Nonlinear Identification Methods for Modeling Biomedical Systems

Robert Kearney & David Westwick

Department of Biomedical Engineering, McGill University
Montreal, Quebec, Canada, H4A 2B3
Email: rob@cortex.biomed.mcgill.ca

Introduction

The development of mathematical models of the dynamics of biomedical systems is an essential component of much biomedical engineering research and development. Models may be used for three reasons:

- *Control.* The objective is to model a plant's input-output response well enough to permit effective control system design. This requires a parsimonious model that nonetheless produces predictions of the system output, which are "good enough" for design purpose. Models for control applications tend to be low-order and parametric. Nonlinear and time-varying properties are often dealt with indirectly through the use of adaptive methods.
- *Insight.* The objective is to gain insight into how the dynamics of the system's components and their interconnections interact to define the overall operating characteristics. Nonlinearities must be dealt with explicitly since they often generate "interesting" aspects of the behavior.
- *Description.* The objective is to capture the characteristics of the system response in as accurate and concise a manner as possible. Nonlinear characteristics of the system response must be dealt with explicitly and the emphasis is on accurate prediction. Models of this type often are important elements in global models developed for insight.

Modeling techniques differ according to the ultimate use foreseen. This review will focus on methods appropriate for descriptive models where the task is to obtain concise, accurate descriptions of system behavior -- usually on the basis of experimental input-output data. Traditional *a priori* modeling techniques often cannot be applied to biomedical systems either because not enough is know about the underlying mechanisms or because the underlying mechanisms are too complex.

Models of the descriptive class must usually be identified using experimental data and so will be subject to the following practical constraints:

- The temporal and amplitude properties of the input signals will be restricted by the nature of the system, or by the apparatus used to perturb it
- Measurements will be corrupted by noise, which may be neither Gaussian nor white.
- Record lengths must be limited to minimize the effects of nonstationarities due to changes in the system with time.

This article will review some of the methods for the identification of nonlinear systems. To provide a context for the discussion, we begin with a formal definition of the various problems considered within the system identification field. Next, the different model structures available for use in describing nonlinear systems are described. The, paper concludes with a consideration of methods available for estimating the different model types.

System Identification

System identification is the process of building mathematical models of systems based on measurements of their input(s) and output(s) [1]. Figure 1 shows a general schema for the system identification problem [2]. The "system" to be identified consists of everything within the dotted box and

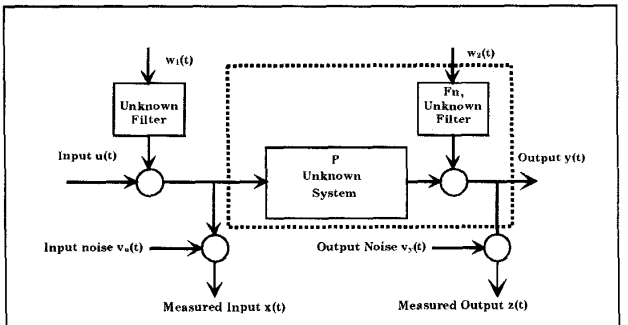

Figure 1 Schema for System identification

has both stochastic and deterministic parts. The stochastic part, F_n, is driven by a white-noise process, $w_2(t)$, which is not available to the experimenter. The deterministic part, P, is driven by the sum of a controlled input, $u(t)$ and a filtered version of an inaccessible white noise process, $w_1(t)$. The experimenter is assumed to have control over $u(t)$ but access

to only a noise corrupted version of the complete input signal, $x(t)$. The noise-free output, $y(t)$, is the sum of the outputs of the deterministic and stochastic parts of the system. However, the experimenter usually only has access to $z(t)$, a noisy version of the output signal,

Given this structure, three system identification problems can be distinguished.

1. *Identify the deterministic model*, P, defining the relationship between $u(t)$ and $y(t)$, assuming that there is no process noise. The input and output may be corrupted by observation noise. This is the problem of primary interest in this paper.
2. *Identify the noise model* F_n, defining the relation between $w_2(t)$ and $y(t)$, given observations of only the system output $y(t)$ [3, 4]. Usually, the input signal, $u(t)$, is assumed to be zero or constant. This approach is used where the inputs are not available to the experimenter. It is related to the study of "chaos" and "nonlinear dynamics" where the system is usually assumed to be autonomous, that is $w_2(t) = 0$.
3. *Identify the complete model* by estimating both the deterministic, P and noise, F_n, models. This problem formulation is used when accurate predictions are desired, such as in the design of model based control systems [1, 5].

Model Types

This section describes the most common types of models and their interrelations. The focus is on nonparametric, time-domain models.

Linear Impulse Response Functions

A linear system can be represented by its impulse response function (IRF) as:

$$y(t) = \int_o^T h(\tau)u(t-\tau)d\tau \quad \text{.....(1)}$$

Quasi-Linear Models

Nonlinear systems may often be linearized about a particular operating point so that

$$y(t) = \int_o^T h(\tau,\lambda)u(t-\tau)d\tau \quad \text{.....(2)}$$

where λ is a vector of parameters that define the operating point.

