ADAPTIVE NONLINEAR WIENER-LAGUERRE-LATTICE MODELS

Zoran Fejzo

Aware, Inc., One Memorial Drive
Cambridge, MA
zfejzo@aware.com

Hanoch Lev-Ari

ECE Department, Northeastern University
Boston, MA
levari@cdsp.neu.edu

ABSTRACT

In this work we develop an adaptive nonlinear estimation technique, polynomial model-based, that has guaranteed stability and makes parsimonious use of coefficients. Our approach to the development of reduced-complexity adaptive nonlinear filters is based on a combination of: (a) The Wiener model of nonlinear systems (both FIR and IIR) and its application to nonlinear estimation from white Gaussian signals; (b) Wiener's notion of fixed (Laguerre) pre-orthogonalization, which we have extended to include adaptive pre-orthogonalization with respect to arbitrary (non-white and non-Gaussian) input signals [2]; (c) Efficient implementation of memoryless nonlinear maps for uncorrelated inputs based on (Hermite) orthogonal polynomials; (d) Application of suitably modified RLS and LMS adaptation techniques to determine the coefficients of such nonlinear maps.

1. INTRODUCTION

Adaptive nonlinear filtering has been dominated by the polynomial-based (Volterra, binomial, etc.) approach, in part because: (i) this approach uses a rather general model of nonlinearity, and (ii) polynomial models are *linear* in their coefficients. By varying the structural indices, the length of memory and the degree of nonlinearity of a polynomial model, one can approximate a large class of nonlinear systems.

The most commonly used approach in adaptive nonlinear filtering is based on the truncated Volterra model due to its guaranteed stability. It exploits the linearity of the truncated Volterra model with respect to its coefficients, to transform adaptive *single-channel nonlinear* estimation problems into adaptive *multichannel linear* estimation problems. The computational complexity of such schemes is either linear in N, the number of coefficients, or quadratic in N. For instance, the adaptation of a quadratic nonlinearity, which involves $N \sim M^2$ (M denotes the length of memory) coefficients, requires $\mathcal{O}(M^2)$ computations for LMS-type schemes and $\mathcal{O}(M^4)$ for RLS-type schemes. By exploiting the (partial) shift structure of the multichannel embedding, one can obtain "fast" RLS-type schemes that require $\mathcal{O}(M^3)$ computations [3]. This high computational cost is the main drawback of adaptive Volterra filters.

This work was supported in part by the National Science Foundation under Grant MIP-9220182

In our work we develop an adaptive nonlinear estimation technique polynomial model-based that has guaranteed stability and makes parsimonious use of coefficients. Our approach to the development of reduced-complexity adaptive nonlinear filters is based on a combination of:

- The Wiener model (Fig. 1) of nonlinear systems (both FIR and IIR) and its application to nonlinear estimation from white Gaussian signals.
- Wiener's notion of fixed (Laguerre) pre-orthogonalization, which we have extended to include adaptive pre-orthogonalization with respect to an arbitrary (non-white and non-Gaussian) input signals [2].
- Efficient implementation of memoryless nonlinear maps for uncorrelated inputs based on (Hermite) orthogonal polynomials.
- Application of suitably modified RLS and LMS adaptation techniques to determine the coefficients of such nonlinear maps.

To be more specific, we construct efficient adaptive implementation of the Wiener model of non-linear systems as a polynomial model consisting of a *Linear-Dynamic* (LD) module and *Nonlinear-Memoryless-Readout* (NMR) module (Fig. 1). The use of a pole-zero (IIR) configuration to represent the LD module can provide significant reduction in implementation complexity compared to Volterra (FIR) models. Moreover, having linear dynamics makes it much easier to ensure the stability of the overall model (compared to models based on non-linear difference equation [3], for example) . This advantage is even more important in adaptive implementations where the model parameters vary in time.

