CODEBOOK PREDICTION: A NONLINEAR SIGNAL MODELING PARADIGM

Andrew C. Singer *Gregory W. Wornell* *Alan V. Oppenheim*

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT

A nonlinear generalization of the family of autoregressive signal models is introduced. This generalization can be viewed as an autoregressive model with state-varying parameters. For such signals, minimum mean-square error prediction can be reformulated as an interpolation problem. A novel interpretation of the signal as a codebook for its own prediction leads to an interpolation strategy resembling a predictive counterpart to vector quantization. The applicability of this model is then demonstrated empirically for a variety of signals.

1 INTRODUCTION

A popular signal modeling technique is to model signals as outputs of dynamic systems excited by either deterministic or random inputs. Once the appropriate form of the signal model is specified, its parameters can be estimated from observations of the signal, and properties of the signal can be inferred from the model. Autoregressive (AR) signal modeling, a particularly useful technique in which the system model is linear and time-invariant, is described in terms of an n^{th} order difference equation of the form,

$$y[k+1] = \sum_{i=0}^{n-1} a_i y[k-i] + u[k], \tag{1}$$

where $y[k]$ is the output signal and input, $u[k]$, is stationary white noise. Not only have these models been shown to be well-suited for many signals of interest, but the linearity of the model also allows for simple analysis, especially when the mean-square error criterion is used. In fact, very efficient algorithms exist for calculating model parameters. For these reasons, AR modeling has been popular in the signal processing literature in, *e.g.*, problems of prediction and spectral estimation [3] [4].

For nonstationary signals, such as speech, a linear time-varying (LTV) generalization of the AR model is frequently used. In particular, an AR model is fit locally in time over a small region of the signal and the parameters of the model are allowed to vary as a function of time.

In a complementary manner, a nonlinear time-invariant (NLTI) generalization of the AR model can be used to account for signals for which the linearity requirement is inappropriate. There are, of course, many physical signals generated by processes that are inherently nonlinear. In such cases, exploitable redundancies may exist in the signal that are transparent to linear methods, so linear modeling may fail to sufficiently capture the underlying structure. As we will show, an effective strategy for nonlinear signal modeling of this case involves fitting an AR model to the signal locally in state space. In effect, the model parameters vary as a function of the state.

2 NONLINEAR AUTOREGRESSIVE MODELS

A broad class of systems, including AR models and both the LTV and NLTI generalizations, can be represented in a common state space form:

$$\begin{aligned} \mathbf{x}[k+1] &= F(\mathbf{x}[k], \mathbf{u}[k], k) && \text{(2a)} \\ \mathbf{y}[k] &= G(\mathbf{x}[k], \mathbf{u}[k], k), && \text{(2b)} \end{aligned}$$

where the $n \times 1$ vector $\mathbf{x}[k]$ is the state, the $p \times 1$ vector $\mathbf{u}[k]$ is the input, and the $m \times 1$ vector $\mathbf{y}[k]$ is the output. When the system is LTI, the state equations (2) may be written in the form

$$\begin{aligned} \mathbf{x}[k+1] &= A\mathbf{x}[k] + B\mathbf{u}[k] && \text{(3a)} \\ \mathbf{y}[k] &= C\mathbf{x}[k] + D\mathbf{u}[k]. && \text{(3b)} \end{aligned}$$

In the AR case, the matrix A from (3) takes companion form,

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_1 & a_0 \end{bmatrix}, \tag{4}$$

while matrices B,C, and D become $\begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}^T$, $\begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}$, and 0 respectively.

This work was funded in part by a subcontract from Lockheed Sanders, Inc., under ONR Contract No. N00014-91-C-0125, in part by the Defense Advanced Research Projects Agency monitored by ONR under Grant No. N00014-89-J-1489, and in part by the U.S. Air Force-Office of Scientific Research under Grant No. AFOSR-91-0034.

