To appear in Journal of Systems Engineering, Vol 6, No. 6, 1996

Nonlinear System Identification Using Discrete Laguerre Functions

Andrew D. Back and Ah Chung Tsoi

Department of Electrical and Computer Engineering
University of Queensland
St Lucia, Queensland 4072. Australia
Phone: +61 7 3365 3965, +61 7 3365 3950
Fax: + 61 7 3365 4999
Email: {back,act}@elec.uq.edu.au

Keywords: Nonlinear System Identification, Orthonormal filter, Laguerre, Volterra Series, Neural Network

Abstract

Volterra functional series approximations are well known techniques for nonlinear system identification. These structures, while quite general, suffer from difficulties in parameterizing the kernels, and may require a large number of parameters for models higher than second order. Discrete Laguerre filters and other orthogonal filters have been re-considered recently for both linear and nonlinear system identification. A class of multilayer perceptrons incorporating an orthogonal filter preprocessing unit is introduced. Derivations for parameter estimation of these models are given, including a consideration of both off-line and on-line methods for estimating the Laguerre pole. The paper concludes with a brief discussion of some open issues concerning such signal models.

1 Introduction

In classical signal processing, signal models are described by linear time-invariant (LTI) models using a rational transfer function or pole-zero formulation. The aim is to obtain a parsimonious model of the process or signal being identified, modelled or predicted using a small number of parameters.

Consider a single-input single-output (SISO) discrete-time, causal, infinite dimensional, time-invariant dynamical system[1] described by $y(t) = \sum_{k=0}^{\infty} h(k)x(t-k) + v(t)$ where $\{h(k)\} \in \Re^1$ is a weighting sequence, $\{v(t)\}$ is a disturbance sequence of zero-mean, independent and identically distributed random variables and $x(t)$ is the input to the model at time t. We seek to model $\{y(t)\}$ by a finite-dimensional signal model, producing an output $\{\hat{y}(t)\}$ and resulting in an output prediction error sequence[2] $\{e(t)\}$. The system output $y(t)$ can be represented as $y(t) = \hat{y}(t) + e(t)$

A linear discrete time time-invariant signal model is given by

$$\hat{y}(t) = G(z,\theta)x(t) \tag{1}$$
$$= \varphi'(t)\theta \tag{2}$$

where $G(z,\theta)$ is a linear transfer function, z is the usual shift operator defined as $zx(t) \stackrel{\Delta}{=} x(t+1)$; and θ is some parameter vector. $\varphi(t)$ is a vector dependent only on the system input $x(t)$ and its past values; and past values of the system output $y(t)$. The nature of this will be made clear as we discuss various possible signal models. The prime operator $'$ represents the vector or matrix transpose.

For the usual moving average model,

$$G(z,\theta) = \sum_{i=0}^{M} b_i z^{-i} \tag{3}$$

where $\theta = [b_0, ..., b_M]'$, $\varphi(t) = [x(t), ..., x(t-M)]'$. This is a finite impulse response (FIR) signal model, which, for high order systems (i.e., M large), can be more efficiently approximated by an infinite impulse response (IIR) filter structure

$$G(z,\theta) = \frac{\sum_{i=0}^{M_1} b_i z^{-i}}{1 + \sum_{i=1}^{M_2} a_i z^{-i}} \tag{4}$$

where M_1 and M_2 are integers. There are a number of disadvantages with both FIR and IIR signal models. These include: determination of the model order, problems in the convergence of gradient descent algorithms for IIR models due to multimodal mean square error surfaces and sensitivity of the model poles and zeros with respect to the model coefficients [1].

Further, it has been recognized that there is often a requirement for different signal models to serve different purposes (see [2] and references therein). For example, on the one hand, in a theoretical

[1] In this paper, this is sometimes synonymously called the plant.

[2] $e(t)$ is different from $v(t)$ because we assume that the plant could be infinite dimensional, and the signal model is finite dimensional.

sense, we may choose one model structure which gives explanations to the behaviour of the plant. On the other hand, we may choose a different model structure if robustness in practical, industrial process control applications is required. Lerner gives an interesting discussion of the differences between models which are best for conveying information about the process and those useful for accurate identification at the expense of understanding the process [3].

In this respect, the use of exponential signal models has proven to be an important development. Consider the impulse response of a rational LTI model which can be described by the summation of P complex exponentials [4]

$$h(t) \quad = \quad \sum_{i=1}^{P} A_i e^{b_i t} \qquad\qquad t = 0, 1, ..., T-1 \tag{5}$$

where A_i, and $e^{b_i t}$, are the amplitudes and complex frequencies respectively, with $b_i = \alpha_i + \jmath 2\pi f_i$, α_i is the damping factor and f_i is the frequency of the ith exponential. T is an integer. If the signal is a sum of undamped exponentials, then $\alpha_i = 0, \forall i$. Methods to estimate A_i, and b_i parameters include Prony's method, least squares Prony's method and the Tufts-Kumaresan method [4].

One problem with the above approach, is how to choose the order P such that the approximation between the model and the true system is sufficiently close to one another[3]. Kautz [5], considered this problem and proposed a related model

$$h(t) \quad = \quad \sum_{i=1}^{N} c_i \psi_i(t) \tag{6}$$

where $\psi_i(t)$ is an orthogonalized exponential. Thus, unlike models such as direct form FIR and IIR structures, the parameters $\{c_i\}$ do not depend on the order N. An orthonormal function $g_i(x)$ is defined by [6]

$$\int_a^b g_i(x) g_j(x) dx \quad = \quad \delta_{ij} \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{7}$$

where δ_{ij} is the Kronecker delta. The condition for $i = j$ is one of normality, while the condition for $i \neq j$ is one of orthogonality. Together they imply orthonormality of $g_i(x)$.

The problem of finding a convergent orthonormal series with complex exponential representation reduces to the problem of seeking an orthonormal function $\psi_i(t)$ which is a linear combination of the exponential functions $\{e^{a_k t}; k = 1, .., n\}$ [5, 6]. We require

$$\psi_k(t) \quad = \quad \sum_{i=1}^{Q} \upsilon_i e^{a_i t} \tag{8}$$

where υ_i is a constant. By defining an admissible weighting function $w(x) = (\alpha x)^n e^{-\alpha x}$, setting $(a, b) = [0, \infty)$ and normalizing the integral

$$\int_a^b \zeta_i(x) \zeta_j(x) w(x) dx \quad = \quad \delta_{ij} \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{9}$$

[3] In the sense that an error criterion does not exceed a certain prescribed threshold.

we obtain the kth order continuous-time Laguerre function[4]

$$\psi_k(t) \quad = \quad \sqrt{2\alpha}\frac{e^{\alpha t}}{(k-1)!}\frac{d^{k-1}}{dt^{k-1}}\left(t^{k-1}e^{-2\alpha t}\right) \tag{10}$$

Evaluation of the Laguerre function for the first few values of k gives (see also [6])

$$\begin{aligned}
\psi_1(t) &= \sqrt{2\alpha}e^{-\alpha t} \\
\psi_2(t) &= \sqrt{2\alpha}e^{-\alpha t}(1-2\alpha t) \\
\psi_3(t) &= \sqrt{2\alpha}e^{-\alpha t}(1-4\alpha t+2\alpha^2t^2)
\end{aligned}$$