This provides a convenient means of describing how the system behavior changes with the operating point but provides no information about how the system behaves when moving between operating points.

The Volterra series

The Volterra series [6] generalizes the linear IRF for nonlinear systems using a series of integration kernels

$$y(t) = \sum_{n=0}^{\infty} \int_0^{\infty} \cdots \int_0^{\infty} h_n(\tau_1 \cdots, \tau_n) u(t-\tau_1) \cdots u(t-\tau_n) d\tau_1 \cdots d\tau_n \quad (3)$$

The first order kernel is similar to the linear IRF; indeed, for linear systems, there is only one term in the Volterra series and it is equal to the IRF.

For practical applications, Volterra models will be restricted to a finite Volterra series implemented in discrete time as:

$$y(t) = \sum_{n=0}^{Q} (\Delta t)^n \sum_{\tau_1=0}^{T-1} \cdots \sum_{\tau_n=0}^{T-1} h_n(\tau_1, \cdots, \tau_n) u(t-\tau_1) \cdots u(t-\tau_n) \quad (4)$$

where Q is the maximum kernel order and T is the memory length. The finite Volterra series can approximate the output of any system with "fading memory" [7] - that is any system which "forgets" about inputs in the distant past.

Once a system's Volterra kernels are known, the Volterra series is useful for computing the system's response to a particular input. However, the outputs of the different Volterra kernels are not orthogonal and consequently the kernels are difficult to determine experimentally since they must all be estimated simultaneously.

The Wiener Series

Wiener used the Gramm-Schmnit technique to orthogonalize the Volterra series under the assumption that the input was Brownian noise [8]. Subsequently this was reformulated for a Gaussian white noise input [9-11]. "White" inputs are impractical in continuous time since they will have infinite power. In discrete time, this problem does not arise since a white input is then simply a series of independent random variables.

The Wiener series can be represented by a set of functionals:

$$y(t) = \sum_{m=0}^{\infty} G_m[k_m(\tau_1, \cdots, \tau_m); u(t'), t' \le t] \quad \text{.....(5)}$$

which use the Wiener kernels, k_n, and the history of the input, to compute the current value of the output. In discrete time, the first four Wiener functionals are

$$G_0[k_0; u(t)] = k_0 \quad \text{.....(6)}$$

$$G_1[k_1(\tau); u(t)] = \sum_{\tau=0}^{T-1} k_1(\tau) u(t-\tau) \quad \text{.....(7)}$$

$$G_2[k_2(\tau_1,\tau_2);u(t)] = \sum_{\tau_1}^{T-1}\sum_{\tau_2=0}^{T-1} k_1(\tau_1,\tau_2)u(t-\tau_1)u(t-\tau_2) \quad \text{.....(8)}$$
$$-\sigma_u^2 \sum_{\tau=0}^{T-1} k_2(\tau,\tau)$$

$$G_3[k_3(\tau_1,\tau_2,\tau_3);u(t)] = \sum_{\tau_1=0}^{T-1}\sum_{\tau_2=0}^{T-1}\sum_{\tau_3=0}^{T-1} k_3(\tau_1,\tau_2,\tau_3)u(t-\tau_1)u(t-\tau_1)u(t-\tau_1) \quad \text{.....(9)}$$
$$-3\sigma_u^2 \sum_{\tau_1=0}^{T-1}\sum_{\tau_2=0}^{T-1} k_3(\tau_1,\tau_2,\tau_2)u(t-\tau_1)$$

Note the terms in the higher order kernels that depend explicitly on σ_u^2, the variance of the input. These are needed to ensure the kernels are orthogonal, that is:

$$E[G_iG_j] = 0 \; for \; i \neq j \; . \quad \text{.....(10)}$$

Relations are available to convert between the Volterra and Wiener series formulations [11, 12]. However, to use these relations all the Volterra (Wiener) kernels must be known since each Wiener (Volterra) kernel depends on all the Volterra (Wiener) kernels.

Block Structured Models

Many systems may be modeled using series interconnections of linear dynamic (L) and zero-memory nonlinear (N) subsystems. Three simple structures are of particular interest.

The Wiener (LN) system structure shown in Figure 2 consists of a single-input single-output (SISO) linear system followed by a SISO static nonlinearity [12, 13].

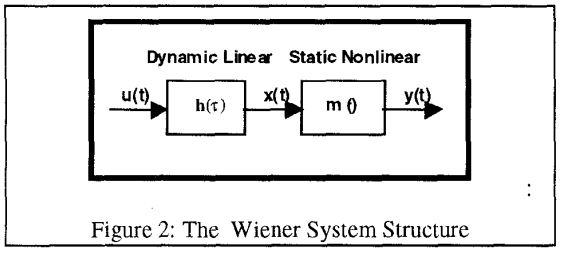

Figure 2: The Wiener System Structure

If the static nonlinearity can be represented by a polynomial:

$$m(x(t)) = \sum_{n=0}^{Q} c_n x^n(t) \quad \text{.....(11)}$$

then the Volterra kernels are given by

$$h_n(\tau_1,\cdots,\tau_n) = c_n h(\tau_1)h(\tau_2)\cdots h(\tau_n) \quad \text{.....(12)}$$

The Hammerstein system structure shown in Figure 3 consists of a static nonlinearity followed by a dynamic linear system [12, 13].