0-7803-0532-9/92 $3.00 © 1992 IEEE

Generalizing the the model to include nonlinear systems, while retaining the companion state variable structure, leads to systems described by n^{th} order nonlinear difference equations of the form

$$y[k+1] = F(y[k], y[k-1], \ldots, y[k-n+1]) + u[k], \quad (5)$$

where $F(\cdot)$ maps $\Re^n$ to $\Re$, and $u[k]$ is stationary white noise. The state space representation can be viewed as a simple extension of (3), *i.e.*,

$$\mathbf{x}[k{+}1] = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \mathbf{x}[k] + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ F(\mathbf{x}[k]) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u[k], \quad (6a)$$

$$y[k] = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \mathbf{x}[k]. \quad (6b)$$

For convenience, we will refer to the class of processes that can be expressed in the form (6) as nonlinear autoregressive (NLAR) processes.

3 PREDICTION OF NLAR PROCESSES

From the Markov structure of NLAR processes, we can see that the statistics for $y[k+1]$ given its entire history depend only on the most recent n values:

$$P(y[k{+}1] \,|\, y[i], 0{\le}i{\le}k) = P(y[k{+}1] \,|\, y[i], k{-}n{+}1{\le}i{\le}k). \quad (7)$$

It is clear from (6), that we can reconstruct the state vector, $\mathbf{x}[k]$, from observations of the the scalar output, $y[k]$, in the following manner,

$$\mathbf{x}[k] = [y[k-n+1], \ldots, y[k-1], y[k]]^T. \quad (8)$$

From (7) and (8), the conditional statistics become,

$$P(y[k{+}1] \,|\, y[i], 0{\le}i{\le}k) = P(y[k{+}1] \,|\, \mathbf{x}[k]). \quad (9)$$

Thus, the minimum mean square error (MMSE) estimate of $y[k+1]$ given its entire signal history is

$$\begin{aligned} \hat{y}[k{+}1] &= E\{y[k+1] \,|\, \mathbf{x}[k]\} \\ &= E\{F(\mathbf{x}[k]) + u[k] \,|\, \mathbf{x}[k]\} \\ &= F(\mathbf{x}[k]). \end{aligned} \quad (10)$$

Although $F(\mathbf{x})$ is part of the system model, and therefore unavailable, the state dynamics of the system can be observed through

$$y[k+1] = F(\mathbf{x}[k]) + u[k]. \quad (11)$$

Thus, given $y[k]$ and recovering $\mathbf{x}[k]$ from (8), the signal history represents a set of noisy samples of $F(\mathbf{x})$, nonuniformly distributed in state space. Consequently, the estimation problem for $y[k+1]$ can be viewed as a problem of interpolating $F(\mathbf{x})$ from noisy samples.

Based on the interpolation viewpoint, several solutions to the prediction problem arise naturally. For example, kernel-based strategies involving splines or radial basis functions can be used to create a global approximation of $F(\mathbf{x})$ [1]. One benefit of such a scheme is that the model for $F(\mathbf{x})$ can be precomputed, making signal prediction a simple function evaluation. However, the performance of such schemes depends critically on the choice of the kernel, since rather strong assumptions are imposed on $F(\mathbf{x})$ between the observed samples.

A philosophically different approach which makes fewer assumptions about the behavior of the function between samples is based on the use of local models. In a manner reminiscent of vector quantization, we can view the signal as a codebook of (state-vector,signal-value) pairs of the form, $(\mathbf{x}[k], y[k+1])$. Because each codebook entry must satisfy (11), the strategy is to use the present state of the system, $\mathbf{x}[k]$, to "lookup" $F(\mathbf{x}[k])$ in the codebook. Assuming only that the state dynamics are locally smooth, and given a long enough signal history, the elements of the codebook near the present state of the system describe a local model of the state dynamics from which $F(\mathbf{x}[k])$ can be inferred. In order for such a scheme to succeed, the codebook must contain entries near any encountered state of the system — therefore only a finite portion of state space can be modeled.

We arrive at the following strategy for predicting $y[k{+}1]$ given $y[i], 0 \le i \le k$:

- Form a codebook of pairs $(\mathbf{x}[i], y[i{+}1])$ from the signal history.
- Select pairs $(\mathbf{x}[i], y[i+1])$ from the codebook for $\mathbf{x}[i]$ near $\mathbf{x}[k]$.
- Fit a local model $y[i+1] \approx \hat{F}(\mathbf{x}[i])$ to the selected pairs.
- Apply the local model to obtain $\hat{y}[k+1] = \hat{F}(\mathbf{x}[k])$.