Summation of the $\psi_k(t)$ functions results in the required exponential sum as specified in (8).
These functions form an orthonormal basis complete on $L_2[0,\infty)$ [6, 7, 8]. A system of functions in the interval $0 \le t < \infty$ is complete if every piecewise continuous function $f(t)$ for which the integral $\int_0^\infty f^2(t)dt$ exists can be approximated arbitrarily well in the mean square sense by a linear combination of the functions [9].
Taking the Laplace transform of the Laguerre function $\psi_k(t)$ gives

$$L_k(s,\alpha) \quad = \quad \sqrt{2\alpha}\frac{(s-\alpha)^{k-1}}{(s+\alpha)^k} \tag{11}$$

where α is referred to as the Laguerre pole or the Laguerre time constant.
The development of models based on the orthonormal Laguerre functions has seen a small, but significant interest over many years, dating back at least to the pioneering work by Lee [6], Kautz [5], and Young and Huggins [2]. More recently, this work has been extended from continuous-time to discrete-time. The discrete Laguerre function $L_k(z,\alpha)$ [10, 11] is given by

$$L_k(z,\alpha) \quad = \quad \frac{\sqrt{1-\alpha^2}}{z-\alpha}\left(\frac{1-\alpha z}{z-\alpha}\right)^{k-1} \qquad k \ge 1 \tag{12}$$

These functions can be interpreted as discrete Laguerre filters where the first section represents a first-order low pass filter, followed by $(k-1)$ all-pass sections. The orthonormal set is obtained as [12]

$$\frac{1}{2\pi j}\oint L_i(z)L_k(z^{-1})\frac{dz}{z} \quad = \quad \delta_{ik} \tag{13}$$

Thus, a signal model based on the Laguerre function can be described as follows [11]

$$y(t) \quad = \quad \sum_{k=1}^{M} b_k L_k(z,\alpha)x(t) \tag{14}$$

[4] In this paper, we will use two alternative notations for the Laguerre functions, $\psi_k(t)$ and $\psi_k(t,\alpha)$. We will normally use the notation $\psi_k(t)$. However, if we wish to emphasize the importance of the Laguerre pole, α, we will use the notation $\psi_k(t,\alpha)$.

where b_k is the kth regression coefficient, and $L_k(z, \alpha)$ is the kth discrete Laguerre function given by eqn (12). By setting $\alpha = 0$, the usual FIR model is obtained. The linear Laguerre model can be re-written as $y(t) = \sum_{k=1}^{M} b_k l_k(t)$ where $l_k(t)$ is defined as

$$\begin{aligned} l_1(t) &\triangleq F_0(z,\alpha)x(t) \\ l_k(t) &\triangleq \bar{L}(z,\alpha)l_{k-1}(t) \qquad k = 2, ..., M \end{aligned} \tag{15}$$

where

$$F_0(z,\alpha) = \frac{\sqrt{1-\alpha^2}}{z-\alpha} \tag{16}$$

$$\bar{L}(z,\alpha) = \left(\frac{1-\alpha z}{z-\alpha}\right) \tag{17}$$

Laguerre-based models give good performance on systems with low frequency characteristics; on the other hand, they do not give good performance for modelling systems with oscillatory behaviour [11]. Kautz functions, due to its ability to model complex poles [5], have been introduced to offer a better performance for systems having a resonant behaviour [13] . The Kautz model is defined in the same manner as the linear Laguerre model (see equation (14)), except the filter function $\Psi_k(z, \alpha, \beta)$ is used instead of $L_k(z, \alpha)$, where

$$\Psi_k(z,\alpha,\beta) = \begin{cases} \frac{\sqrt{1-\beta^2}(z-\alpha)}{z^2+\alpha(\beta-1)z-\beta} \left[\frac{-\beta z^2+\alpha(\beta-1)z+1}{z^2+\alpha(\beta-1)z-\beta}\right]^{\frac{k-1}{2}} & k \text{ odd} \\ \frac{\sqrt{1-\beta^2(1-\alpha^2)}}{z^2+\alpha(\beta-1)z-\beta} \left[\frac{-\beta z^2+\alpha(\beta-1)z+1}{z^2+\alpha(\beta-1)z-\beta}\right]^{\frac{k-2}{2}} & k \text{ even} \end{cases} \tag{18}$$

The system identification problem using discrete Laguerre and Kautz filters has been considered recently by Wahlberg et al. [13, 11]. Discrete-time Laguerre models have been shown to be insensitive to convergence problems resulting from rapid sampling or equivalently, plants having poles too close to the unit circle [11].

Zervos, Belanger and Dumont used a continuous-time Laguerre series expansion to model the closed loop plant in a control application [38]. In this case, the model provided a robust representation given little prior information as compared with a usual transfer function model. Others have considered various aspects of discrete-time modelling and control, see [14, 15] for some examples.

This paper considers a special class of nonlinear systems, viz., one which is based on neural network architectures. In particular, our proposed signal model is based on the classic multilayer perceptron (MLP) model and orthogonal filters, e.g., Laguerre filters. It will be shown that this class of models can be used to approximate a Volterra signal model to an arbitrary degree of accuracy. Further, we derive on-line parameter estimation algorithms for the models. Our approach differs from previously closed form methods [16, 17, 7] used to determine the Laguerre time-constants, in that it is an on-line approximation method and does not require impulse response information from the system being modelled.

The aims of the paper are to provide a brief overview of the use of Laguerre models for system identification, and in particular, to focus on the very recent consideration of nonlinear Laguerre models. In section 2 we review a class of nonlinear models, viz.,the Volterra orthogonal filter models. In section 3 we define the MLP Laguerre model and show the links between the Volterra Laguerre and MLP Laguerre models. Section 4 contains reviews of algorithms for determining the Laguerre pole. Derivations are given for an on-line learning algorithm for the Laguerre pole in the proposed model. Open issues are considered in section 5 and conclusions are given in section 6.

2 Volterra Orthogonal Filter Models

2.1 Background

It is well known that many physical processes are intrinsically nonlinear, and cannot be adequately approximated by linear models. A popular method of nonlinear modelling is to use functional series approximation techniques, e.g., the Volterra series [18]. This model is readily seen to be a generalization of the linear impulse response model to the nonlinear case. A discrete-time Volterra series model may be written as follows:

$$\begin{aligned} y(t) &= h_0 + \sum_{\tau_1=-\infty}^{\infty} h_1(\tau_1)u(t-\tau_1) + \\ & \sum_{\tau_1=-\infty}^{\infty}\sum_{\tau_2=-\infty}^{\infty} h_2(\tau_1,\tau_2)u(t-\tau_1)u(t-\tau_2) + \cdots \\ & \cdots + \sum_{\tau_1=-\infty}^{\infty}\cdots\sum_{\tau_n=-\infty}^{\infty} h_n(\tau_1,\tau_2,...,\tau_n)\prod_{k=1}^{n} u(t-\tau_k) + \cdots \end{aligned} \tag{19}$$

where h_i is known as the ith Volterra kernel. The impulse response corresponds to the case when $h_n = 0$, for $n > 1$. Methods of determining the kernels in the Volterra series are based on cross-correlation [19, 20], and frequency domain methods [21].

While the Volterra series is capable of modelling nonlinear dynamic systems, it has some major disadvantages. These include

1. The difficulty in obtaining the required number of kernels in order to approximate a given system [19]. In a sense, the number of kernels required to approximate a nonlinear system gives a measure of the "degree of nonlinearity" of the system. Thus, a system which is "mildly" nonlinear may require only a second order Volterra series to give satisfactory results, while a "highly" nonlinear system may require a large number of kernels to obtain a satisfactory approximation. This is a significant problem as including higher order kernels require far more memory storage. For this reason, most practical applications of Volterra series are often limited to only 2nd order terms.

