TECHNIQUES FOR BILINEAR TIME SERIES ANALYSIS

V. John Mathews
Electrical Engineering Department
University of Utah
Salt Lake City, UT 84112

Junghsi Lee
Advanced Technology Center
CCL/ITRI
Hsinchu, Taiwan

ABSTRACT

This paper reviews the general problem of nonlinear time series analysis. The special case of bilinear time series analysis is discussed in detail. The stability of the estimated nonlinear system models is of particular importance. We discuss a simple sufficient condition for the stability of such models. Several approaches to the estimation of the parameters of a bilinear time series are then reviewed. The paper concludes with a discussion of some theoretical aspects associated with using nonlinear time series models in applications involving predictive data compression.

I. INTRODUCTION

A large class of parametric spectrum estimation techniques is based on the assumption that the underlying model that generates the time series is linear. While linear models provide adequate performance in many applications, there are several instances where they can give misleading or inadequate results. Consequently, there has been a large amount of activity in the recent past on fitting signals with nonlinear models. Examples of nonlinear models include threshold linear models, polynomial models and others. An excellent source that discusses various nonlinear time series models is the book by Priestley [9].

The purpose of this paper is to review a particular type of nonlinear time series analyses using polynomial models. A general polynomial time series model describes the relationship between the time series $X(n)$ and $e(n)$, the driving signal that generates the time series in the following manner.

$$X(n) = f_i(X(n-1), X(n-2), ..., X(n-M), e(n), e(n-1), ..., e(n-1)) + e(n) \tag{1}$$

where $f_i(\cdot\cdot\cdot\cdot)$ is an $i-th$ order polynomial in the entries within the parenthesis. Note that $e(n)$ is not measurable. The objective of the time-series analysis here is to identify the parameters of the polynomial model from a single time-limited measurement of $X(n)$. Often such analyses assume that $e(n)$ belongs to a particular class of processes (for example, zero mean, white Gaussian processes).

A special case of polynomial system modeling that has attracted the attention of several researchers is the bilinear system model. In the most general form, a discrete-time random signal $X(n)$ that can be fitted with a bilinear model satisfies the following difference equation.

$$X(n) = \sum_{i=1}^{p} a_i X(n-i) + \sum_{i=1}^{s} \sum_{j=1}^{q} b_{ij} X(n-i)e(n-j) + \sum_{j=1}^{r} c_j e(n-j) + e(n) \tag{2}$$

where $e(n)$ belongs to a stationary and zero-mean white process. In statistical literature, the above model is often denoted as the $BL(p, r, s, q)$ model [12]. This model is attractive since it shares several features with linear system models. In particular, if $b_{i,j} = 0$ for all i and j, we get the familiar ARMA model.

In spite of the simplicity of bilinear system models, there is a large class of nonlinear systems that can be adequately modeled as bilinear systems. For example, it has been shown under fairly mild conditions that a bilinear system model with finite number of coefficients can be used to approximate any Volterra system with arbitrary precision [1]. Consequently, such system models have found a variety of applications including those in control systems, population models, biological systems, economic models, etc. An overview of continuous-time bilinear models and their applications can be found in [7].

In this paper, we will restrict our attention to bilinear time series analysis, even though most of the results we discuss can be directly applied to the general polynomial models also. Two aspects of bilinear time series models will be considered in some detail

1058-6393/93 $03.00 © 1993 IEEE

here – the stability of the models and methods for estimating the parameters of the model.

II. STABILITY OF THE TIME SERIES MODELS

A time invariant model generates a stationary time series only if the model itself is stable. However, determining if a bilinear or the general polynomial model is stable is no trivial task. Exact analysis often requires evaluating the eigenvalues of very large matrices. See Subba Rao [12] for an example. Consequently, researchers in this area have often resorted to developing and using simpler sufficient conditions for stability of such systems [8]. As an example, consider the bilinear time series model in equation (2). If we assume that the driving signal $e(n)$ is zero-mean and Gaussian, exact analysis requires calculating the eigenvalues of a $n^2m \times n^2m$ matrix [12] where $n = Max(p,s)$ and $m = Max(q,r,s)$. However, if we assume that the driving signal is bounded in magnitude (by say, M), we can use a recent result obtained by the authors [3] and show that the system is stable (in the sense that $X(n)$ will be bounded in magnitude) if

i. The roots of the polynomial $(1-\sum_{i=1}^{p}a_iz^{-i})z^p$ (say, $r_1, r_2, ...r_p$) are all strictly less than 1 and

ii.
$$M\sum_{i=1}^{s}\sum_{j=1}^{q}|b_{ij}| \leq \Pi_{i=1}^{p}(1-|r_i|) \tag{3}$$

