ADAPTIVE VOLTERRA FILTERS USING ORTHOGONAL STRUCTURES

V. John Mathews

Department of Electrical Engineering
University of Utah
Salt Lake City, Utah 84112, USA

ABSTRACT

This paper presents an adaptive Volterra filter that employs a recently developed orthogonalization procedure of Gaussian signals for Volterra system identification. The algorithm is capable of handling arbitrary orders of nonlinearity P as well as arbitrary lengths of memory N for the system model. The adaptive filter consists of a linear lattice predictor of order N, a set of Gram-Schmidt orthogonalizers for N vectors of size $P+1$ elements each, and a joint process estimator in which each coefficient is adapted individually. The complexity of implementing this adaptive filter is comparable to the complexity of the system model when N is much larger than P, a condition that is true in many practical situations. Experimental results demonstrating the capabilities of the algorithm are also presented in the paper.

1. INTRODUCTION

Truncated Volterra series models have become very popular in adaptive nonlinear filtering applications [6], [9]. Several stochastic gradient (SG) and recursive least-squares (RLS) adaptive Volterra filters have been developed in the last fifteen years or so [1], [4], [5], [8], [10]. The stochastic gradient algorithms are, in general, easy to derive and implement. However, they show slow and input-signal-dependent convergence characteristics. The recursive least-squares algorithms, on the other hand, exhibit fast convergence characteristics that are more or less independent of the input signal statistics. However, unlike their linear counterparts, even the most efficient RLS Volterra filters have significantly larger computational complexity than the SG Volterra filters.

One approach to improving the convergence characteristics of the stochastic gradient adaptive filters is to employ structures that orthogonalize the input signal. Unfortunately, the lattice realizations of Volterra systems for arbitrary inputs are over-parameterized [10]. For example, the lattice realization of a second-order Volterra system with N-sample memory requires $O(N^3)$ parameters even though the system model itself has only $O(N^2)$ parameters. Consequently, SG adaptive filters employing such structures will have computational complexity that is comparable to the RLS algorithms. Thus, there is a real need for developing stochastic gradient adaptive filters with better convergence properties than currently available techniques without exorbitant increases in the computational complexity. This paper presents an approach to developing such algorithms using a recently developed method for orthogonalizing Gaussian input signals for Volterra system identification problems [7].

The rest of the paper is organized as follows. The next section describes the method for orthogonalizing Gaussian signals for Volterra system identification problems. An adaptive filtering algorithm that makes use of the structure derived from the orthogonalization approach is presented in Section 3. Experimental results are presented in Section 4. The concluding remarks are made in the last section.

2. ORTHOGONALIZATION OF GAUSSIAN SIGNALS FOR VOLTERRA SYSTEM IDENTIFICATION

Consider a finite-memory and finite order Volterra system represented by the input-output relationship:

$$y(n) = h_0 + \sum_{p=1}^{P} \overline{h}_p[x(n)], \tag{1}$$

where $x(n)$ is the input signal to the system, $y(n)$ is the output of the system, and

$$\overline{h}_p[x(n)] = \sum_{m_1=0}^{N-1} \sum_{m_2=m_1}^{N-1} \cdots \sum_{m_p=m_{p-1}}^{N-1} h_p(m_1, m_2, \cdots, m_p) \cdot x(n-m_1)x(n-m_2)\cdots x(n-m_p). \tag{2}$$

The above model incorporates the kernel symmetry without any loss of generality. Also, the upper limit in the summations above are all identical only for convenience. The methodology for arbitrary upper limits is identical to the one presented in the paper. The symmetric form coefficients of the expression in (1) can be uniquely estimated under some very mild conditions on the input signal.

It is convenient to represent the system of (1) using vector notations for our derivations. Let

$$\overline{h_p}[x(n)] = \mathbf{X}_{pr}^T(n)\mathbf{H}_{pr} \tag{3}$$

This work was supported in part by the National Science Foundation under Grant MIP-8922146 and an IBM Departmental Grant.