2. An additional problem associated with the Volterra series representation is in requiring that the function to be approximated be analytic. It is known that modelling certain nonlinearities (such as a saturating element) presents significant difficulties [19] to the extent that the series may not converge [20]. Although the Volterra series may not converge, the application of the

Weierstrass theorem to the Volterra series [20] means that the approximation to any continuous function $f(x)$ can take place over a finite interval (a, b) with an error $\epsilon > 0$ (ϵ can be made arbitrarily small). Palm [22] showed that the Volterra series can represent all polynomial systems if the kernels are distributions. This still excludes all nonlinear systems which are not analytic.

It is apparent that while of considerable theoretical interest, Volterra series models leave much to be desired in terms of practical nonlinear system identification. These disadvantages have contributed to the development of techniques aimed specifically to overcome such problems. In particular, there has been recent interest in using discrete-time Laguerre functions as a basis for nonlinear models. While the classical Wiener model [21, 23] also used Laguerre functions in its development, the approach used in this paper appears to be fundamentally different.

2.2 Comments on the General Structures

It may be useful to give a general formulation of the ideas which had been expressed in the literature, and which will form the material in the following subsections. There are a number of ways in which the nonlinearity may enter the system, for example

1. as a memoryless nonlinearity at the input end,
2. as a memoryless nonlinearity at the output end,
3. as a memoryless nonlinearity somewhere in the internal structure of the model.

Alternatively, we may have nonlinearities which have memory, e.g., hysteresis nonlinearity in any of these locations. In addition, the dynamics in the model may be linear or nonlinear. In the literature, it appears that much work is concentrated on the following structure:

a) The nonlinearity is memoryless, and occurs only at the output

b) The dynamic subsystem is linear, and occurs at the input end

This special case can be broadly interpreted as a generalization of the usual linear system description, except that the output becomes nonlinear in order to take into account of the possible nonlinear behaviour of the observed system. In view of the overwhelming importance of this special case, in the following subsections, we will consider only this case (see [21, 29] for discussions on related structures).

A general description of this class of models is given as follows:

$$y(t) = f(\hat{x}(t)) \tag{20}$$

$$\hat{x}(t) = H(z)x(t) \tag{21}$$

where $x(t)$ is the input signal, and $y(t)$ is the output signal. $\hat{x}(t)$ is the output, and $H(z)$ is the transfer function of the linear subsystem.

So far, most of the papers in the literature (see e.g., [10, 25]) considered the case when $H(z)$ is a Laguerre filter. A disadvantage with the Laguerre model is that it employs a pole only on the real axis, and is therefore not particularly well suited for modelling systems with transfer functions having complex poles [11]. In addition, the Laguerre model may have convergence problems for systems with widely spaced poles [11]. There are a number of extensions which can be considered as a means of overcoming these problems. The differences among the proposed extensions depend on the nature of the dynamic linear subsystem. For example,

1. Fejzo used a lattice Laguerre filter as the linear subsystem [24] (see also [26, 27]).
2. Wahlberg used a linear system which consists of multiple Laguerre filters [11].

We will denote each of the above mentioned models by the more generic name: *Volterra orthogonal filter* model and in particular, each model as: the Volterra Laguerre model, the Volterra Lattice model and the Volterra Multiple Laguerre model. It is simple to observe that it is also possible to use a Kautz filter as the linear dynamic subsystem and hence we introduce the Volterra Kautz model in a later section.

In the following subsections, we will briefly review the most relevant of these models: Volterra Laguerre, Volterra Multiple Laguerre and Volterra Kautz models [5]

2.3 The Volterra Laguerre Model

Dumont and Fu [10] and independently Alshebeili, Mertzios and Venetsanopoulos [25] derived a Volterra Laguerre model representing the nonlinear system in (21). In this approach, the model is derived by firstly considering the usual Volterra series, where $f(\cdot)$ in (20) is represented by a power series expansion. Next, the linear system $H(z)$ is described by its impulse response function to obtain the Volterra series model. Instead of leaving it in this form however, the Volterra kernels are approximated by orthonormal functions, allowing the final Volterra Laguerre model to be obtained. A representation of the Volterra Laguerre model is given in Figure 1.

More specifically, consider the second order Volterra model proposed in [10]

$$y(t) = y_1(t) + y_2(t) \tag{22}$$

where $y_i(t)$, $i = 1, 2$ are the outputs of the i th Volterra kernel; these Volterra kernels can be approximated by orthonormal functions

$$h_1(\tau_1) = \sum_{k=1}^{N} c_k \psi_k(\tau_1) \tag{23}$$

[5]The Volterra lattice model [24] is interesting in itself. However, since our main aim is to consider the MLP orthogonal filter model (see later sections), to describe the Volterra lattice model would require us to introduce many more notations, which may obscure the flow of the presentation in this paper. It suffices to indicate that the ideas developed in this paper can be applied equally to formulate a MLP lattice model, except that the treatment would involve more complicated notations.

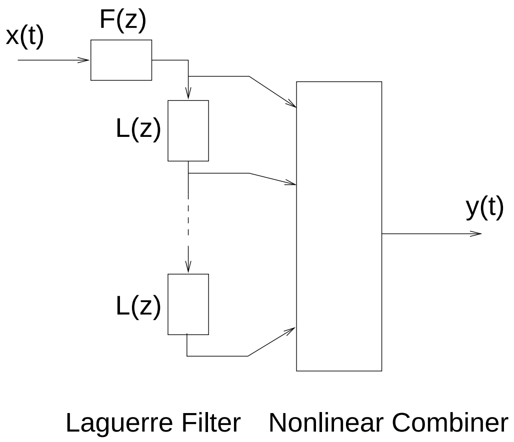


Figure 1: The Volterra Laguerre model.

$$h_2(\tau_1, \tau_2) = \sum_{n=1}^{N} \sum_{m=1}^{N} c_{nm} \psi_n(\tau_1) \psi_m(\tau_2) \tag{24}$$

where $\psi_k(\tau)$ is the kth order Laguerre function with pole α, defined in (10). For an input $x(t)$ to the kth order Laguerre filter, the output is given by

$$l_k(t) = \sum_{i=0}^{\infty} \psi_k(t) x(t-i) \tag{25}$$

Hence we obtain

$$\begin{aligned} y_1(t) &= \sum_{k=1}^{N} c_k \left[\sum_{i=0}^{\infty} \psi_k(t) x(t-i) \right] \\ &= \sum_{k=1}^{N} c_k l_k(t) \end{aligned} \tag{26}$$

Similarly, for $y_2(t)$ we obtain

$$y_2(t) = \sum_{n=1}^{N} \sum_{m=1}^{N} c_{nm} l_n(t) l_m(t) \tag{27}$$

Implementing the nonlinear Laguerre model is easily achieved by means of a "nonlinear combiner" summing all weighted product term combinations, $l_i(t)l_j(t)$, $i = 1, .., N$, $j = 1, .., N$. Alshebeili et al [25] further introduce a matrix decomposition based implementation well suited for an efficient VLSI implementation.

This model structure overcomes to some extent, the usual problem of Volterra based models requiring a large number of Volterra kernels.

2.4 The Volterra Multiple Laguerre Model

The problem of slow convergence of the linear Laguerre model, in some cases, has been considered by Wahlberg [11]. He noted that convergence can be slowed as a result of (1) the plant poles being close to the unit circle, (2) the plant consists of resonant poles, and/or (3) the plant has widely dispersed poles. This problem can also be expected in the Volterra Laguerre model. One method of overcoming the last problem, is to introduce a Laguerre model with multiple poles [11]

$$H(z) = \sum_{j=1}^{P} \sum_{k=1}^{M} b_{jk} L_k(z, \alpha_j) \tag{28}$$

In this case, there are P Laguerre poles α_j, $j = 1, 2, \ldots, P$, all real variables, instead of a simple Laguerre pole.