The single-input/multiple-output LD module $\mathbf{F}(z)$ can be interpreted as a *filter bank*. If this filter bank is orthogonal it maps a white input signal $x(n)$ into white and uncorrelated output signals $\xi_0(n), \cdots, \xi_M(n)$. In particular, if the input signal is Gaussian (and white) then the outputs $\xi_0(n), \cdots, \xi_M(n)$ are independent of each other (and white). Wiener suggested using a Laguerre based filter bank, because this makes it possible to implement the entire filter bank as a cascade of identical first-order all-pass filter sections preceded by a single low-pass section. However, if the input is stationary Gaussian but colored (not white), the filter bank has to be orthogonal with respect to the power spectrum of the input signal $S_x(e^{j2\pi f})$. Since often $S_x(e^{j2\pi f})$ is not known a priori, adaptive orthogonalization is needed. We refer to the configuration of an LD module

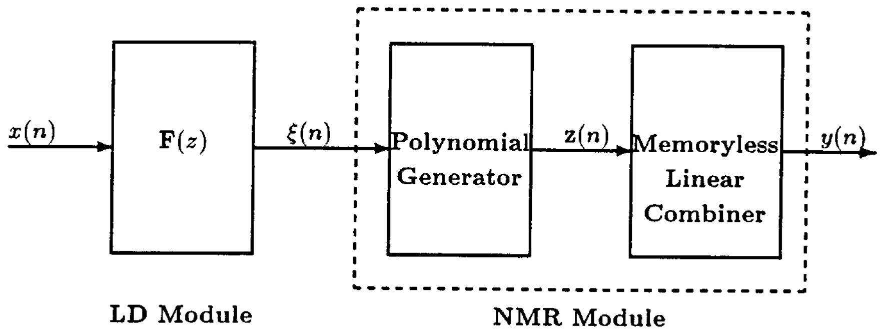

Figure 1: The Wiener Model of Non-linear Systems

that produces uncorrelated outputs (for arbitrary input signals) as a linear-dynamic-uncorrelated (LDU) module.

Since orthogonalization is a linear operation, the outputs of the LDU with a Gaussian input $x(n)$, are jointly Gaussian, and thus they are independent. Consequently when we use Hermitian polynomials in the Polynomial Generator, its outputs $z_1(n)\cdots z_{N_u}(n)$ are also uncorrelated. Now since the signals $z_1(n)\cdots z_{N_u}(n)$ are uncorrelated the coefficients of the memoryless linear combiner can be determined independently for each z_i so that the adaptation of these coefficients can be carried out, for instance, in an order recursive fashion using a modified gradient adaptive lattice algorithm (which we call the ORGA algorithm [1]).

For correlated signals $z_1(n)\cdots z_{N_u}(n)$ the use of the ORGA algorithm for adaptation of the NMR module may result in an unacceptable misadjustment error. As an alternative we consider two adaptive mechanisms:

- stochastic gradient adaptation, using a version of the normalized LMS algorithm with time varying step size, or
- least squares technique, using the RLS algorithm

2. ADAPTIVE WIENER-LAGUERRE-HERMITE MODEL

As discussed in the introduction we opt to use the LDU/NMR model structure consisting of a Laguerre-lattice LDU module and a Hermitian NMR module (Hermite Polynomial Generator + Adaptive Linear Combiner).

The adaptation of the LDU module is affected only by the statistics of the input signal $x(n)$ and is independent of the statistics relating the desired output signal $d(n)$ to the input signal $x(n)$. Therefore, there is no global feedback between the modules and we can consider the adaptation of each one of the three modules separately.

2.1. Adaptive LDU Module

For the LDU module we use the NGALL (normalized gradient adaptive Laguerre-lattice) algorithm [2] that produces normalized outputs $b_i^{(N)}(n)$. These outputs represent the variance-normalized Laguerre-lattice (backward) residuals and their variance is [2]

$$E|b_i^{(N)}(n)|^2 \approx \frac{1-\lambda}{1-\lambda^{n+1}} := \frac{1}{\alpha(n)}$$

which converges very slowly to its non-unity steady state value of $1-\lambda$; it takes about $\frac{1}{1-\lambda}$ iterations to reach 90% of this steady-state value.