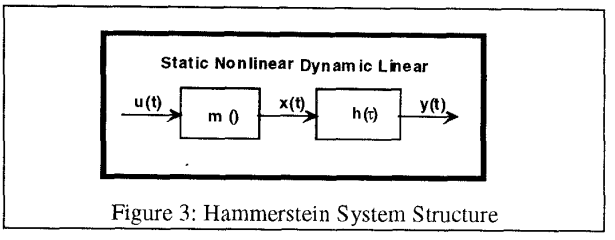

Figure 3: Hammerstein System Structure

In this case, the Volterra kernels are given by:

$$h_n(\tau_1,\cdots,\tau_n) = c_n h(\tau_1)\delta_{\tau_1,\tau_2}\delta_{\tau_1,\tau_3}\cdots\delta_{\tau_1,\tau_n} \quad \text{.....(13)}$$

and the kernels have non-zero values only on their diagonals

The LNL or Sandwich System structure shown in Figure 4 consists of two linear systems separated by a static nonlinearity [12, 14].

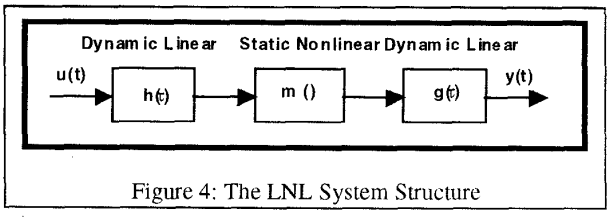

Figure 4: The LNL System Structure

The Volterra kernels of this structure are given by:

$$h_n(\tau_1,\cdots,\tau_n) = c_n \sum_{\sigma=0}^{T-1} g(\sigma)h(\tau_1-\sigma)h(\tau_2-\sigma)\cdots h(\tau_n-\sigma) \quad (14)$$

Parallel Cascades

Any system which can be described by a finite Volterra series can also be represented exactly by a finite number of Wiener pathways in parallel as show in Figure 5 [15, 16].

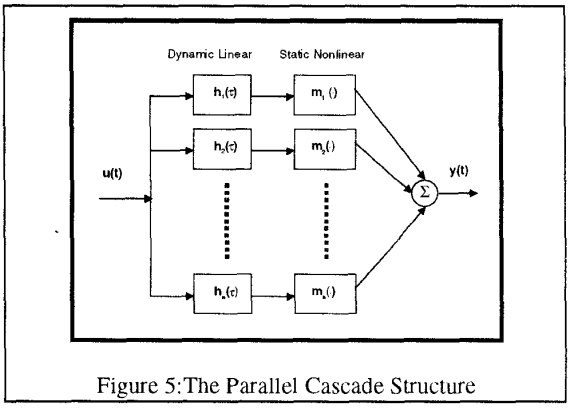

Figure 5:The Parallel Cascade Structure

The Volterra kernels of a parallel cascade system can be computed by noting that the n^{th} order Volterra kernel is simply the sum of the n^{th} order kernels of the individual pathways. Equation (11) can be used to determine the kernels for each pathway; summing the contributions of all pathways to each order then gives the overall Volterra kernel

Wiener-Bose Model

The Wiener-Bose model structure [17, 18] illustrated in Figure 6 is one in which the input signal is first passed through a bank of linear filters. The filter outputs are then applied to a multiple-input static nonlinearity, which is often expanded using an orthogonal polynomial set such as Grad-Hermite polynomials. The structure has been shown to very general and is capable of describing any "fading" memory nonlinear system [7].

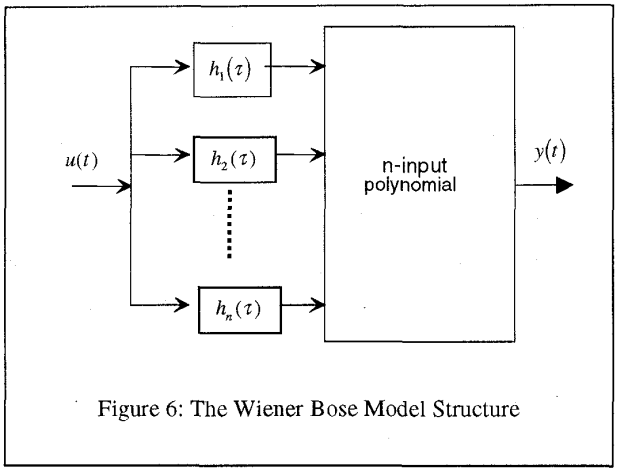

Figure 6: The Wiener Bose Model Structure

The structure of a Wiener-Bose model is not unique since it depends upon the choice of choice of filter bank and on the polynomial used to combine the filter outputs. Any set of filters may be used providing their outputs form a complete basis of a vector space whose dimension is equal to that of the system memory. Moreover, applying a linear transformation the filter bank of a Wiener-Bose model will generate an equally valid model with different polynomial coefficients.