The above condition requires computation of the roots of a $p-th$ order polynomial and is obviously a significantly simpler problem than solving for the eigenvalues of much larger matrices. Similar stability results for time varying bilinear systems can be found in [10].

III. PARAMETER ESTIMATION FOR BILINEAR TIME SERIES

A. Numerical Search

The most common approach to the estimation of the unknown parameters is to perform a numerical search over the space of the unknown parameters. Consider the bilinear time series model given by equation (2). Assume that N samples of $X(n)$ are available for estimating the parameters of the model. Let Θ denote a vector of unknown parameters and

$$\hat{\Theta} = \begin{array}{l}[\hat{a}_1, \hat{a}_2, ..., \hat{a}_p, \hat{b}_{1,1}, \hat{b}_{1,2}, ..., \\ \quad \hat{b}_{s,q}, \hat{c}_1, \hat{c}_2, ..., \hat{c}_r, \hat{\sigma}_e^2]^T\end{array} \tag{4}$$

denote an estimate of Θ. We will assume for convenience that the driving signal $e(n)$ is white and Gaussian with unknown variance σ_e^2. Also, assume for convenience that p, s, q and r are known.

$$\begin{array}{c}\hat{X}(n,\hat{\Theta}) = \sum_{i=1}^{p}\hat{a}_iX(n-i) + \sum_{i=1}^{s}\sum_{j=1}^{q}\hat{b}_{i,j} \\ X(n-i)\hat{e}(n-j,\hat{\Theta}) + \sum_{i=1}^{r}\hat{c}_i\hat{e}(n-i,\hat{\Theta})\end{array} \tag{5}$$

is an estimate of $X(n)$ using the previous samples $X(n-1), X(n-2), ...$ of the time series and the prediction errors values $\hat{e}(n-1,\hat{\Theta}), e(n-2,\hat{\Theta})$. Here, the prediction error $\hat{e}(n,\hat{\Theta})$ is given by

$$\hat{e}(n,\hat{\Theta}) = X(n) - \hat{X}(n,\hat{\Theta}). \tag{6}$$

One approach to formulating the parameter estimation problem is to choose Θ so as to minimize a cost function defined as

$$Q(\Theta) = \sum_{n=1}^{N}\hat{e}^2(n,\Theta). \tag{7}$$

Since $Q(\Theta)$ is in general multimodal, closed form solutions are not normally available for this minimization problem. In such cases, one has to resort to numerical search procedures. Subba Rao [12] has used the Newton-Raphson technique to iteratively estimate the parameters of a bilinear time-series.

The advantages of iterative techniques are that they are relatively straightforward to formulate and that they provide satisfactory results in general. However, it is possible that the iterative techniques can converge to local minima and even diverge at times [13]. Consequently, manual supervision is almost always required when using this approach.

B. Closed Form Solutions for Special Cases

It turns out that the closed form solution for the bilinear time series estimation problem is possible for certain cases. We consider the $BL(p,0,p,1)$, model given by

$$X(n) = \sum_{i=1}^{p}a_iX(n-i) + \sum_{i=1}^{p}b_iX(n-i)e(n-1) + e(n) \tag{8}$$

when $e(n)$ belongs to a zero-mean and white Gaussian process with unknown variance σ_e^2. It has been shown in [6] and [11] that the first, second and third order statistics of $X(n)$ satisfy the following relationships.