While the slow convergence of the variance of $b_i^{(N)}(n)$ has no effect on the dynamics of the NGALL algorithm itself, it has an adverse effect on the performance of the Hermite polynomial generator. Consequently the outputs of the NGALL algorithm need to be rescaled, viz., $\xi_i(n) = \sqrt{\alpha(n)}b_i^{(N)}(n)$ so that $E|\xi_i(n)|^2 \approx 1$ for all n. This produces unit-variance inputs to the Hermite polynomial generator module.

2.2. Hermite Polynomial Generator

The Hermite polynomials satisfy the recursive relation [4]

$$H_{P+1}(X) = XH_P(X) - \sigma_X^2 PH_{P-1}(X) \qquad (1)$$

with $H_0(X) = 1$. Thus the outputs of the Hermite polynomial generator are products of terms of the form $H_p(\xi_l(n))$, viz.,

$$\hat{d} = \sum_{\substack{k_0,k_1,\ldots,k_M \\ k_i\in[0,\,L],\quad 0\le\sum k_i\le L}} \eta_{k_0,k_1,\ldots,k_M}\, H_{k_0}(\xi_0)H_{k_1}(\xi_1)\ \ldots\ H_{k_M}(\xi_M)$$

Since the outputs of the LDU module are normalized i.e., $E\mid\xi_i(n)\mid^2 = 1$, it follows from (1) that the coefficients of the Hermite polynomials are constant. However the outputs of the Hermite polynomial generator are not normalized because $E\mid H_P(\xi_i(n)\mid^2 = E|\xi_i(n)|^{2P}P! = \sigma_X^{2P}P!$ (for Gaussian signals). This can be remedied by scaling the Hermite polynomial of order P by $\frac{1}{\sigma_X^P\sqrt{P!}}$, viz., $\bar{H}_P(X) := \dfrac{H_P(X)}{\sigma_X^P\sqrt{P!}}$ so that $E|\bar{H}_P(X)|^2 = 1$ for all P and for a Gaussian input signal. Such normalized Hermite polynomials satisfy the recursive relation

$$\bar{H}_{P+1}(X) = \frac{1}{\sqrt{P+1}}\left[\frac{X}{\sigma_X}\bar{H}_P(X) - \sqrt{P}\bar{H}_{P-1}(X)\right].$$

In practice the variance of the scaled residuals $\xi_i(n) = \sqrt{\alpha(n)}b_i^{(N)}(n)$ fluctuates around unity, so that the decorrelation achieved by the Hermite polynomial generator is less than perfect. However, this slight imperfection has a negligible effect on the overall performance of the adaptive algorithm.

2.3. Adaptive Linear Combiner

The outputs $z_1(n), \cdots, z_{N_u}(n)$ from the Hermite Polynomial Generator submodule span the linear subspace from which we want to estimate the desired signal $d(n)$. This means that the estimate $\hat{d}(n)$ is obtained as a linear combination of $z_i(n)$, viz.,

$$\hat{d}(n) = [w_1(n) \cdots w_{N_u}(n)] \left[z_1(n) \cdots z_{N_u(n)}\right]^T = \mathbf{w}(n)\mathbf{z}(n)$$

and the coefficients of the adaptive linear combiner $w_i(n)$ can be determined using standard techniques developed for adaptive linear filters. All of these techniques attempt to determine the parameters $\{w_i\}_{i=1}^{N_u}$ that minimize the global estimation error

$$J = E\{\mid d(n) - \hat{d}(n) \mid^2\} = E\{\mid d(n) - \sum_{i=1}^{N_u} w_i z_i(n) \mid^2\} \,. \quad (2)$$

The techniques vary in the assumptions made about the degree of statistical correlation between the variables $\{z_i(n)\}_{i=1}^{N_u}$, and in the method of estimating the various probabilistic moments needed to determine the optimal solution for (2).