If the filters in the filter bank are orthogonal, and the input is white Gaussian noise, their outputs will also be orthogonal. If the polynomial is a Hermite polynomial, the output of each term will be orthogonal to the outputs generated by all other terms in the static nonlinearity. As a result the Wiener-Bose model can form the basis for an orthogonal expansion of the kernels. Wiener called the subsystems generated by the individual polynomial terms "instrumental kernels" [8].

In this context is should be noted that the other nonlinear model structures we have considered may be viewed as special cases of the Wiener-Bose model. Thus:

- *The Volterra Kernel Model* may be viewed as a Wiener-Bose model with a filter bank consisting of a series of delays whose outputs are combined by a multi-input polynomial;
- *The Wiener Kernel Model* may be regarded similarly as consisting of a filter bank of delays whose outputs are combined by means of an orthogonal Grad-Hermite polynomial.
- *The Parallel Cascade Model* is simply a Wiener-Bose model in which the filter-bank corresponds to the IRFs of the individual pathways and the polynomial is chosen to have no cross-terms.

Identification Methods

This section will review the methods available to identify the different model structures from experimental data. The intent is to develop a common theoretical framework and to show how each of the methods fits within that structure.

Kernel Methods

Wiener's Original Method

Wiener's original approach was developed in a continuous-time framework with the intention of performing the identification in real-time using analog computers [8]. The approach was based on the Wiener-Bose model, with a filter bank consisting of orthogonal Laguerre filters and the multiple-input polynomial expanded using Grad-Hermite basis functions. The polynomial coefficients were to be estimated, one by one, from the cross-correlation between the measured output and the output of the appropriate instrumental kernel. There appears to have been only one physiological application of this approach; Stark used it to study nonlinearities in the pupillary control system [19].

Lee-Schetzen cross-correlation}

Lee and Schetzen [10] showed that Wiener kernels could be estimated using a series of cross-correlation provided the input was white Gaussian noise. If this is the case, the zero order Wiener kernel will be equal to the output mean.

$$\hat{k}_0 = \frac{1}{N}\sum_{t=1}^{N} y(t) = \hat{\mu}_y \quad \text{.....(15)}$$

This can be computed, and subtracted from the output, producing the zero-order residuals,

$$v_0(t) = y(t) - \hat{k}_0 \quad \text{.....(16)}$$

The first-order Wiener kernel is then simply the first-order cross-correlation between the input and the zero-order residuals, normalized with respect to the input power level.

$$\begin{aligned}\hat{k}_1(\tau) &= \frac{1}{\sigma_u^2}\hat{\phi}_{uv_0}(\tau) \\ &= \frac{1}{N\sigma_u^2}\sum_{t=0}^{N} u(t-\tau)v_0(t)\end{aligned} \quad \text{.....(17)}$$

The output of the first-order kernel is evaluated using a discrete convolution, and subtracted from the remaining system output to give the first-order residuals:

$$v_1(t) = v_0(t) - \sum_{\tau=0}^{T} \hat{k}(\tau)u(t-\tau) \quad \text{.....(18)}$$

The second-order kernel is then given by the normalized second-order cross-correlation between the input and the first-order residuals.

$$\begin{aligned}\hat{k}_2(\tau_1,\tau_2) &= \frac{1}{2\sigma_u^4}\hat{\phi}(\tau_1,\tau_2) \\ &= \frac{1}{2N\sigma_u^4}\sum_{t=0}^{N} u(t-\tau_1)u(t-\tau_2)v_1(t)\end{aligned} \quad \text{.....(19)}$$

Similarly, the n^{th} order Wiener kernel is equal to the n^{th} order cross-correlation between the input and the $(n-1)^{th}$ order residuals.

Correlation Methods

In the general case, the input-output correlation functions are not simply scaled copies of the kernels. Rather they will be the convolution of the kernels with the autocorrelation properties of the input signal [20]. Thus, the first-order cross-correlation function will be:

$$\phi_{ux}(\tau) = \sum_{i=0}^{T-1} k_1(i)\phi_{uu}(\tau - i) \quad \text{.....(20)}$$

and the second-order cross-correlation function will be:

$$\phi_{uuv}(\tau_1,\tau_2) = 2\sum_{i_1=0}^{T-1}\sum_{i_2=0}^{T-1} k_2(i_1,i_2)\phi_{uu}(\tau_1 - i_1)\phi_{uu}(\tau_2 - i_2) \quad (21)$$

Similar relations hold for higher-order correlations.

The Lee-Schetzen cross-correlation technique relies on the input spectrum being white so that input autocorrelation function will be a delta function:

$$\phi_{uu}(\tau) = \delta(\tau) \quad \text{.....(22)}$$

If this is the case, the kernels will be proportional to the input-output cross-correlation functions. However, if the input is not white, this simplification will not hold and the Lee-Schetzen method will fail.