0-7803-2431-5/95 $4.00 © 1995 IEEE

where $\mathbf{X}_{pr}(n)$ and $\mathbf{H}_{pr}$ represent the vectors containing all pth order products of the input signal appearing in (2) and the corresponding coefficients, respectively. Let us also define the input and coefficient vectors as

$$\mathbf{X}(n) = [1, \mathbf{X}_{1r}^T(n), \mathbf{X}_{2r}^T(n), \cdots, \mathbf{X}_{Pr}^T(n)]^T \qquad (4)$$

and

$$\mathbf{H} = [h_0, \mathbf{H}_{1r}^T, \mathbf{H}_{2r}^T, \cdots, \mathbf{H}_{Pr}^T]^T, \qquad (5)$$

respectively. Now, we can rewrite (1) in a very compact form as

$$y(n) = \mathbf{X}^T(n)\mathbf{H}. \qquad (6)$$

The basic problem considered in this section is the orthogonalization of the elements of the input vector $\mathbf{X}(n)$ in (4). The orthogonality is in the minimum mean-square error sense. We will assume that the input signal is Gaussian and has zero mean value. The assumption that the input signal has zero mean value is not restrictive in any way since the mean value can be removed from any signal and the bias term h_0 in (1) can account for any contribution from the non-zero mean value of the input signal.

In order to derive the orthogonalizer, let us define a smaller input vector

$$\mathbf{X}_L(n) = [x(n), x(n-1), \cdots, x(n-N+1)]^T, \qquad (7)$$

which consists only of the linear components in the input signal set in (4). Using a lattice predictor [2], find an orthonormal basis set for the elements of $\mathbf{X}_L(n)$. Let $u_i(n)$; $i = 1, 2, \cdots, N$ represent the orthogonal basis signals generated by the linear lattice predictor. Then

$$E\{u_i(n)u_j(n)\} = \delta(i-j), \qquad (8)$$

where $\delta(n)$ represents the Dirac delta function.

The elements of the set $\{u_1(n), u_2(n), \cdots, u_N(n)\}$ are Gaussian, zero-mean and uncorrelated with each other. Since all of them have unit variance, they also have identical distribution functions. Furthermore, since uncorrelated Gaussian processes are also independent processes, $u_1(n)$, $u_2(n)$, $\cdots$, $u_N(n)$ are mutually independent random processes. In particular,

$$E\{f(u_i(n))g(u_l(n))\} = E\{f(u_i(n))\}E\{g(u_l(n))\} \qquad (9)$$

whenever $i \neq l$.

Now, let us define a vector $\mathbf{U}_{P,i}(n)$ as

$$\mathbf{U}_{P,i}(n) = [1,\ u_i(n),\ u_i^2(n),\ \cdots,\ u_i^P(n)]^T. \qquad (10)$$

Let $\mathbf{Q}_P$ be a lower triangular, $(P+1)\times(P+1)$ element matrix that orthogonalize $\mathbf{U}_{P,i}(n)$. Since all $u_i(n)$'s have identical distributions, the same $\mathbf{Q}_P$ will orthogonalize $\mathbf{U}_{P,i}(n)$ for all values of i. Furthermore, since the statistics of $\mathbf{U}_{P,i}(n)$ are known, we can pre-compute $\mathbf{Q}_P$. An example of calculating $\mathbf{Q}_P$ is given in [7].

Let $\mathbf{V}_{P,i}$ be an orthogonalized vector obtained as

$$\mathbf{V}_{P,i} = \mathbf{Q}_P\mathbf{U}_{P,i}. \qquad (11)$$

Let $v_{P,i,j}$ denote the jth element of $V_{P,i}$.

Theorem 1.

$$\{v_{P,1,m_1}(n)v_{P,2,m_2}(n)\cdots v_{P,N,m_N}(n) \;\Big|\; m_1 + m_2 + \cdots + m_N \le P\}$$

is an orthogonal basis set for

$$\{x^{m_1}(n)x^{m_2}(n-1)x^{m_3}(n-2)\cdots x^{m_N}(n-N+1) \;\Big|\; m_1 + m_2 + \cdots m_N \le P\}.$$

Note that $v_{P,i,0}(n) = 1$ for all i and that each m_i takes values from $0 \le m_i \le P$.