The most conservative technique is the Recursive Modified Gram-Schmidt (RMGS), which makes no assumptions on the correlation between the variables $z_i(n)$. Using a deterministic (exponentially-weighted) estimate of the correlation matrix $R = E\{\mathbf{z}(n)\mathbf{z}^*(n)\}$, the RMGS produce a decorrelated equivalent of $\mathbf{z}(n)$ which is then used to form an estimate of $d(n)$. The main disadvantage of this method is its high computational cost: $\mathcal{O}(N_u^2)$ operations per a single time-instant are required. We consider two alternative realizations of the adaptive linear combiner submodule:

- assuming that the outputs from the Hermite Polynomial Generator are uncorrelated, the coefficients of a linear combiner can be adapted in an order recursive fashion using gradient adaptive type (ORGA [1]) algorithm
- if decorrelation between variables $z_i(n)$ cannot be assumed, we can still use a stochastic gradient approach (NLMS with time varying step size [1]) for adaptation of the linear combiner coefficients.

Both approaches achieve reduced complexity ($\mathcal{O}(N_u)$ operations per a time-instant) at the cost of a modest increase in the steady-state error.

The outputs of the Hermite polynomial generator are perfectly decorrelated only when the outputs of the LDU module are: (i) Gaussian, and (ii) uncorrelated. While the first condition (Gaussianity) may be occasionally met, the second condition is never perfectly met as the consequence of the ongoing adaptation in the self-orthogonalizing LDU module. Consequently, using the ORGA algorithm for adaptation of the linear combiner results in an excess error J_{ex}. An upper bound on this error can be expressed in terms of the eigenvalues (λ_i) of the data correlation matrix ($\mathbf{R} = E[\mathbf{z}(n)\mathbf{z}^*(n)]$) as follows

$$J_{ex} \leq N_u(\sigma_d^2 - J_{min}) \max_i \left(\frac{1}{\lambda_i} + \lambda_i - 2\right)$$

where $\sigma_d^2 = E|d(n)|^2$. In particular, if the $z_i(\cdot)$ are indeed uncorrelated (so that $\mathbf{R} = \mathbf{I}$), then the upper bound vanishes (because $\lambda_i = 1$ for all i) and $J_{ex} = 0$, as expected.

3. SIMULATION RESULTS

We compare the performance of adaptive Wiener-Laguerre and adaptive Volterra, second order non-linear models by computer simulation. The RLS algorithm, which in this case requires ($\mathcal{O}(M^3)$) computations, is used to adapt the second order Volterra model (VT/RLS) coefficients. For the adaptation of Wiener-Laguerre model coefficients RLS (WLT/RLS), ORGA (WLL/ORGA) and NLMS (WLL/NLMS) algorithms are used.

The performance of the various adaptive non-linear models is evaluated in several experiments, in which the adaptive filters are used in a system identification scenario. The system to be identified is a second order non-linear system as shown in Fig. 2(a), where $H(z)$ is a linear