Time-Domain Solutions

One way to correct for this problem is to remove the effects of input-correlation from the input-output cross-correlation functions. The relation for the first-order kernel can be written as the matrix equation:

$$\phi_{uy}(\tau) = \Phi_{uu}k_1 \quad \text{.....(23)}$$

where Φ_{uu} is a symmetric Toeplitz matrix [21]. The first-order Wiener kernel is then given by:

$$k_1 = \Phi_{uu}^{-1}\phi_{uy} \quad \text{.....(24)}$$

This approach also corrects for finite-length error in the correlation estimates and so gives better first-order Wiener kernel even with white inputs, as compared to simple cross-correlation [22]. If the bandwidth of the input is far from white, the conditioning of the matrix inversion may become bad leading to noise amplification. Replacing the exact inverse with a pseudo-inverse based on its singular value decomposition circumvents this problem [22].

Higher order kernels may be corrected by recognizing that each one-dimensional slice of the second-order cross correlation is given by:

$$\phi_{uuv}(\tau_1,i_2) = 2\sum_{i_1=0}^{T-1} k_2(i_1,i_2)\phi_{uu}(\tau_1 - i_1)\phi_{uu}(0) \quad \text{.....(25)}$$

so that the slice of the kernel can be recovered as

$$k_2(\tau_1,i_2) = \Phi_{uu}^{-1}\phi_{uuv}(\tau_1,i_2) \quad \text{.....(26)}$$

The complete kernel can be recovered by repeating the deconvolution for all slices [20].

Similar procedures can be used to correct higher-order Wiener kernels for non-white inputs. In general, computing the n^{th} order kernel will require that each one-dimensional slice of the n^{th} order cross-correlation, taken parallel to its n axes, be multiplied by the inverse of the Toeplitz structured auto-correlation matrix.

Frequency-Domain Solution

The Lee-Schetzen cross-correlation method can also be implemented in the frequency domain, resulting in N -fold computational savings, if the fast Fourier transform (FFT) was used to switch between domains.[23]. This approach can also correct for the effects of a non-white input spectrum.[24]

Orthogonal Algorithms

Another approach to estimating kernels with colored inputs is to formulate the problem as a linear regression and estimate all coefficients at once[25, 26]. For example, a system described by a second-order, Volterra series of memory length T will be given by:

$$y(t) = h_0 + \sum_{o}^{T-1} h_1(\tau)u(t-\tau) + \sum_{\tau_1=0}^{T-1}\sum_{\tau_2=0}^{T-1} h_2(\tau_1,\tau_2)u(t-\tau_1)u(t-\tau 2) \quad \text{.....(27)}$$

a relation with $N = T^2 + T + 1$ unknowns. If there are more than N observations it is possible to define a regression problem of the form:

$$\bar{\mathbf{y}} = \mathbf{X}\theta \quad \text{..... (28)}$$

where **X** is a matrix containing shifted copies of the input and products of shifted terms and θ is a vector of kernel weights. The difficulty with this approach is that the size of the problem rapidly becomes enormous since X will have $\frac{(T+Q)!}{T!Q!}$ columns where Q is the kernel order and T is the memory length.

Korenberg reformulated the problem in an equivalent form that could be solved implicitly without the need to generate the X matrix directly [27, 28]. This *fast orthogonal algorithm* results in a dramatic reduction in computation time and storage requirements.

Block Structured Models

The identification of simple block-structured models is based on Bussgang's theorem [29] that states that:

> *For two Gaussian signals, the cross-correlation function taken after one of them has undergone a nonlinear amplitude distortion is identical, except for a factor of proportionality, to the cross-correlation function taken before the distortion.*

Hammerstein Models

May be estimated using an iterative method [13] in which:

1) A linear system is first estimated between the output, $y(t)$, and the input, $u(t)$ to produce $\hat{h}^{-1}$, an estimate of the inverse of the linear dynamics.
2) $\hat{h}^{-1}$ is convolved with $y(t)$ to produce $\hat{x}(t)$, an initial estimate, of the intermediate signal.
3) The static nonlinearity $m(\bullet)$ is estimated by fitting a polynomial between the input and $\hat{x}(t)$.
4) The response of $m(\bullet)$ to the input, $u(t)$, produces an updated estimate of the intermediate signal, $\hat{x}'(t)$.
5) The inverse of the linear element, $\hat{h}^{-1}(\tau)$, is updated by fitting a linear system between the output and the updated estimate of the intermediate signal, $\hat{x}'(t)$.
6) Return to step 2 and continue until the process converges.

Note that the inverse filter may be non-causal, and the static nonlinearity never needs to be inverted.

Wiener Systems

Wiener systems may be estimated in a similar manner [13]. In this case,

1) A linear filter, $\hat{h}(\tau)$, is estimated between the input, u(t) , and the output, y(t).
2) The input $u(t)$ is convolved with $\hat{h}(\tau)$ to give an estimate of the intermediate signal $x(t)$.
3) A static nonlinearity is fit between $y(t)$ and $\hat{x}(t)$ to give $m^{-1}(\bullet)$, an estimate of the inverse of the static nonlinearity in the original system.
4) The output, $y(t)$, is transformed by this inverse estimate, producing an updated estimate, $\hat{x}(t)$, of the intermediate signal, $x(t)$.
5) The linear element $\hat{h}(\tau)$ is then re-estimated, this time between the input and $\hat{x}(t)$, the iteration returns to step 2 and continues until it converges.