2.5 The Volterra Kautz Model

It is simple to observe that a Volterra Kautz model can be introduced following the same approach as the Volterra Laguerre model above. In this case, however, $L_k(z, \alpha)$ in (12) is replaced by $\Psi_k(z, \alpha, \beta)$ from (18). The major difference in this case compared with the Volterra Laguerre model is that the input linear dynamic filter stage is now capable of more accurately approximating systems with complex poles.

Similarly, it is possible to reformulate the multiple pole Laguerre model in terms of Kautz filters to deal with systems having widely spaced complex poles.

3 Laguerre Neural Network Model

In recent years, with the growth of interest in neural networks, a major result which has emerged is that a multilayer perceptron (MLP), with a single hidden layer, is capable of approximating any smooth nonlinear input output mapping to an arbitrary degree of accuracy, provided that a sufficient number of hidden layer neurons is used [28]. This is often referred to as the universal approximation theorem. Since the nonlinearity in the Volterra orthogonal filter models is memoryless, it is simple to observe that this memoryless nonlinearity can be approximated using a MLP structure. However, as will be made clear in this and later sections, there are some differences in using a memoryless nonlinearity and the MLP approximations.

To follow the nomenclature used in the previous section, we will call this class of neural networks the MLP orthogonal filter model, and the individual MLP architectures as MLP Laguerre model, MLP Multiple Laguerre model, and the MLP Kautz model.

In addition, in the following sections, we will deal with mainly Laguerre based models, rather than Kautz based models. The exposition here can be easily extended to the Kautz based models or the multiple Laguerre based models. However, this will involve more complicated notation and in general, may distract from the main points that we wish convey in the paper. Hence we have chosen to describe the methods using a Laguerre based model instead of a Kautz based model.

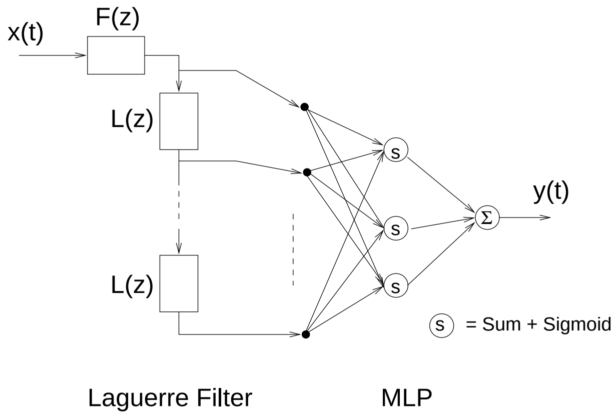


Figure 2: The MLP Laguerre model.

3.1 The MLP Laguerre model

A nonlinear model may be defined using a MLP with the Laguerre filters at the input stage. A MLP Laguerre model having L layers with $N_0, N_1, ..., N_L$ nodes per layer, is defined in the same manner as a usual MLP, with the input vector $\mathbf{z}^0(t)$ defined as

$$\mathbf{z}^0(t) \quad = \quad [L_1(z,\alpha)x(t), ..., L_M(z,\alpha)x(t)]' \tag{29}$$

where $x(t)$ is the input to the system, $L_i(z,\alpha)x(t)$ is the output from the ith Laguerre filter, $N_0 = M$ and

$$z_k^l(t) \quad = \quad g\left(\hat{x}_k^l(t)\right) \tag{30}$$

$$\hat{x}_k^l(t) \quad = \quad \sum_{i=1}^{N_{l-1}} w_{ki}^l z_i^{l-1}(t) \tag{31}$$

where $l = 1, 2, \ldots, L$; each neuron i in layer l has an output of $z_i^l(t)$; a layer consists of N_l neurons ($l = 0$ denotes the input layer, and $l = L$ denotes the output layer, $z_0^l = 1.0$ may be used as a bias); and $g(\cdot)$ is a sigmoid-type function typically evaluated as $\tanh(\cdot)$, and a synaptic connection between unit i in the previous layer and unit k in the current layer is represented by w_{ki}^l. The notation t is used to represent a discrete time instance. Note that this model is obtained by essentially substituting the memoryless nonlinearity in a nonlinear Laguerre model by the corresponding MLP structure with N_L layers. The MLP Laguerre model is shown in Figure 2. It is also possible to derive a more general architecture by defining subsequent layers as having Laguerre filters, in a manner similar to those indicated in [29]. Note that it is simple to define a MLP Multiple Laguerre model or a MLP Kautz model by modifying the input $\mathbf{z}^0(t)$ to the MLP appropriately.

3.2 Relationship between the Volterra orthogonal filter Models and MLP orthogonal filter Models

A fundamental question concerning the use of MLPs in place of the nonlinear combiner, is whether the MLP model is capable of the same representational structure as the Volterra-series type model. Wray and Green [30] have indicated that in some practical computer simulations, the sigmoid function is in fact approximated by a polynomial. Thus, one could consider a neural network model to be composed of sums of polynomials. In this section, we take a different approach, by considering the theoretical approximation of a Volterra series by a MLP model. It will be shown that a MLP can approximate arbitrarily closely a Volterra series expansion.

To demonstrate this result we make use of the following fact.

> From the universal approximation theorem, it is possible to find a MLP with a single hidden layer which will approximate the following nonlinear mapping: $F : x \longrightarrow x^2$ to an arbitrary degree of prescribed accuracy. For lack of a better name, we will call such an architecture a squarer.

From this fact, we have the following lemma.

Lemma 3.1 *The following mapping $f : (x_i, x_j) \mapsto x_i x_j$, where x_i, and x_j are inputs restricted to the space $E([0,1])$, $f(\cdot)$, a smooth mapping, can be approximated to an arbitrary degree of prescribed accuracy by using two squarers, and operations involving only additions and multiplications.*

Proof. This result follows essentially from Wiener [23]. The proof is by construction.

Prescribe two squarers, $F_i : x \mapsto x^2$ for $x \in E([0,1])$, $i = 0, 1$.

Construct a first layer $\mathbf{z} = W\mathbf{x}$, where $\mathbf{z} = [z_0 z_1]'$, $\mathbf{x} = [x_0 x_1]'$, and W is a 2×2 weight matrix with values $w_{00} = 1$, $w_{01} = 1$, $w_{10} = 1$, $w_{11} = -1$. Hence $z_0 = x_0 + x_1$, $z_1 = x_0 - x_1$. Pass $\{z_i : i = 1, 2\}$ through sub-networks F_0, F_1 respectively. Then $y_0 = F_0(z_0) = F_0(x_0 + x_1) \approx (x_0 + x_1)^2 = x_0^2 + 2x_0x_1 + x_1^2$; $y_1 = F_1(z_1) = F_1(x_0 - x_1) \approx (x_0 - x_1)^2 = x_0^2 - 2x_0x_1 + x_1^2$. Construct an output layer $v = \mathbf{b}'\mathbf{y}$, where v is a scalar, $\mathbf{y}' = [y_0 y_1]$, and $\mathbf{b}' = [0.25 \quad -0.25]$. Thus $v = x_0x_1 + \epsilon$ where ϵ is governed by the error in $\{F_i\}$ approximating $F : x \mapsto x^2$.

Remark 1. By repeatedly using this lemma, it is possible to obtain higher order polynomial terms in x_i, x_j, $\ldots$ x_k.

From Lemma 3.1 we can now prove the following theorem.