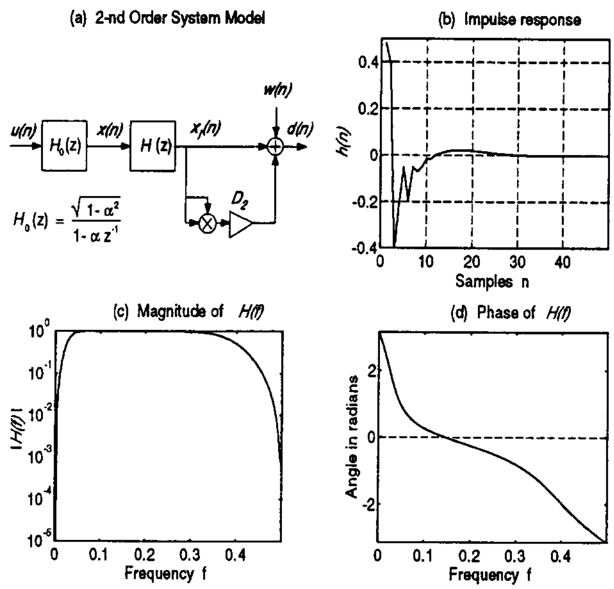

Figure 2: Unknown System Used in Experiments

4-th order band-pass Butterworth filter with frequency characteristics and impulse response as presented in Fig. 2(b), (c), (d). The amount of non-linear distortion introduced by our model is characterized in terms of the parameter $D_2 = 10\log_{10}\frac{\sum_k |x_l(k)|^2}{\sum_k |x_l^2(k)|^2}$ $[dB]$. We use $M = 10$, $a = 0.5$ and $P = 2$, for the adaptive Wiener-Laguerre model while $M = 22$ and $P = 2$ are used with the adaptive Volterra model.

The input signal $x(n)$ to the unknown system is a unit variance AR(1) process obtained by filtering zero-mean white noise process $u(n)$ through the filter $H_0(z)$ with $\alpha = 0.8$. The distribution of the white noise signal $u(n)$ is selected from the family of double sided generalized Gamma distributions where we control the kurtosis $\gamma_4 = \frac{E|u(n)|^4}{[E|u(n)|^2]^2}$ of the distribution, while maintaining a unit variance. The desired signal $d(n)$ is obtained by adding zero-mean white Gaussian noise $w(n)$ to the output of the unknown system. The noise $w(n)$ is independent of the input signal $x(n)$. Experiments are performed with a signal-to-measurement-noise ratio of $SNR_i = 20\,dB$.

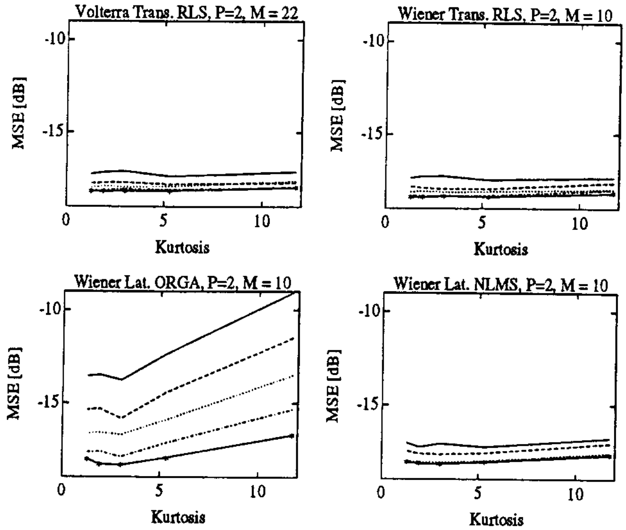

Figure 3: Comparison between 4 adaptive models while varying kurtosis and D_2; $SNR_i = 20\ dB$, Solid line ($D_2 = 0\ dB$), Dashed line ($D_2 = 4\ dB$), Dotted line ($D_2 = 7\ dB$), Dash-dotted line ($D_2 = 10\ dB$) and Marked line (*) ($D_2 = 13\ dB$).

The results presented are ensemble averages of 50 independent runs of 2000 samples each. Performance evaluation is carried out by plotting the ensemble average of the squared modeling error (MSE) and computing the steady state error (MMSE). Based on the simulation results we observe that:

• The Laguerre models perform well even without exact knowledge of the optimal Laguerre parameter a; for $a \in [0.35 - 0.6]$ ($a_{opt} = 0.5$) the excess error is always less then $1.5dB$.
• The Wiener Laguerre model with $M = 10$ and $a = 0.5$ has similar modeling capabilities to that of the Volterra model with $M = 22$.
• The models with RLS adaptation exhibit a robust behavior with respect to changes in signal statistics (in particular kurtosis) as well as changes in the amount of second order non-linear distortion D_2 (Fig. 3).
• The WLL/NLMS model is comparable to the VT/RLS and WLT/RLS models in terms of steady-state error and robustness with respect to kurtosis and D_2.
• The WLL/ORGA model is more sensitive to changes in kurtosis and level of non-linear distortion. Its steady-state performance is comparable to the other models only for a narrow range of kurtosis and D_2.
• The WLT/RLS model and the WLL/ORGA model exhibit the fastest convergence rate (Fig. 4).

