finland, in 1993.

From 1993–1994, he worked as an Alberta Heritage Foundation Medical Research Fellow at the University of Alberta, Edmonton, AB, Canada. From 1994–1996, he was a Postdoctoral Fellow at Dalhousie University, Halifax, NS, Canada. In 1995, he was made a Docent of Neurophysiology at the University of Oulu. In 1997–1999, he worked as an Academy of Finland Research Fellow in the Physiological Laboratory, University of Cambridge, Cambridge, U.K. He is currently a Royal Society University Research Fellow and the Principal Investigator of Drosophila Vision Laboratory in the Physiological Laboratory, University of Cambridge. His studies span from neural coding and probabilistic analysis to naturalistic stimulation, molecular biology, and genetics.