$$\mu = E\{X(n)\} = \left(1-\sum_{i=1}^{p}a_i\right)^{-1} b_1\sigma_e^2$$

Let

$$C(n) = X(n) - \mu. \tag{9}$$

Then

$$r_{cc}(m) = E\{C(n)C(n-m)\}$$

$$= \begin{cases} \sum_{i=1}^{p} a_i r_{cc}(m-i); m > 1 \\ \sum_{i=1}^{p} a_i r_{cc}(m-i) + \sigma_e^2 \sum_{i=1}^{p} \sum_{j=1}^{p} \\ b_i b_j r_{cc}(i-j) \\ +\mu \left(\sum_{i=1}^{p} b_i\right) \sigma_e^2 \left\{a_1 + \mu \sum_{i=1}^{p} b_i\right\} \\ +b_1^2 \sigma_e^4 + \sigma_e^2 \ ; m = 0 \\ \sum_{i=1}^{p} a_i r_{cc}(m-i) + \mu \left(\sum_{i=1}^{p} b_i\right) \sigma_e^2 \ ; m = 1 \end{cases} \tag{10}$$

and

$$\begin{aligned} \Gamma_{ccc}(1,m) &= E\left\{C(n)C(n-1)C(n-m)\right\} \\ &= \sum_{i=1}^{p} a_i \Gamma_{ccc}(i-1, m-1) + \\ &\quad \sigma_e^2 \sum_{i=1}^{p} b_i r_{cc}(m-i) \ ; m \geq 2 \end{aligned} \tag{11}$$

We can solve for the parameters $a_1, a_2, ..., a_p$ and $\sigma_e^2 b_1, \sigma_e^2 b_2, ..., \sigma_e^2 b_p$ from equations (11) (for $m > 1$) and (12). σ_e^2 can be calculated from the equation for $r_{cc}(0)$ (after substituting the estimated values of $a_i's$ and $\sigma_e^2 b_i's$ in the equation and then solving for the positive root of the resulting quadratic equation).

The advantage of this approach is that it is indeed a closed form solution. The disadvantage is that the estimates of the higher order statistics require a fairly large number of data samples to provide reliable results, and consequently, the estimation variance associated with this approach tends to be somewhat higher than for many iterative schemes. Of course, the method is applicable only for a smaller class of models than the iterative schemes.

C. Extended RLS Approach

In the extended recursive least squares (RLS) approach, we try to adaptively predict $X(n)$ using the bilinear system model. The equation satisfying the bilinear predictor is given by

$$\begin{aligned} \hat{X}(n) = \sum_{j=1}^{p} a_i(n) X(n-i) + \sum_{i=1}^{s} \sum_{j=1}^{q} \\ b_{i,j}(n) X(n-i) \hat{e}(n-j) \\ + \sum_{j=1}^{r} c_i(n) \hat{e}(n-i) \end{aligned} \tag{12}$$

where $\hat{e}(n)$ is the prediction error defined as

$$\hat{e}(n) = X(n) - W^T(n) Z(n) \tag{13}$$

and the coefficient vectors $W(n)$ and data vector $Z(n)$ are defined as

$$W(n) = \begin{matrix} [a_1(n), a_2(n), ..., a_p(n), b_{11}(n), ..., \\ b_{sq}(n), c_1(n), ...c_r(n)]^T \end{matrix} \tag{14}$$

and

$$\begin{aligned} Z(n) = [X(n-1), X(n-2), ..., X(n-p), \\ X(n-1)\hat{e}(n-1), ..., X(n-s)\hat{e}(n-q), \\ \hat{e}(n-1), ..., \hat{e}(n-r)]^T \end{aligned} \tag{15}$$

respectively. In the extended RLS approach, we solve for the coefficient vector $W(n)$ at each time, so that a weighted sum of squared prediction errors, given by

$$J(n) = \sum_{k=1}^{n} \lambda^{n-k} \hat{e}^2(k) \tag{16}$$

is minimized. In the above, λ is a number between 0 and 1 that controls the memory span of the adaptive filter. It is well known that the optimum solution for the above problem is given by

$$W(n) = R^{-1}(n) P(n) \tag{17}$$

where

$$R(n) = \sum_{k=1}^{n} \lambda^{n-k} Z(k) Z^T(k) \tag{18}$$

and

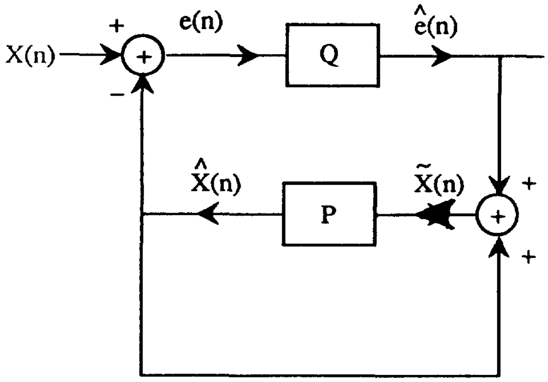
$$P(n) = \sum_{k=1}^{n} \lambda^{n-k} Z(k) X(k) \tag{19}$$

