Orthogonalization of Correlated Gaussian Signals for Volterra System Identification

V. John Mathews, *Senior Member, IEEE*

***Abstract*— This letter presents a simple method for orthogonalizing correlated Gaussian input signals for identification of truncated Volterra systems of arbitrary order of nonlinearity P and memory length N. The procedure requires a Gram–Schmidt orthogonalizer for a vector containing N elements and some nonlinear processing of the output elements of the Gram–Schmidt processor. However, the nonlinear processors do not depend on the statistics of the input signals and, consequently, are easy to design and implement.**

I. Introduction

NONLINEAR signal processing techniques employing truncated Volterra and other types of polynomial system models have become very popular in recent years [4], [5]. One disadvantage of many such system models and, in particular, of truncated Volterra system models, is the large number of parameters they often require to satisfactorily represent many nonlinear systems. Consequently, problems involving parameter estimation tend to become extremely complex for all but very simple cases. Similarly, gradient adaptive Volterra filters exhibit very slow convergence properties because of the large eigenvalue spread of the autocorrelation matrices of the input vectors. In both of these situations, orthogonalization of the input data will substantially reduce the complications associated with direct formulations of the problem. The objective of this letter is to present a very simple method to orthogonalize the input signals for identifying truncated Volterra systems when the input signal is known to be Gaussian.

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 and $y(n)$ is the output of the system and

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

Note that the above model incorporates the kernel symmetry without any loss of generality. It is convenient to represent the system of (1) using vector notation for our derivation. Let $\mathbf{X}_p(n)$ represent the vector containing all the pth order products of the input signal appearing in (2) and define $\mathbf{X}(n)$ as

$$\mathbf{X}(n) = [1, \mathbf{X}_1^T(n), \mathbf{X}_2^T(n), \cdots, \mathbf{X}_P^T(n)]^T. \tag{3}$$

The basic problem considered in this letter is the orthogonalization of the elements of the input vector $\mathbf{X}(\mathbf{n})$ in (3). We assume that the input signal is zero-mean and Gaussian. 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 nonzero mean value of the input signal. Orthogonalization of Gaussian signals for the truncated second-order Volterra system identification problem has been studied before by Koh and Powers [3] in the context of adaptive filters. Orthogonalization of Gaussian input signals using G-functionals [5] require that the input signal is white. The Schetzen L-functionals described in [5] may be used to identify Volterra systems with colored Gaussian inputs. This letter presents a new procedure for orthogonalizing colored Gaussian input signals for application in the identification of truncated Volterra systems of arbitrary but finite order of nonlinearity and finite memory. As with most orthogonalization procedures, system identification using the orthogonal basis set derived in this paper does not directly estimate the Volterra kernels of the unknown system. Instead, an equivalent set of parameters is estimated during the procedure.

II. The Orthogonalization Procedure

In order to derive the orthogonalizer, consider $\mathbf{X}_1(n)$ given by

$$\mathbf{X}_1(n) = [x(n), x(n-1), \cdots, x(n-N+1)]^T \tag{4}$$

which consists only of the linear components in the input signal set in (3). Let $\mathbf{Q}$ be a lower trinagular matrix such that the elements of the transformed vector

$$\mathbf{U}_1(n) = \mathbf{Q}\mathbf{X}_1(n) \tag{5}$$

have unit variance and are mutually orthogonal, i.e.

$$E\{\mathbf{U}_1(n)\mathbf{U}_1^T(n)\} = \mathbf{I}. \tag{6}$$

One way of efficiently implementing the Gram–Schmidt orthogonalizer for the linear part is to use a lattice predictor [2].

Manuscript received January 31, 1995. This work was supported by a Departmental Grant from IBM. The associate editor coordinating the review of this letter and approving it for publication was Prof. A. E. Yagle.

V. J. Mathews is with Department of Electrical Engineering, University of Utah, Salt Lake City, UT 84112 USA.

IEEE Log Number 9415013.

1070–9908/95$04.00 © 1995 IEEE

The backward prediction error signals of order $0, 1, \cdots, N-1$ orthogonalize the signal set $\{x(n), x(n-1), \cdots, x(n-N+1)\}$. One additional constraint we have imposed on the orthogonalization of the linear part is that the orthogonal signals also have unit variance.

Let $u_i(n);\ i = 1,\ 2,\ \cdots,\ N$ represent the orthogonal signals that make up $\mathbf{U}_1(n)$. Then

$$E\{u_i(n)u_j(n)\} = \delta(i-j). \tag{7}$$

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

$$E\{f(u_i(n))g(u_l(n))\} = E\{f(u_i(n))\}E\{g(u_l(n))\} \tag{8}$$

whenever $i \neq l$ for arbitrary functions f and g.

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

$$\mathbf{U}_{P,i}(n) = [1,\ u_i(n),\ u_i^2(n),\ \cdots,\ u_i^P(n)]^T. \tag{9}$$

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 precompute $\mathbf{Q}_P$. As an example, for $P = 5$, one possibility for $\mathbf{Q}_5$ is

$$\mathbf{Q}_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -3 & 0 & 1 & 0 & 0 \\ 3 & 0 & -6 & 0 & 1 & 0 \\ 0 & 15 & 0 & -10 & 0 & 1 \end{bmatrix}. \tag{10}$$

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

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

Theorem 1: The elements of

$$\mathbf{V}(n) = \mathbf{V}_{P,1}(n) \bigotimes \mathbf{V}_{P,2}(n) \bigotimes \cdots \bigotimes \mathbf{V}_{P,N}(n) \tag{12}$$

where $\otimes$ denotes the Kronecker product [1], are mutually orthogonal.