Proof. A proof for this theorem may be found in [7]. It should be noted that the lattice structure for second-order Volterra filters that was presented in [3] is a special case of the above procedure.

3. AN EFFICIENT ADAPTIVE LATTICE VOLTERRA FILTER FOR GAUSSIAN SIGNALS

Let $x(n)$ and $d(n)$ represent the input and desired response signals, respectively, of an adaptive filter. The objective of the adaptive Volterra filter is to model the relationship between $x(n)$ and $d(n)$ adaptively using the truncated Volterra series representation of 1. For the sake of computational simplicity, we are interested in stochastic gradient adaptive filters. In order to improve the convergence characteristics of the adaptive filter, we wish to employ an adaptive version of the orthoganalization structure derived in Section 2. Such an approach will result in the lattice filter with different, but equivalent parameterization of the system model.

The adaptive lattice Volterra filter consists of three stages. The first stage is an adaptive linear lattice predictor for the input signal $x(n)$. This predictor can be implemented using one of several algorithms available [2]. When the memory of the system model is N samples long, the lattice predictor has $N-1$ stages. A normalized LMS lattice linear predictor can be realized using the following equations:

$$f_i(n) = f_{i-1}(n) - \rho_i(n)b_{i-1}(n-1), \qquad (12)$$

$$b_i(n) = b_{i-1}(n-1) - \rho_i(n)f_{i-1}(n), \qquad (13)$$

$$\rho_i(n+1) = \rho_i(n) + \frac{\mu}{\hat{\sigma}_{i-1}^2(n)} \cdot \{f_i(n)b_{i-1}(n-1) + b_i(n)f_{i-1}(n)\} \qquad (14)$$

and

$$\hat{\sigma}_i^2(n) = \beta\hat{\sigma}_i^2(n-1) + (1-\beta)\{f_i^2(n) + b_i^2(n-1)\}. \quad (15)$$

In the above equations, $f_i(n)$ and $b_i(n)$ represent the ith order forward prediction error and backward prediction error values, respectively, at time n, $\rho_i(n)$ is the ith reflection coefficient at time n, and μ is a small positive constant that controls the rate of convergence of the various stages of the lattice predictor. The parameter β is bounded above and below by 1 and 0, respectively, and controls the behavior of the adaptive power estimators. Usually, β is chosen as $(1-\mu)$. The above equations are implemented for each stage in a sequential manner. When the adaptive coefficients filter converge in some sense to a neighborhood of their optimal values, the backward prediction error sequences will be nearly orthogonal to each other in the minimum mean-square error sense. However, the backward prediction error signals are not unit variance signals.

The second stage of the adaptive lattice Volterra filter creates N vectors of $P+1$ elements each as

$$\mathbf{B}_{P,i}(n) = [1,\ b_i(n),\ b_i^2(n),\ \cdots,\ b_i^P(n)]^T; \quad i = 0, 1, \cdots, N-1. \quad (16)$$

As discussed in the previous section, it is possible to design a Gram-Schmidt orthogonalizer for $\mathbf{B}_{P,i}(n)$ that is independent of the signal statistics when the input signals are Gaussian. However, to account for potential variations from the Gaussian distribution of the elements of $\mathbf{B}_{P,i}(n)$, we will employ adaptive Gram-Schmidt orthogonalizers for each $\mathbf{B}_{P,i}(n)$. Let $u_{i,j,0}(n)$ denote the jth element of $\mathbf{B}_{P,i}(n)$, i.e.,

$$u_{i,j,0}(n) = b_i^j(n). \quad (17)$$

Then, the equations that describe the Gram-Schmidt orthogonalizers that employ a normalized LMS adaptation algorithm are as follows:

$$u_{i,l,m}(n) = u_{i,l,m-1}(n) - \alpha_{i,l,m-1}(n)u_{i,m-1,m-1}(n) \quad ; l = m+1, \cdots, P \quad (18)$$

$$\alpha_{i,l,m}(n+1) = \alpha_{i,l,m}(n) + \frac{\mu}{\hat{\gamma}_{i,m}^2(n)} u_{i,l,m}(n)u_{i,m,m}(n) \quad (19)$$

and

$$\hat{\gamma}_{i,m}^2(n) = \beta\hat{\gamma}_{i,m}^2(n-1) + (1-\beta)u_{i,m}^2(n). \quad (20)$$