Theorem 3.1 *A SISO MLP Laguerre model, with input $x(t)$ and output $y(t)$ can approximate a finite order Volterra Laguerre model, to an arbitrary degree of prescribed accuracy, defined by*

$$y(t) = \sum_{k=1}^{N} y_k(t) \tag{32}$$

$$y_k(t) = \sum_{i_1=1}^{N} \cdots \sum_{i_k=1}^{N} c_{i_1 \cdots i_k} l_{i_1}(t) \cdots l_{i_k}(t) \tag{33}$$

where $l_j(t)$ is the output of the jth discrete Laguerre filter given by

$$l_j(t) = L_j(z,\alpha)x(t) = \frac{\sqrt{1-\alpha^2}}{z-\alpha}\left(\frac{1-\alpha z}{z-\alpha}\right)^{j-1} x(t) \tag{34}$$

Proof. The linear input stage is formed by approximating the Volterra kernels with orthonormal Laguerre functions as in Dumont and Fu [10]. The model is obtained by forming a weighted sum of power series terms of the outputs $\{l_i(t)\}$ from the Laguerre filters. From Lemma 3.1, the power series terms are easily obtained.

Remarks

1. In the proof of the theorem, we have not made explicit use of any information concerning the nature of the linear subsystem. Hence, the proof can be carried through to linear dynamical subsystem using other orthogonal filters, e.g., Kautz filters. Thus, the theorem is also true for linear dynamic subsystems involving other orthogonal filters.

2. Note that the proof requires an MLP having three hidden layers of units. However, it can be easily seen that since two of these layers are formed by linear summations, the approximation can be formed using an MLP with a single hidden layer. Hence, in general, one would expect a single layered MLP architecture for the MLP orthogonal filter model will be sufficient to approximate a Volterra orthogonal filter model to an arbitrary degree of accuracy.

3. An obvious question is whether the MLP structure offers any advantages over the Volterra model. It is known that a power series approximation has difficulties in modelling hard limiting [31]. It is observed in [31] that polynomial models exibit a pathological behaviour with inputs outside the space for which the approximation is defined[6]. The MLP does not appear to be exhibiting the same problems. Instead, a finite error is associated with any implementation using a MLP. In fact, it has been shown by a number of researchers using simulations that the MLP model can approximate "wild" nonlinearities, e.g., saw tooth nonlinearity. This indicates that the MLP orthogonal filter models potentially may be more useful for the modelling of nonlinear systems than the corresponding Volterra orthogonal filter models.

[6] There is a strict dependence between the input space and the mapping space in polynomial models such as the Volterra and Wiener series models [22]. In general, a smaller input space will result in a larger mapping space. This implies that the model identification process is governed by the choice of test inputs, to the extent that the inputs determine the class of systems to which the model is associated with [22].

4. The system described by (21) is also known as a Wiener system which is one of the block-oriented class of nonlinear systems [21, 19].

5. Suppose that we wish to obtain a better system approximation using more hidden units in the MLP Laguerre model. Can we obtain a higher degree of accuracy using an MLP Laguerre model than using a Volterra Laguerre while retaining a constant complexity in the linear input stage (see (21)) ? The Volterra Laguerre models has the order of the Laguerre input stage *directly coupled* with the order of the power series used to form the memoryless nonlinear mapping. As such, the Volterra Laguerre model appears to be best suited for 'mild' nonlinear functions [10, 19]. On the other hand, for the MLP Laguerre model, the complexity of the memoryless nonlinear mapping is determined by the MLP which is *decoupled* from the linear Laguerre input stage. This indicates that the MLP Laguerre model is a potentially more versatile structure than the Volterra Laguerre model.

4 Parameter estimation of the MLP Laguerre models

In the MLP Laguerre model, there are a number of parameters which need to be estimated from the input put data. These are: the Laguerre pole, α, and the order of the Laguerre filter; M, the parameters associated with the MLP section, viz., the constant weights, and the number of hidden layer neurons.

Since the MLP subsystem is the same as the classic MLP architecture, the parameter estimation techniques, e.g., backprop algorithm [32] can be applied directly. Hence, in the interests of space we will not consider in detail the training algorithm for the MLP subsystem.

4.1 Algorithms for Determining the Laguerre Pole

The Laguerre filter is parameterized by the Laguerre pole α, and the order M. Since the Laguerre filter is an orthogonal filter, the required degree of accuracy can be determined to some near optimal value by successively closer estimates to the optimal order of the model M, without such changes in the order affecting the existing model parameters $\{c_i\}$.

In the literature, it is apparent that most researchers favor a two stage approach, viz., first find the Laguerre pole, and then find the associated weights of the nonlinear combiner, based on this estimate of the Laguerre pole. The logic in using this two phase approach is that once the Laguerre pole is known, the parameter estimation problem for the associated weights is linear in parameters. Thus, any linear estimation technique could be employed, e.g. least mean square (LMS), recursive least squares (RLS), or Levinson-Durbin algorithms [33]).

There have been a number of approaches to the determination of the Laguerre pole, most of these require some a priori information on the signal to be approximated. Most methods can be grouped under the following headings: off-line determination and on-line determination methods.

In the category of off-line determination methods, for example, Clowes [16] described a method for finding the Laguerre pole by minimizing an integrated square error function of the impulse response[7]. This technique has since been extended to other cases, e.g., in Fu and Dumont [10]. On the other hand, Parks [7] proposed a method for determining the Laguerre pole based on two empirical observations of the signal to be approximated, viz., how fast the signal decays, and how smoothly it decays. This technique has been extended also by a number of researchers [34, 17].

There has been some discussion in the literature as to whether it is necessary to find the Laguerre pole accurately or not. One reason for this is the belief that because Laguerre functions form an orthonormal basis complete on $L_2[0, \infty)$, and hence can represent any function f for which $\int_a^b f^2(x)w(x)dx$ is square integrable, the value of the Laguerre time constant is not critical [35]. Fejzo found that, based on simulations, the modelling error of the Laguerre model is reasonably insensitive to the position of the Laguerre pole [24]. On the other hand, it has been found that the choice of the time constant can in fact have a significant impact on the performance of the model [11, 34].

Many of the above approaches for determining the Laguerre time constant are not so useful in the problem of system identification since they either require impulse response data and/or are performed off-line. There has been less work done in developing on-line algorithms for determining the Laguerre pole. The only work which we are aware of is the work by Lindskog and Wahlberg [13] (see also the work by Wahlberg and Hannan [36] where they derive a Levison-type recursion for computing the Laguerre pole in a linear model). They proposed that the poles in a Kautz model could be estimated by performing some preliminary analysis on the system, either by step response tests, or spectral analysis. Then an on-line non-linear Gauss-Newton search strategy could be used, though this does not guarantee convergence to an optimal solution. In the next section we present the derivation of on-line algorithms for estimating the Laguerre pole.

4.2 Derivation of On-line Algorithms for Laguerre Pole Estimation

Methods for determining the optimal coefficients in orthonormal functions have been studied over some time. This raises the question - "is it possible to modify the off-line methods for on-line application" ? In this case, we can identify two main approaches:

1. **Direct:** Adapt the off-line methods requiring step/impulse inputs to on-line methods which use some other input signals which can be taken into account in the formulation.

2. **Indirect:** In this case, a secondary system identification experiment is set up, such that the model is effectively determined off-line, with the resultant Laguerre time constant value being transferred back to the on-line model.

In the work presented here, in contra distinction to both the direct or indirect approaches, we are interested in adaptively computing estimates for the Laguerre time constant directly. This direct estimation of the Laguerre pole does not presuppose any information, e.g., impulse response, on the system under study.

[7] This is based on the assumption that an analytic expression of impulse response of a rational transfer function of the signal is available.