	M	(MULT. + DIV.)/ITER.
VT/RLS	22	$\mathcal{O}(5M^3)$ 6.235 10^4
WLT/RLS	10	$\mathcal{O}(\frac{3}{8}M^4)$ 6.619 10^3
WLL/ORGA	10	$\mathcal{O}(4M^2)$ 6.226 10^2
WLL/NLMS	10	$\mathcal{O}(\frac{3}{2}M^2)$ 2.943 10^2

Table 1: Comparison of Computational Complexity for Quadratic Models

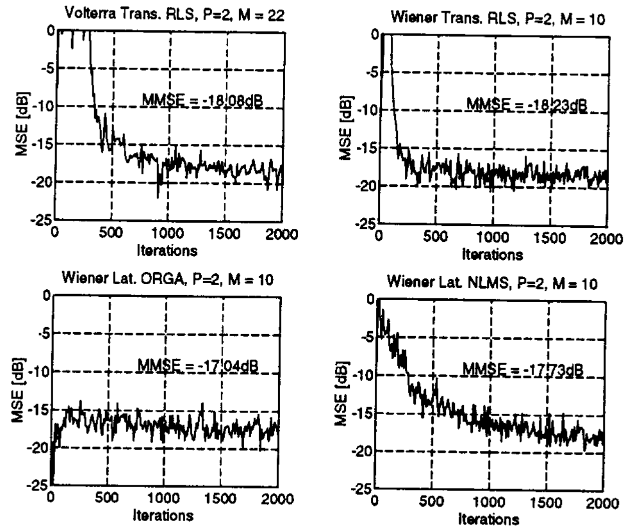

Figure 4: Comparison between 4 adaptive models for a non-Gaussian input ($\gamma_4 = 5.4$), $SNR_i = 20\ dB$ and $D_2 = 10\ dB$.

A more detailed evaluation of the computational requirements is given (Table 1) by the number of multiplications and divisions per time iteration executed by each of the above mentioned algorithms.

4. CONCLUSION

In this work we have developed adaptive non-linear estimation technique that use the Laguerre filter bank to possibly reduce the structural index M of the non-linear model. This approach offers an alternative performance-complexity trade-off. In particular for a quadratic non-linear model: (i) the WLT/RLS has excellent steady state performance and fast initial convergence at the cost of $\mathcal{O}(M^4)$ computations per time iteration; (ii) the WLL/ORGA exhibits excellent convergence behavior and requires only $\mathcal{O}(M^2)$ computations however, it introduces a moderate steady state performance degradation; (iii) WLL/NLMS performs only $\mathcal{O}(M^2)$ computations per time instant and has very good steady state performance but with somewhat slower initial convergence.

5. REFERENCES

[1] Z. Fejzo and H. Lev-Ari. *Adaptive Non-linear Filtering with Laguerre-Lattice Orthogonalisation.* CDSP Technical Report TR-CDSP-94-24, Northeastern University, Boston, MA, June 1994.

[2] Z. Fejzo and H. Lev-Ari. *Adaptive Laguerre-Lattice Filters.* CDSP Technical Report TR-CDSP-93-21, Northeastern University, Boston, MA, September 1993. Also, submitted to IEEE Trans. on Signal Processing.

[3] V.J. Mathews. "Adaptive Polynomial Filters". *IEEE Signal Processing Magazine*, pages 10–26, July 1991.

[4] M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems.* Robert E. Krieger Publishing Company, Malabar,Florida, 1989.