Recursive and much more efficient implementations of the above scheme are possible.

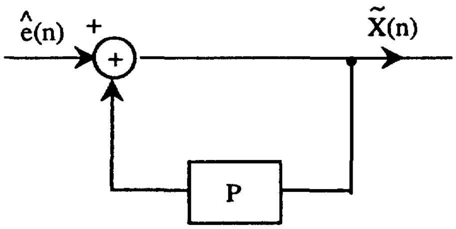
While very little can be stated about the overall stability of the bilinear models generated by the extended RLS approach, one can show that the algorithm is stable in the sense that the time average of the squared prediction error values is bounded [4]. This result is a relatively straightforward extension of a similar result available for RLS adaptive filters employing linear system models [5].

IV. APPLICATION TO PREDICTIVE CODING

Predictive coding [2] is a very popular technique for data compression. Linear prediction is employed in a majority of such coders. Several researchers have



(a) Prediction at encoder.



(b) Reconstruction at receiver.

Figure 1: Block diagram for a predictive coding system.

found that using nonlinear system models (including bilinear models) provides additional coding gains with many types of speech and image signals. In this section, we will address the issue of whether polynomial system models employed along with extended RLS predictors can operate in a stable manner when used in predictive coding applications. The system block diagram is shown in Figure 1.

The key issue in predictive coding systems is that the input signals $X(n)$ is not available at the receiver, and consequently, the predictor P must use the reconstructed signal $\tilde{X}(n)$ at both the encoder and receiver in order to ensure proper operation. In the block diagram, P denotes the bilinear predictor implemented using an extended RLS algorithm and Q is the quantizer for the error signal. The relevant equations that constitute the algorithm are as follows.

ENCODER

$$\begin{gathered} Z(n) = [\tilde{X}(n-1), \tilde{X}(n-2), ..., \tilde{X}(n-p), \\ \tilde{X}(n-1)\hat{e}(n-1), ..., \tilde{X}(n-s)\hat{e}(n-q), \\ \hat{e}(n-1), ..., \hat{e}(n-r)]^T \end{gathered} \tag{20}$$

$$\alpha(n) = \tilde{X}(n) - Z^T(n)W(n-1) \tag{21}$$

$$R(n) = \lambda R(n-1) + Z(n)Z^T(n) \tag{22}$$

$$P(n) = \lambda P(n-1) + Z(n)\tilde{X}(n) \tag{23}$$

$$k(n) = R^{-1}(n)P(n) \tag{24}$$

(Equation (22-24) can be implemented more efficiently using fast algorithms).

$$W(n) = W(n-1) + k(n)\alpha(n) \tag{25}$$

$$\hat{X}(n) = Z^T(n)W(n) \tag{26}$$

$$e(n) = X(n) - Z^T(n)W(n) \tag{27}$$

$$\hat{e}(n) = Q\{e(n)\} \tag{28}$$

$$\tilde{X}(n) = \hat{e}(n) + \hat{X}(n) \tag{29}$$

At the receiver, one can reconstruct $\tilde{X}(n)$ by implementing equations (20), (21), (22), (23), (24), (25), (26) and (29) in order at each instant. It should be clear from the above discussion that the receiver can track the behavior of the predictor at the encoder with knowledge of $\hat{e}(n)$, the quantized prediction error signal alone. Now, the key question becomes: Can the predictor operate in a stable manner when its input is $\tilde{X}(n)$ instead of $X(n)$? To answer this question, we now prove the following result:

Theorem: Consider the bilinear predictive coder of Figure 1. The time average of the prediction error given by

$$\left(\frac{1}{n}\right)\sum_{k=1}^{n}\left(X(k) - \hat{X}(k)\right)^2 \tag{30}$$

is bounded if $\frac{1}{n}\sum_{k=1}^{n} X^2(k)$ and $\frac{1}{n}\sum_{k=1}^{n}(e(k) - \hat{e}(k))^2$ is bounded.