The major difficulty with this approach lies in the estimation of the inverse of the static nonlinearity. If the static nonlinearity is not a one-to-one function, its inverse will not exist since there will be information present in the input to the nonlinearity, $x(t)$, that cannot be recovered from its output, $y(t)$.

LNL systems

Methods for identifying LNL cascade models depend on a extension of Bussgang's theorem [29] which states that the Wiener and Volterra kernels of LNL cascades are proportional. Hence, for a white Gaussian input signal:

$$\begin{aligned} \phi_{uy}(t) &= k\int_0^T g(\sigma)h(\tau-\sigma)d\sigma \\ \phi_{uuy}(\tau_1,\tau_2) &= k_2\int_o^T g(\sigma)h(\tau_1-\sigma)h(\tau_2-\sigma)d\sigma \end{aligned} \quad \text{.....(29)}$$

If $g(\sigma)$, the first linear element is not high-pass, this relation will be sufficient to identify both linear sub-systems. However, it the first system is high pass alternative approaches are needed.

Korenberg and Hunter [14] described an iterative procedure for LNL identification that proceeds as follows:

1) The convolution of the two linear elements is estimated from the first-order input-output cross-correlation.

2) A first-order unity-gain filter is constructed with a time constant that best fits this correlation, and used as an initial estimate for the first linear element, $\hat{h}(\tau)$.

3) Its output, $\hat{x}(t)$, is generated by convolving $\hat{h}(\tau)$ with $u(t)$.

4) A Hammerstein system is estimated between $\hat{x}(t)$ and the output, $y(t)$, using the iterative technique described above.

5) A relaxation technique is used to modify the estimate of the initial linear system, and the process repeated

Another approach is based on first estimating the system's kernels [30, 31] using, for example, the fast orthogonal algorithm. The first non-zero slice of the second-order kernel, $h_2(j,\tau)$, provides an estimate of the first linear subsystem $h(\tau)$. The IRF of the second linear element, $g(\sigma)$ can be obtained by deconvolving $h(\tau)$ from the first-order kernel which is the convolution of $g(\sigma)$ and $h(\tau)$.

Parallel Cascades

Methods for the identification of parallel cascade models proceeds iteratively as well [15, 16, 30-33]:

1) A block-structured, nonlinear system is estimated between the input and the output.

2) The output of this system is computed, and subtracted from the measured output.

3) A second block structured model is fit between the input and the output residuals.

4) This process is repeated from step 2 until the variance of the output residuals is reduced to the point where no additional statistically significant paths can be added to the model.

The key to the parallel cascade method's success lies in the estimation of the linear parts of the cascade paths. An early method used slices of input/output cross-correlation[16] functions, selected at random, as estimates of the linear subsystems of Wiener cascades. A subsequent approach estimated the Wiener pathway to reduce the cross-correlation functions to zero sequentially [32]. This yielded a unique and in some sense optimal solution but did not necessarily minimize the residual variance unless the input is white. Indeed, unless a complete system description is available it is not possible to find the Wiener cascade that minimizes the residual variance.

However, a constrained optimization is possible in which a two-step procedure is used to find "locally" optimal Wiener cascades [33]. The nonlinearity is first constrained to be a Hermite polynomial, of order equal to that of the lowest-order, non-zero input-residual cross-correlation function, and the MMSE linear element is found. Then the constraint is removed, and a high-order polynomial is fitted between the linear system output and the residuals. The optimal, first order pathway is given by:

$$h_1 = \Phi_{uu}^{-1}\phi_{uy} \quad \text{.....(30)}$$

Provided that the nonlinearity, $m(\bullet)$, is fit using a least-squares technique, the first-order input-residual cross-correlation will be reduced to zero. Furthermore, provided that least-squares fitting procedures are used for the nonlinearities, the first-order input-residual correlation will continue to be zero [30].

Once the first-order input-residual cross-correlation has been reduced to zero, the linear elements must minimize the residual variance when followed by a second-order Hermite polynomial. This is achieved by solving the generalized eigenvalue problem:

$$\phi_{uuv}h = \lambda\Phi_{uu}h \quad \text{.....(31)}$$

and choosing the generalized eigenvector, h, associated with the largest generalized eigenvalue, λ, as the IRF for the next pathway. Once the IRF has been calculated, a high-order polynomial is fit between its output and the current residuals using a least-squares fitting technique. This procedure will reduce the generalized eigenvalue associated with the IRF to zero, while leaving the other eigenvalues unaffected. Once an eigenvalue has been reduced to zero, it will remain zero. Thus, T paths will reduce the second-order input-residual cross-correlation to zero. In this form, the parallel cascade method produces an exactly orthogonal parallel Wiener cascade representation of the second-order Volterra kernel.

However, as in the linear case [22], the deconvolution, accomplished implicitly by the generalized eigenvalue problem, can become ill conditioned, producing large noise terms in the IRF estimates. These noise terms have little effect on the output, since they correspond to frequency components that are not present in the input. They can be eliminated by restricting the linear elements in the parallel cascade to the class of filters that are strongly excited by the input. This may be accomplished by projecting the IRFs onto a subspace spanned by the first few singular vectors of the input auto-correlation matrix thus forcing the IRF to contain only strongly excited components.