The algorithm derivation proceeds by defining the cost function $J = E[\frac{1}{2}e^2(t)]$, where $E[\cdot]$ is the expectation operator, and the output prediction error $e(t)$ is given by $e(t) = (y(t) - \hat{y}(t;\alpha))$, where $\hat{y}(t;\alpha)$ is the output of the MLP Laguerre model, and $y(t)$ is the desired output. To a first approximation, the cost function can be approximated by an instantaneous error function as $J_i(t) = \frac{1}{2}e^2(t)$. For completeness sake, we will describe briefly the algorithm derivation for the usual feedforward weights associated with the MLP even though these update algorithms are standard [32]. We will then concentrate on the derivation of the update equation for the Laguerre pole.

The Laguerre pole can be updated according to a steepest descent procedure, where we have

$$\begin{aligned} \hat{\alpha}(t+1) &= \hat{\alpha}(t) + \Delta\hat{\alpha}(t) \\ \Delta\hat{\alpha}(t) &= -\eta \nabla_\alpha J_i(t) \end{aligned} \tag{35}$$

where $\hat{\alpha}(t)$ is an estimate of α at time t. The weights in the MLP are adjusted in the same manner

$$w_{ik}^l(t+1) = w_{ik}^l(t) + \Delta w_{ik}^l(t) \tag{36}$$

$$\begin{aligned} \Delta w_{ikj}^l(t) &= -\eta \frac{\partial J_i(t)}{\partial w_{ik}^l(t)} \\ &= -\eta \frac{\partial J_i(t)}{\partial \hat{x}_k^l(t)} \frac{\partial \hat{x}_k^l(t)}{\partial w_{ik}^l} \\ &= \eta \delta_k^l(t) z_k^{l-1}(t) \end{aligned} \tag{37}$$

where

$$\begin{aligned} \delta_k^l(t) &= -\tfrac{\partial J_i(t)}{\partial \hat{x}_k^l(t)} \\ &= e_k(t) f'(\hat{x}_k^l(t)) && l = L \\ &= f'\left(\hat{x}_k^l(t)\right) \textstyle\sum_{p=1}^{N_{l+1}} \delta_p^{l+1}(t) w_{kp}^{l+1} && 1 \le l \le L-1 \end{aligned} \tag{38}$$

Hence, it is not too difficult to see that

$$\begin{aligned} \Delta\hat{\alpha}(t) &= \eta \sum_{j=1}^{M} \frac{\partial J_i(t)}{\partial l_j(t)} \frac{\partial l_j(t)}{\partial \alpha} \\ &= \eta \sum_{j=1}^{M} \beta_j(t) \delta_j^0(t) \end{aligned} \tag{39}$$

where $\delta_j^0(t)$ is the backpropagated error at the jth node of input layer, where $\delta_j^0(t) = -\frac{\partial J_i(t)}{l_j(t)}$. $\beta_i(t)$ is the first order sensitivity vector of the model operator parameters, defined as $\beta_j(t) = \frac{\partial l_j(t)}{\partial \alpha}$. Now, $l_j(t)$ can be expressed (see (15)) as follows:

$$l_j(t) = \begin{cases} \sqrt{1-\alpha^2}\, l_{j-1}(t-1) + \alpha l_j(t-1) & j = 1 \\ l_{j-1}(t-1) + \alpha\left(l_j(t-1) - l_{j-1}(t)\right) & j = 2, ..., M \end{cases} \tag{40}$$

Hence, β_j can be obtained by direct differentiation of eqn (40) as follows:

$$\beta_j(t) = \begin{cases} \frac{-\alpha}{\sqrt{1-\alpha^2}} l_{j-1}(t-1) + l_j(t-1) + \alpha\beta_j(t-1) & j = 1 \\ l_j(t-1) - l_{j-1}(t) - \alpha\beta_{j-1}(t) + \beta_{j-1}(t-1) + \alpha\beta_j(t-1) & j = 2, ..., M \end{cases} \tag{41}$$

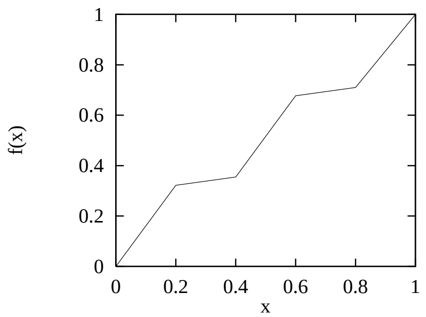


Figure 3: Piecewise nonlinear function $f(x)$ used in system model.

Hence, the algorithm for updating α is given in eqns (38),(39),(40),(41). This algorithm is a first order gradient algorithm. At present, little is known about the convergence properties for the learning algorithm.

More general updating equations are obtained by using the Gauss-Newton method [33].

Note that it is quite simple in concept to derive a corresponding set of online training algorithms for other MLP orthogonal filter models, e.g., the MLP Kautz model, except that the notation will be more complicated.

4.3 An Example

In this section, we will give a simulation example indicating the performance of the MLP Laguerre algorithm and the Volterra Laguerre model. The problem is to model a nonlinear system given in (21). Here, the linear section $H(z)$ is a first-order system given by

$$H(z) = \frac{0.5z^{-1}}{1-0.9z^{-1}} \tag{42}$$

The nonlinearity $f(x) : \mathcal{R}^1 \rightarrow \mathcal{R}^1$ is selected as shown in Figure 3. This is a nonlinearity which the Volterra series expansion might have some difficulties as indicated previously.

The input signal to the system consists of a white noise process $x(t)$ with $\mu = 0.1$, $\sigma = 0.1$, where μ and σ are the mean and standard deviation respectively[8]. The results shown here are for SISO models. In the case of the MLP Laguerre model, the order of the Laguerre structure was selected to be $M = 6$ while the MLP structure has 5 hidden units. The Volterra Laguerre model was selected to have up to second order nonlinearities, with a pre-selected Laguerre parameter of $\alpha = 0.75$.

To ensure stability in the estimation of α, it is important to choose a sufficiently small learning rate for this parameter. In addition, it is often found that a different learning rate for different sections

[8]This specifies the bounds on the normal operating range of the plant. For nonlinear models of this type, it is necessary to identify the plant over a specified range of input signal magnitudes, otherwise a different model would be learned for each different range of input signals (in constrast to linear systems where the magnitude of the input signal does not matter).

of the estimation process would be advantageous. Consequently, we use different learning rates for the Laguerre parameter ($\eta_\alpha = 0.0035$) and the MLP weights ($\eta_{MLP} = 0.05$). Apart from the usual stability constraints, these learning rates did not appear to be particularly sensitive. The Laguerre pole was not adapted at all for the Volterra Laguerre model. To select the remaining parameters for the Volterra Laguerre model, a least squares method [10] was used on 1000 data points.

Learning in neural networks, especially in recurrent neural networks, is often notoriously slow and difficult. The architecture we have proposed here requires careful attention to practical convergence issues such as finding suitable initial weights. Various methods may be devised to overcome these difficulties, including the application of various "tricks" to speed-up backpropagation convergence and constructive algorithms to initialize weights. In our case, we used a three-step method[9].

The training algorithm then becomes:

1. Determine initial parameters for the linear subsystem using usual methods.
2. Initialize the MLP weights by means of any suitable approaches.
3. Apply the on-line algorithm.