Proof: By extending a result in [4], we can show that

$$\tfrac{1}{n}\sum_{k=1}^{n}\left(\tilde{X}(k)-\hat{X}(k)\right)^2 \leq \tfrac{1}{n}\sum_{k=1}^{n}\tilde{X}^2(k) + \tfrac{\lambda}{n}W^T(0)R(0)W(0) \tag{31}$$

Now, it is easy to show that

$$\tilde{X}(k) = X(k) - (e(k) - \hat{e}(k)) \tag{32}$$

Substituting this in (31), we get

$$\frac{1}{n}\sum_{k=1}^{n}\left(X(k) - (e(k) - \hat{e}(k)) - \hat{X}(k)\right)^2 \leq \frac{1}{n}\sum_{k=1}^{n}(X(k) - (e(k) - \hat{e}(k)))^2 + \frac{\lambda}{n}W^T(0)R(0)W(0) \tag{33}$$

The result of the theorem follows immediately from the above.

When there are a finite number of quantizer levels or codewords, it is not always possible to guarantee that the quantizer error will be bounded in the sense of the theorem. However, there are adaptive quantization algorithms that can guarantee that the quantization error is bounded. The theorem implies that at least when used in conjunction with such quantizers, bilinear predictive coding will operate in a stable manner. In the preliminary experiments we have conducted, bilinear predictive coders with fixed quantizer levels have worked satisfactorily.

V. CONCLUDING REMARKS

In this paper, we reviewed some important aspects of bilinear time series analysis. In particular, we discussed a relatively simple test for the stationarity of the process defined by a bilinear system. We showed that a bilinear predictive coder employing the extended RLS predictor operates in a stable manner. We also discussed various approaches to estimation of the parameters of the bilinear time series model.

References

[1] R.W. Brockett, "Volterra series and geometric control theory," *Automation*, Vol. 12, pp. 167-176,1976.

[2] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, Englewood Cliffs, NJ, 1984.

[3] J. Lee and V. J. Mathews, "A stability condition for certain bilinear systems," *Proceedings of IEEE Winter Workshop on Nonlinear DSP*, Tampere, Finland, January 1993.

[4] J. Lee and V. J. Mathews, "Adaptive bilinear predictors," *Proceedings ICASSP'94*, Adelaide, Australia, May 1994.

[5] L. Ljung and T. Soderstrom, *Theory and practice of recursive identification*, The MIT Press, Cambridge, MA, 1983

[6] V. J. Mathews and T. K. Moon, "Parameter estimation for a bilinear time series model," *Proc. ICASSP*, pp. 3513 - 3517, Toronto, Canada, May 1991.

[7] R. R. Mohler and W. J. Kolodziej, "An overview of bilinear system theory and applications," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-10, No. 10, pp. 683-688, October 1980.

[8] T. Pham-Dinh and L. T. Tran, "On the first-order bilinear time series model," *Journal of Applied Probability*, 18, pp. 617-627, 1981

[9] M. B. Priestley, *Non-Linear and Non-Stationary Time Series Analysis*, Academic Press Inc., New York, 1988.

[10] T. Siu and M. Schetzen, "Convergence of Volterra series representation and BIBO stability of bilinear systems," *International Journal of Systems Science*, Vol. 22, No. 12, pp. 2679-2684, 1991.

[11] S. A. O. Sesay and T. Subba Rao, "Difference equations for higher-order moments and cumulants for the bilinear time series model," *Journal of Time Series Analysis*, February 1991, pp. 159-177.

[12] T. Subba Rao, "On the theory of Bilinear time series models," *Journal of Royal Statistical Society* (b), Vol. 43, No. 2, pp. 244-255: 1981.

[13] S. J. Wissler, "Parameter estimation for a bilinear time series model," *Proceedings of the National Conference on Undergraduate Research*, Minneapolis, MN, March 1992.