When the coefficients of the processor are close to the optimum values, the signals $\{u_{i,m,m}; m = 0, 1, \cdots, P\}$, which are the output signals of the adaptive orthogonalizer for $\mathbf{B}_{P,i}(n)$ will be approximately uncorrelated with each other. Furthermore, as a consequence of the results of the previous section, when the input signal is Gaussian, all the elements in the set $\{u_{i,m,m}; i = 1, 2, \cdots, N-1,\ \ m = 0, 1, \cdots, P\}$ are approximately uncorrelated with each other. For the rest of the discussion we will denote $u_{i,m,m}(n)$ using $v_{i,m}(n)$.

The third stage of the adaptive filter is the joint process estimator. The signal set that is used for joint process estimation is obtained by nonlinearly combining the various $v_{i,m}(n)$ as

$$s_{i_1,i_2,\cdots,i_N}(n) = v_{1,i_1}(n)v_{2,i_2}(n)\cdots v_{N,i_N}(n); \quad i_1 + i_2 + \cdots + i_N \le P \quad (21)$$

According to Theorem 1, the elements of the set described by the above equation will be orthogonal, or at least close to orthogonal, when the adaptive filter has converged to nearly optimal values and the input signal is Gaussian. Therefore, it is reasonable to develop the adaptive filter by individually adapting the coefficients of $s_{i_1,i_2,\cdots,i_N}(n)$. Let $\{y_k(n); k = 1, 2, \cdots, M\}$ represent an ordered arrangement of all signals $s_{i_1,i_2,\cdots,i_N}(n)$ involved in the joint process estimation. Here M represent the total number of coefficients in the in the joint process estimator. The following equations represent a normalized LMS joint process estimator for the adaptive lattice Volterra filter.

$$e_k(n) = e_{k-1}(n) - w_k(n)y_k(n) \quad (22)$$

$$w_k(n+1) = w_k(n) + \frac{\mu}{\hat{\kappa}_k^2(n)} e_k(n)y_k(n) \quad (23)$$

and

$$\hat{\kappa}_k^2(n) = \beta\hat{\kappa}_k^2(n-1) + (1-\beta)y_k^2(n). \quad (24)$$

In the above equations, the sequence of estimation errors are initialized using $e_0(n) = d(n)$.

4. EXPERIMENTAL RESULTS

In this section, we will present the results of an experiment that demonstrates the properties of the adaptive lattice Volterra filter when its input signals have narrow band characteristics. The results presented are ensemble averages over fifty independent simulations of a system identification problem. The unknown system was a second-order Volterra filter described by the following input-output relationship:

$$\begin{aligned} y(n) = & -0.78x(n) - 1.48x(n-1) + 1.39x(n-2) \\ & + 0.04x(n-3) + 0.54x^2(n) + 3.72x(n)x(n-2) \\ & + 1.86x(n)x(n-2) - 0.76x(n)x(n-3) \\ & - 1.62x^2(n-1) + 0.76x(n-1)x(n-2) \\ & - 0.12x(n-1)x(n-3) + 1.41x^2(n-2) \\ & - 1.52x(n-2)x(n-3) - 0.13x^2(n-3) \end{aligned} \quad (25)$$

This system is identical to the one used in Example 2 of [5]. Four different types of input signals were used in the simulations. Each signal set was generated as the output of a linear system with input-output relationship