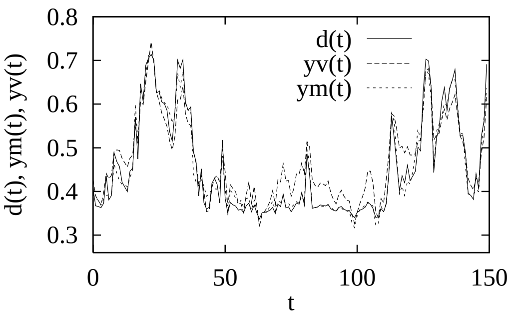
Various experiments were performed using a range of different starting weights. In Figure 4(a) results are shown for a simulation with $\alpha(0) = 0.5$ running for 10000 training iterations. This resulted in a normalized mean square error (NMSE) and variance of 0.0074 and 0.0005 respectively (computed over the last 500 data points) for the MLP Laguerre model. The Volterra Laguerre model obtained figures of 0.014 and 0.012 respectively. The convergence properties of the Laguerre α parameter are indicated in Figure 4(b) where 50000 training iterations were used ($\alpha(0) = 0$). For this example, it can be seen that the MLP Laguerre model approximates the system with a high degree of accuracy, while the Volterra Laguerre model on the other hand, does not approximate the system as well.

This simple example is given to indicate the performance of the model on a very straightforward problem. There are many practical issues which need to be resolved in order to successfully apply this architecture and on-line learning algorithm to a range of practical problems. Some of these are discussed in Section 5. Clearly, a detailed study of the practical performance of the model on various problems would be of value. Results of this nature will be published elsewhere.

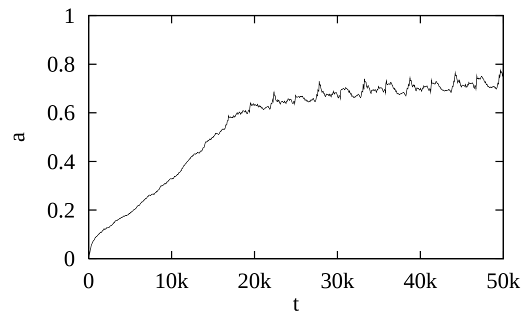
5 Open Issues

There are a number of issues that need to be resolved in order for these techniques to be accepted as providing solutions to a wide range of industrial problems. In this section, we highlight some of these problems and indicate areas requiring further research.

[9] This method assumes a model which has the general structure similar to (21). In situations where the plant structure is unknown, structural identification techniques together with techniques indicated in this paper would need to be considered. In this example, knowledge of the plant structure allows a slightly simpler model, viz., the reduced complexity model to be used. For more details concerning the reduced complexity model, see e.g. [26].



(a)



(b)

Figure 4: (a) Performance results for the MLP Laguerre model and Volterra Laguerre model used to approximate the first order nonlinear system described in the text showing desired and actual model outputs ($y_v(t)$ is the Volterra Laguerre output, $y_m(t)$ is the MLP Laguerre output). (b) Convergence properties of α in the MLP Laguerre model.

5.1 Effect of Laguerre Pole on Model Convergence

Wahlberg has examined the condition number of the data covariance matrix for the linear Laguerre model and proven that it has advantages over conventional parametric models [11]. However, there has been little consideration given to the problem of convergence characteristics of nonlinear models. A preliminary examination of these problems has been undertaken in [37].

5.2 Mean Square Output Error Surface

To be useful, gradient descent algorithms depend on several criteria. These include the shape of the cost function surface (typically mean square output error) versus model parameters and initial starting point. It is generally suspected that the MSOE surface for the MLP orthogonal filter model is multimodal. This could cause problems in the training of such models.

This area deserves further consideration, particularly methods for overcoming the multimodalities which may typically arise. In conjunction with learning algorithms, it is also possible that the problem of multimodal error surfaces may be overcome by introducing completely different approaches, e.g.,

1. The Hooker-Jeeves pattern search method has been used for determining the Laguerre pole by Zervos et. al. [38]. However this approach needs to be examined more closely for system identification experiments and compared with other optimization techniques.

2. Cost functions other than mean square output error, including regularization methods, may be of value in order to find improved on-line algorithms.

3. It is conjectured that more efficient algorithms could be derived by extending the off-line methods such that they do not require an exact knowledge of the system's characteristics. Instead, it may be possible to modify these approaches to give an estimate of the optimal Laguerre pole

and from there adapt the model pole on-line. Updating could proceed by periodic "checking" against this modified off-line (e.g. Parks-type) algorithm, thus allowing the gradient-descent algorithm to continue updating in suitable region of the plane.

4. Genetic algorithms may be useful in determining suitable regions within which the on-line algorithms could operate.

5.3 Model Structure Design

We term this section model structure design as opposed to the more well known issue of model structure selection, since there is a slightly different issue we wish to highlight. In some preliminary numerical experiments with the MLP orthogonal filter models and Volterra orthogonal filter models, we noted the problem of too few degrees of freedom in the model which may lead to a multimodal MSOE surface. However, too many degrees of freedom may lead to over-fitting. Thus, this calls for some criterion which will allow an "intelligent" determination of the number of parameters used in the model, finding an optimal trade-off in some sense between too few and too many parameters. To-date, no such criterion has been formulated for this class of models.

5.4 Sensitivity and Robustness Issues

In a real-world implementation of a system identification model, issues such as sensitivity to parameter error and system disturbances are of interest. Wahlberg [11] has considered these issues for linear Laguerre models. Wahlberg and Hannan have proposed the use of Laguerre functions as alternative discrete-time operators [39] in order to obtain better convergence properties [11]. It would be interesting to find out if the application of the MLP orthogonal filter model will lead to a reduction in sensitivity to parameter variations and system disturbances.

5.5 Adaptive Nonlinear Control Systems

Nonlinear adaptive control using neural network models is an area of significant recent interest. In view of the advantages the Laguerre based models present in both linear and nonlinear settings, it may be fruitful to consider the use of these architectures in nonlinear adaptive control. Some work in the area of using the Volterra Laguerre model in an adaptive control setting has been performed by Dumont and Fu [10]. Clearly it is of much interest to examine the issues of adaptive nonlinear control using these models, and the MLP Laguerre structure proposed here. A main impetus in applying the MLP orthogonal filter model to adaptive control lies in their efficient parametrization of the underlying nonlinear system. Thus, by using this nonlinear model, it may lead to more efficient adaptive control algorithms.

5.6 Other Orthonormal Basis Functions

The development of all the models discussed in this paper are based on the orthonormal Laguerre functions derived from the weighting function $w(x) = e^{-x}$ in the orthonormal basis function $\zeta_i(x)$ defined by (9).

Other orthonormal models which can be developed are well known and include Chebychev, Legendre etc. It would be interesting to compare the development of nonlinear models based on these orthonormal functions and compare their properties with that of the Laguerre models considered here. For example, given certain types of signals, is it possible to obtain a better model by choosing the orthonormal basis function appropriately ? Given the advantages of the Kautz models having complex poles, one would expect that this should be so.

6 Conclusions

The advantages of using orthonormal basis functions has been well established for some time, though recently there has been much interest in discrete-time Laguerre and Kautz models. The problem of nonlinear system modelling is an important one and so it is natural to ask how we may effectively use orthonormal functions for nonlinear system modelling. This problem has been considered previously using various approaches. In this paper, we have examined the use of discrete-time Laguerre functions for nonlinear system identification. In particular, we have considered the relationship between the Volterra orthogonal filter models and the MLP orthogonal filter models. It is shown that the MLP orthogonal filter model can approximate to an arbitrary degree of accuracy, a Volterra orthogonal filter model. In addition, it is argued that the MLP orthogonal filter model[10] is better in approximating a nonlinear system than the corresponding Volterra orthogonal filter model. An online training algorithm for the estimation of the Laguerre pole has been formulated. This method of estimating the Laguerre pole is different to those online algorithms discussed in the literature.