Proof: Let $v_{P,i,k}(n)$ denote the kth element of $\mathbf{V}_{P,i}(n)$. Recall that $v_{P,i,k}(n)$ and $v_{P,l,m}(n)$ are independent random processes if $i \neq l$. Any element of $\mathbf{V}(n)$ has the form

$$v_{P,1,m_1}(n)v_{P,2,m_2}(n) \cdots v_{P,N,m_N}(n). \tag{13}$$

Now, let us evaluate the cross-correlation of any two elements of $\mathbf{V}(n)$. The expectation will have the form

$$\begin{aligned} E\{&(v_{P,1,m_1}(n)v_{P,2,m_2}(n) \cdots v_{P,N,m_N}(n)) \cdot \\ &(v_{P,1,l_1}(n)v_{P,2,l_2}(n) \cdots v_{P,N,l_N}(n))\} \\ = &E\{v_{P,1,m_1}(n)v_{P,1,l_1}(n)\}E\{v_{P,2,m_1}(n)v_{P,2,l_2}(n)\} \\ &\cdots E\{v_{P,N,m_N}(n)v_{P,N,l_N}(n)\}. \end{aligned} \tag{14}$$

The expectations can be separated as on the right-hand side of the above equations because of the independence of the various elements involved. The expectations of the form $E\{v_{P,i,m_i}(n)v_{P,i,l_i}(n)\}$ are zero whenever $m_i \neq l_i$ since $v_{P,i,m_i}(n)$ and $v_{P,i,l_i}(n)$ are uncorrelated processes and at least one of $v_{P,i,m_i}(n)$ and $v_{P,i,l_i}(n)$ have zero-mean value. Consequently, the only situation under which (14) is nonzero is when $m_1 = l_1, m_2 = l_2, \cdots$ and $m_N = l_N$. This implies that the elements of $\mathbf{V}(n)$ are mutually orthogonal.

It is relatively straightforward to show that a linear transformation exists between the elements of $\mathbf{V}(n)$ and those of $\mathbf{X}_{aug}(n)$ defined as

$$\mathbf{X}_{aug}(n) = \mathbf{Y}_P(n) \bigotimes \mathbf{Y}_P(n-1) \bigotimes \cdots \bigotimes \mathbf{Y}_P(n-N+1) \tag{15}$$

where

$$\mathbf{Y}_P(n) = [1\ x(n)\ x^2(n)\ \cdots\ x^P(n)]^T. \tag{16}$$

It follows immediately that the elements of $\mathbf{V}(n)$ is an orthogonal basis set for the elements of $\mathbf{X}_{aug}(n)$.

While the above result is satisfying in many ways, one should not overlook the fact that $\mathbf{X}_{aug}(n)$ as well as $\mathbf{V}(n)$ has far greater number of elements in them than the number of terms in a Pth order Volterra series expansion with $N-$sample memory. It would be useful to derive an orthogonal basis set for the signals involved in the general Pth order Volterra series expansion. The next theorem provides a solution to this problem.

Theorem 2:

$$\{v_{P,1,m_1}(n)v_{P,2,m_2}(n) \cdots v_{P,N,m_N}(n) \,\big|\, m_1 + m_2 + \cdots + m_N \leq P\}$$

is an orthogonal basis set for

$$\{x^{m_1}(n)x^{m_2}(n-1) \cdots x^{m_N}(n-N+1) \,\big|\, m_1 + m_2 + \cdots m_N \leq P\}.$$

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

Proof: The proof is a straightforward consequence of the fact that $v_{P,i,m_i}(n)$ can be written as a linear combination of $1, x(n), x(n-2), \cdots x(n-i+1)$ and their products of order up to m_i. Consequently, there exists a one-to-one linear transformation between the elements of the two sets defined in the theorem. This implies that the elements of both sets span the same space. The result follows immediately.

III. Concluding Remarks

This letter presented a simple method to orthogonalize Gaussian input signals for identifying truncated Volterra systems. The complexity of implementing the orthogonalizer is comparable to that of the system model itself. This property is different from that of lattice orthogonalizers for Volterra system identification using arbitrary input signals. Such systems are over-parameterized [4] and have significantly higher complexity than the system models. Applications of the orthogonalizer in system identification and adaptive filtering problems are currently being investigated.

REFERENCES

[1] J. L. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 9, pp. 772–781, Sept. 1978.

[2] S. Haykin, *Adaptive Filter Theory,* 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[3] T. Koh and E. J. Powers, "An adaptive nonlinear filter with lattice orthogonalization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Boston, MA, Apr. 1983, pp. 37–40.

[4] V. J. Mathews, "Adaptive polynomial filters," *IEEE Signal Processing Mag.*, vol. 8, no. 3, pp. 10–26, July 1991.

[5] M. Schetzen, *The Volterra and Wiener Theory of Nonlinear Systems,* updated ed. Malabar, FL: R. E. Krieger, 1989.

[6] G. L. Sicuranza, "Quadratic filters for signal processing," *Proc. IEEE,* vol. 80, no. 8, pp. 1263–1285, Aug. 1992.