Wiener-Bose Models

Laguerre Expansion

One approach to the identification of Wiener-Bose model is to utilize a filter bank made up of discrete Laguerre filters [34-37]. The filters depend on a single parameter so their outputs can be computed efficiently recursively

The filter outputs are computed, and stored in a matrix, X whose first column contains ones and whose next n columns contain the linear outputs $x_i(t)$ for $i = 1 \rightarrow n$. If the second order kernel is desired, these are followed by the outputs of the second-order Hermite polynomials applied to the linear filter outputs. If the higher order kernels are desired, the outputs of the appropriate order Hermite polynomials must be computed and appended to X

The output of the general Wiener model can then be written as

$$\hat{y}(t) = X\theta \quad \text{.....(32)}$$

where θ is a vector containing the coefficients of the polynomials. Identification of the model is then achieved by solving the least squares problem as:

$$\theta = X^{+} y(t) \quad \text{.....(33)}$$

where X^{+} is the pseudoinverse of X.

Neuronal Modes

As noted above, a Wiener-Bose model can never be unique since the polynomial coefficients can be modified to reverse any linear transformation applied to the filter bank. Marmarelis and Orme [38] and Marmarelis [39] used this fact to develop a method that minimized the number of elements in the filter bank, and therefore polynomial coefficients. An eigendecomposition diagonalized a matrix containing the zero through second-order Laguerre expansion coefficients. In many cases, this coefficient matrix had few significant eigenvalues, and could therefore be approximated using relatively few terms. In such cases the model may be simplified by discarding components associated with small eigenvalues without significantly affecting its predictive power. These components were constructed by applying transformations, defined by the eigenvectors, to the filter bank. Those transformed filters associated with significant eigenvalues were termed "principal dynamic modes", and retained in the model. All others were discarded prior to least squares fitting the nonlinearity. Thus, changing the basis from the Laguerre filters to the principal dynamic modes produced a more compact model leading to better-conditioned parameter estimates and more efficient output computations.

Conclusion

This review has described a variety of model structures and estimation methods suitable for the identification of nonlinear systems in biomedical engineering applications. The Wiener-Bose model structure provides a common reference since all other model structures can be viewed as special cases. Moreover, it simplifies the interpretation of different estimation methods since they may be all interpreted in terms of identifying different components of Wiener-Bose models

Acknowledgements

Supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Medical Research Council of Canada.

References

[1]. L. Ljung, *System Identification: Theory for the user,* Englewood Cliffs, New Jersey, Prentice Hall Inc., 1987.

[2]. M. Verhaegen and P. DeWilde, "Subspace model identification part 1. The output-error state-space model identification class of algorithms," *International Journal of Control,* vol. 56, pp. 1187--1210 , 1992.

[3]. P.E. Caines, *Linear Stochastic Systems,* New York, John Wiley and Sons, 1988, 874 pages.

[4]. D.R. Brillinger, *Time Series Data Analysis and Theory,* New York, Holt, Rhinehart & Winston, 1975, 540 pages.

[5]. T. Soderstrom and P. Stoica, *System Identification,* New York, Prentice Hall, 1989, 612 pages.

[6]. V. Volterra, *Theory of functionals and of integral and integro-differential equations,* New York, Dover, 1959.

[7]. S. Boyd and L.O. Chua, "Fading Memory and the Problem of Approximating Nonlinear Operators with Volterra Series," *IEEE transactions on Circuits and Systems,* vol. CAS(32), pp. 1150--1161 , 1985.

[8]. N. Wiener, *Nonlinear problems in random theory,* New York, Wiley, 1958.

[9]. Y. Goussard, W.C. Krenz, and L. Stark, "An Improvement of the Lee and Schetzen Cross-Correlation Method," *IEEE Transactions on Automatic Control,* vol. 30, pp. 895--898 , 1985.

[10]. Y.W. Lee and M. Schetzen, "Measurement of the Wiener kernels of a non-linear system by cross-correlation," *International Journal of Control,* vol. 2, pp. 237--254 , 1965.

[11]. P.Z. Marmarelis and V.Z. Marmarelis, *Analysis of Physiological Systems,* New York, Plenum Press, 1978, 487 pages.

[12]. W.J. Rugh, *Nonlinear system theory. the Volterra/Wiener approach,* Baltimore, The Johns Hopkins University Press, 1981, 325 pages.

[13]. I.W. Hunter and M.J. Korenberg, "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models," *Biological Cybernetics,* vol. 55, pp. 135--144 , 1986.

[14]. M.J. Korenberg and I.W. Hunter, "The Identification of Nonlinear Biological Systems: LNL Cascade Models," *Biological Cybernetics,* vol. 55, pp. 125--134 , 1986.

[15]. G. Palm, "On representation and approximation of nonlinear sytems,," *Biological Cybernetics,,* vol. 34, pp. 49--52 , 1979.

[16]. M.J. Korenberg, "Parallel cascade identification and kernel estimation for nonlinear systems," *Annals of Biomedical Engineering,* vol. 19, pp. 429--455 , 1991.