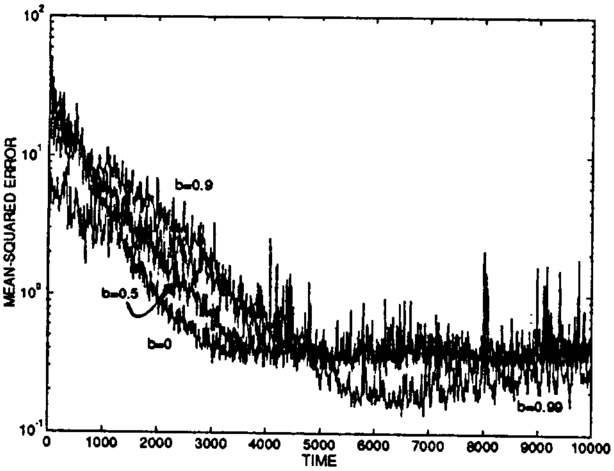

Figure 1: Mean-squared estimation error of the adaptive lattice Volterra filter for four different input signals

$$x(n) = bx(n-1) + \sqrt{1-b^2}\xi(n), \qquad (26)$$

where $\xi(n)$ was zero-mean and white Gaussian noise with unit variance and a was a parameter between 0 and 1 that determined the level of correlation between adjacent samples of the process $x(n)$. Experiments were conducted with a set to 0.00, 0.50, 0.90 and 0.99. When $b = 0$, the input signal is white. As the parameter b approaches 1, the signal characteristics become highly lowpass in nature. The desired response signals were generated by passing the input signals described above through the unknown system, and corrupting the output signals with additive zero-mean and Gaussian noise with variance 0.1. The measurement noise sequence and the input signal x(n) were mutually uncorrelated. In all the experiments, μ and β were chosen to be 0.001 and 0.999, respectively. Figure 1 displays overlayed plots of the squared estimation error signal, averaged overaged over the fifty runs. These error curves were further smoothed by time-averaging over ten consecutive samples. It can be seen from the figure that the rate of convergence of the adaptive filter is similar in all cases, in spite of the fairly large disparity in the spectra of the signals employed. It appears from the results of this experiment that the objective of designing an adaptive filter that is relatively insensitive to the statistics of the input signals has been achieved.

5. CONCLUDING REMARKS

This paper presented an adaptive lattice Volterra filter. The filter is based on a recent result for orthogonalizing Gaussian signals for Volterra system identification problems. The results of a limited number of experiments presented seem to indicate that the filter has good convergence characteristics. Further performance evaluations are necessary to understand the properties of the adaptive filter when higher-order system models are employed and also when the input signals are not Gaussian distributed.

6. REFERENCES

[1] C. E. Davila, A. J. Welch, and H. G. Rylander, III, "A second-order adaptive Volterra filter with rapid convergence," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-35, No. 9, pp. 1259-1263, September 1987.

[2] S. Haykin, *Adaptive Filter Theory* (Second Edition), Prentice Hall, Englewood Cliffs, New Jersey, 1991.

[3] T. Koh and E. J. Powers, "An adaptive non-linear filter with lattice orthogonalization," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Proc.*, Boston, Massachusetts, April 1983.

[4] T. Koh and E. J. Powers, "Second-order Volterra filtering and its application to nonlinear system identification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-33, No. 6, pp. 1445-1455, December 1985.

[5] J. Lee and V. J. Mathews, "A fast recursive least squares adaptive second-order Volterra filter and its performance analysis," *IEEE Transactions on Signal Processing*, Vol. 41, No. 3, pp. 1087-1102, March 1993.

[6] V. J. Mathews, "Adaptive polynomial filters," *IEEE Signal Processing Magazine*, Vol. 8, No. 3, pp. 10-26, July 1991.

[7] V. J. Mathews, "Orthogonalization of Gaussian signals for Volterra system identification," *Proc. Sixth IEEE Digital Signal Processing Workshop*, Yosemite, California, pp. 105-108, October 1994.

[8] G. L. Sicuranza and G. Ramponi, "Adaptive nonlinear digital filters using distributed arithmetics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-34, No. 3, pp. 518-526, June 1986.

[9] G. L. Sicuranza, "Quadratic filters for signal processing," *Proc. IEEE*, Vol. 80, No. 8, pp. 1263-1285, August 1992.

[10] M. A. Syed and V. J. Mathews, "Lattice algorithms for adaptive Volterra filtering," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 41, No. 3, pp. 202-214, March 1994.