Clearly much remains to be done in developing a better understanding of how discrete-time orthonormal models can be used for nonlinear system identification. It is hoped that the present work will stimulate further investigation into this interesting field.

Acknowledgements

The first author acknowledges financial support from the Australian Research Council. The second author acknowledges partial support from the Australian Research Council. Our appreciation is expressed to Prof. Y. Fu for providing us with software for the Volterra Laguerre model. We would also like to thank the anonymous reviewers for helpful suggestions leading to a much improved paper.

[10] A model similar to this has apparently also been considered independently by Peng [40], referenced in [24], though this was not available to us at the time of writing. Although not discussed in this paper, there is a close relationship between Laguerre models and the recently developed Gamma model [41].

References

[1] Cowan CFN, Grant PM. Adaptive Filters. Prentice-Hall Englewood Cliffs, 1985.

[2] Young TY, Huggins WH. Representation of electrocardiogram by orthogonalized exponentials. IRE Int Conv Rec 1961; 9:145-153

[3] Lerner RM. The representation of signals. IRE Trans. Circuit Theory 1959; CT-6:197-216

[4] Tufts DW, Kumaresan R and Kirsteins I. Data adaptive signal estimation by singular value decomposition of a data matrix. Proc IEEE 1982; 70:684-685

[5] Kautz WH. Transient synthesis in the time domain. IRE Trans. Circuit Theory 1954; CT-1:23-39

[6] Lee YW. Statistical Theory of Communication. Wiley New York, 1960.

[7] Parks TW. Choice of time-scale in Laguerre approximations using signal measurements. IEEE Transactions on Automatic Control 1971; 511-513

[8] Clement PR. On completeness of basis functions for signal analysis. Siam Review 1963; 5:131-139

[9] Courant R, Hilbert D, Methods of Mathematical Physics. Interscience Publishers New York, 1953.

[10] Dumont GA, Fu Y. Non-linear adaptive control via Laguerre expansion of Volterra kernels. Int J Adaptive Control and Signal Proc 1993; 7:367-382

[11] Wahlberg B. System identification using Laguerre models. IEEE Trans. Automat. Control 1991; 36:551-562

[12] Wahlberg B. On the use of orthogonalized exponentials in system identification. Technical Report LiTH-ISY-I-1099, Dept of Elec. Eng., Linkoping University, Sweden, 1990.

[13] Lindskog P, Wahlberg B. Applications of Kautz models in system identification. Proc. IFAC 12th Triennial Conference, Sydney, Australia, pp. 41-44, 1993.

[14] Den Brinker AC. Adaptive orthonormal filters. Proc IFAC 1993; 5:287-292

[15] Nurges U. Laguerre models in problems of approximation and identification of discrete systems. Automation and Remote Control 1987, 48:346-352

[16] Clowes CJ. Choice of the time-scaling factor for linear system approximations using orthonormal Laguerre functions. IEEE Trans on Automat Control 1965; 487-489

[17] Masnadi-Shirazi MA, Ahmed N. Optimum Laguerre networks for a class of discrete-time systems. IEEE Trans Signal Processing 1991; 39:2104-2108

[18] Volterra V. Theory of Functionals and of Integral and Integro-Differential Equations. Blackie & Son Ltd Glasgow, 1930.

[19] Norton JP. Introduction to Identification. Academic Press London, 1986.

[20] Schetzen M. Nonlinear system modeling based on the Wiener theory. Proc IEEE 1981; 69:1557-1573

[21] Billings SA. A survey of nonlinear systems. IEE Proc 1980; Pt. D, 127:272-285

[22] Palm G. On representation and approximation of nonlinear systems. Biological Cybernetics 1978; 31:119-124

[23] Wiener N. Nonlinear Problems in Random Theory. The MIT Press New York, 1958.

[24] Fejzo Z, Adaptive Non-linear Laguerre-Based Filtering. PhD Thesis, Department of Electrical and Computer Engineering, Northeastern University, 1994.

[25] Alshebeili SA, Mertzios BG Venetsanopoulos AN. Efficient implementation of nonlinear Volterra systems via Laguerre orthogonal functions. In: Vanderwalle J, Boite R, Moonen M, Oosterlinck A (eds.) Signal Processing VI: Theory and Applications. Elsevier Science Publishers B.V. 1992; 685-688

[26] Back AD, Tsoi AC. An adaptive lattice architecture for dynamic multilayer perceptrons. Neural Computation 1992; 4:922-931

[27] Malin PS, Palaniswamy M. A partially correlated higher order recurrent neural network. Proc Third Aust Conf on Neural Networks, Canberra, pp. 45 - 48, 1992.

[28] White H. Artificial Neural Networks: Approximation and Learning Theory. Blackwell Cambridge Ma, 1992.

[29] Tsoi AC, Back AD. Locally recurrent globally feedforward networks, a critical review of architectures. IEEE Trans Neural Networks 1994; 5:229-239

[30] Wray J, Green GGR. Neural networks, approximation theory and finite precision computation. Neural Networks 1995; 8:31-37

[31] Matthews MB, Moschytz GS. The identification of nonlinear discrete-time fading-memory systems using neural network models. IEEE Trans Circuits and Systems-II 1994; 41:740-751

[32] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE and McClelland JL (eds). Parallel Distributed Processing, Vol. 1. Foundations. The MIT Press Cambridge MA, 1986.

[33] Ljung L, Söderström T. Theory and Practice of Recursive Identification. The MIT Press Cambridge Massachusetts, 1983

[34] Fu Y, Dumont GA. An optimum time scale for discrete Laguerre Network. IEEE Trans Automat Control 1993; 38:934-938

[35] Olivier PD. Online system identification using Laguerre series. IEE Proc.-Control Theory Appl. 1994; 141:249-254

[36] Wahlberg B, Hannan EJ. Parametric signal modelling using Laguerre filters. The Annals of Applied Probability 1993; 3:467-496

[37] Back AD, Tsoi AC. Constrained pole-zero filters as discrete-time operators for system approximation. In: Girosi F, Makhoul J, Manolakos E and Wilson E (eds). Neural Networks for Signal Processing V, IEEE Press New York, 1995, pp. 191-200

[38] Zervos C, Belanger PR, Dumont GA. On PID controller tuning using orthogonal series identification. Automatica 1988; 24:165-175

[39] Back AD, Tsoi AC, A comparison of discrete-time operator models for nonlinear system identification. In: G. Tesauro, D. S. Touretzky and T. K. Leen (eds.) Advances in Neural Information Processing Systems 7, MIT Press, Cambridge MA, 1995, pp. 883-890

[40] Peng M. PhD Thesis. CDSP Center, Northeastern University Boston MA, 1994.

[41] Olivier e Silva T. On the equivalence between Gamma and Laguerre filters. Proc Int Conf Acoust Speech Signal Proc. 1993; 3, pp. 109-112

Notation

a_i	Filter parameter (feedback)
α	Laguerre pole
$\hat{\alpha}$	Estimate of Laguerre pole
b_i	Filter parameter (feedforward)
$\beta_i(t)$	First order sensitivity vector of the model parameters
c_i	Volterra Laguerre model parameter
$\delta_j^0(t)$	Backpropagated error at the jth node of input layer
$G(z, \theta)$	Linear transfer function
J	Cost function
$J_i(t)$	Instantaneous cost function at time t
$L_i(s, \alpha)$	Laplace transform of the Laguerre function
$L_i(z, \alpha)$	Discrete Laguerre function
$l_i(t)$	Output of the ith discrete Laguerre filter
$\psi_k(\tau)$	kth order continuous-time Laguerre function
$\Psi_k(z, \alpha, \beta)$	Discrete Kautz function