[17]. M.G. Paulin, "A method for constructing data-based models of spiking neurons using a dynamic linear-static nonlinear cascade," *Biological Cybernetics,* vol. 69, pp. 67--76 , 1993.

[18]. V.Z. Marmarelis, "Signal Transformation and Coding in Neural Systems," *IEEE Transactions on Biomedical Engineering,* vol. 36, pp. 15--24 , 1989.

[19]. L. Stark, "The pupillary control system, its nonlinear adaptive and stochastic engineering design characteristics," *Automatica,* vol. 5, pp. 655--676 , 1969.

[20]. M.J. Korenberg and I.W. Hunter, "The identification of nonlinear biological systems: Wiener kernel approaches,," *Annals of Biomedical Engineering,* vol. 18, pp. 629--654 , 1990.

[21]. I.W. Hunter and R.E. Kearney, "Two-sided linear filter identification," *Medical and Biological Engineering and Computing,* vol. 21, pp. 203--209 , 1983.

[22]. D.T. Westwick and R.E. Kearney, "Identification of Physiological Systems: A Robust Method for Non-Parametric Impulse Response Estimation," *Medical & Biological Engineering & Computing,* vol. 35, pp. 83-90 , 1997.

[23]. A.S. French and E.G. Butz, "Measuring the Wiener Kernels of a Non-Linear System using the Fast Fourier Transform Algorithm," *International Journal of Control,* vol. 17, pp. 529--539 , 1973.

[24]. A.S. French, "Practical Nonlinear System Analysis by Wiener Kernel Estimation in the Frequency Domain," *Biological Cybernetics,* vol. 24, pp. 111--119 , 1976.

[25]. M.J. Korenberg, "Functional Expansions, Parallel Cascades, and Nonlinear Difference Equations,," vol. , pp. 221--240 , 1987.

[26]. M.J. Korenberg, S.B. Bruder, and P.J. McIlroy, "Exact Orthogonal Kernel Estimation from Finite Data Records: Extending Wiener's Identification of Nonlinear Systems," *Annals of Biomedical Engineering,* vol. 16, pp. 201--214 , 1988.

[27]. M.J. Korenberg, "Identifying nonlinear difference equation and functional expansion representations: The fast orthogonal algorithm," *Annals of Biomedical Engineering,* vol. 16, pp. 123--142 , 1988.

[28]. M.J. Korenberg, "Fast Orthogonal Algorithms for Nonlinear System Identification and Time-Series Analysis," in *Advanced Methods of Physiological Modelling,* Vol. 2, V.Z. Marmarelis, Editor, Plenum Press, New York, 1989, p. 165--178.

[29]. J.J. Bussgang, *Crosscorrelation functions of amplitude-distorted Gaussian signals,,* MIT Electrical Research Lab, 1952, 14 pages.

[30]. M.J. Korenberg, "Statistical identification of parallel cascades of linear and nonlinear systems," *IFAC Symp. Ident. Sys. Param. Est.,* vol. 1, pp. 669--674 , 1982.

[31]. M.J. Korenberg, "Statistical identification of Volterra kernels of high order systems,," *ICAS'84,* vol. 1, pp. 570--575 , 1984.

[32]. D.T. Westwick and R.E. Kearney, "Identification of Multiple-Input Nonlinear Systems Using Non-White Test Signals," in *Advanced Methods of Physiological Systems Modeling,* Vol. 3, V.Z. Mzrmarelis, Editor, Plenum Press, New York, 1994, p. 163--178.

[33]. D.T. Westwick and R.E. Kearney, "Generalized Eigenvector Algorithm for Nonlinear System Identification with Non-White Inputs," *Annals of Biomedical Engineering,* vol. 25, pp. 802-814 , 1997.

[34]. A. Watanabe and L. Stark, "Kernel method for nonlinear analysis: identification of a biological control system," *Mathematical Biosciences,* vol. 27, pp. 99--108 , 1975.

[35]. J. Amorocho and A. Brandstetter, "Determination of nonlinear functional response functions in rainfall runoff processes,," *Water Resources Research,,* vol. 7, pp. 1087--1101 , 1971.

[36]. V.Z. Marmarelis, "Identification of Nonlinear Biological Systems Using Laguerre Expansions of Kernels," *Annals of Biomedical Engineering,* vol. 21, pp. 573--589 , 1993.

[37]. V.Z. Marmarelis, "On Kernel Estimation Using Non-Gaussian and/or Non-White Input Data,," in *Advance Methods of Physiological Systems Modeling,* Vol. 3, V.Z. Marmarelis, Editor, Plenum Press, New York, 1994, p. 229--242.

[38]. V.Z. Marmarelis and M.E. Orme, "Modeling of Neuronal Systems by Use of Neuronal Modes," *IEEE Transactions on Biomedical Engineering*, vol. 40, pp. 1149--1158 , 1993.

[39]. V.Z. Marmarelis, "Nonlinear Modeling of Physiological Systems using Principal Dynamic Modes," in *Advanced Methods of Physiological Systems Modeling*, Vol. 3, V.Z. Marmarelis, Editor, Plenum Press, New York, 1994, p. 1--27.