Although the problems can be considered independently, it will be easier to approach pair selection once the local models are determined.

3.1 LOCAL MODELS

If the state update function is sufficiently smooth that we can closely represent $F(\mathbf{x})$ in the vicinity of $\mathbf{x}[k]$ by the first few terms of its multidimensional Taylor series expansion, *i.e.*,

$$\begin{aligned} F(\mathbf{x}) &= F(\mathbf{x}[k]) + \nabla F^T(\mathbf{x}[k])(\mathbf{x} - \mathbf{x}[k]) + \cdots \\ &\approx b + \mathbf{a}^T\mathbf{x}, \end{aligned} \quad (12)$$

then we may approximate $F(\mathbf{x})$ as either a constant or as a linear function near $\mathbf{x}[k]$. If our selected pairs are close enough to $\mathbf{x}[k]$, the local model will be a good approximation. Generating a locally linear approximation $(F(\mathbf{x}) \approx \hat{F}(\mathbf{x}) = b + \mathbf{a}^T\mathbf{x})$, amounts to fitting the parameters b and $\mathbf{a}$ to the selected pairs, $(\mathbf{x}[i], y[i+1])$, in the region of state space near $\mathbf{x}[k]$. The model will generally provide an overdetermined set of linear equations in the model parameters,

$$y[i+1] = b + \mathbf{a}^T\mathbf{x}[i], \quad (13)$$

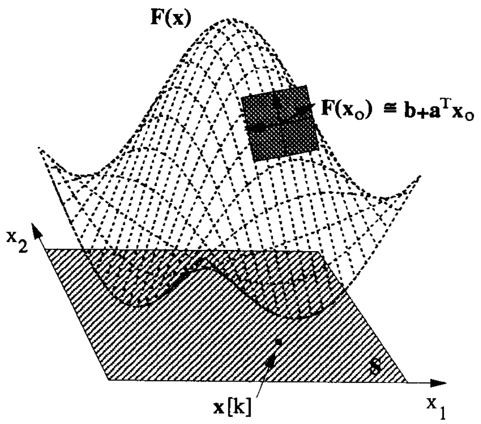

Figure 1: Locally Linear State Space Model

which can be solved in a least-squares sense.

Our locally linear model can be seen as a generalization of the AR model of (4), with the matrix A varying as a function of the state. We also have the addition of the constant, b, to account for any local mean of the process. For example, in the case of a second order system ($n = 2$), our state space, $\mathcal{S}$, is a plane with axes x_1 and x_2. As shown in Fig. 1, we can look at $F(\mathbf{x})$ as a surface over the plane, $\mathcal{S}$. In Fig. 1, we model $F(\mathbf{x}_0)$ for $\mathbf{x}_0$ near $\mathbf{x}[k]$ by a plane tangent to $F(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}[k]$. In contrast, traditional AR modeling of the system would approximate the function $F(\mathbf{x})$ with a single plane through the origin — thus the locally linear model reduces to the AR model when the "local" region extends to include all of state space.

3.2 PAIR SELECTION

Pair selection involves two conflicting requirements. On one hand, our selection process must choose pairs that are close to $\mathbf{x}[k]$ in order that the local model hold; on the other, we need to ensure that the local fitting procedure is taken over a large enough number of measurements. Using nearest-neighbor pair selection, various trade-offs can be made. For a large class of signals, including many stable AR and Markov processes [7], an estimation procedure that selects a number of nearest-neighbors that increases at a rate slower than the data length can be shown to converge in the mean-square sense to $F(\mathbf{x}[k])$. As we will show, however, in other cases it is preferable to hold fixed the number of nearest neighbors as the data length increases.

4 RESULTS

The NLAR prediction procedure has been tested on a variety of signals, with traditional AR modeling (linear prediction) used as a means of comparison. To demonstrate the viability of the algorithm, the first set of experiments involve predicting the following NLAR process,

$$y[k+1] = .2y[k]^2 + .25y[k] - .2y[k]y[k-1] + .125y[k-1] + u[k], \tag{14}$$

where $u[k]$ is a unit variance white Gaussian noise sequence. Theoretically, we expect the NLAR model to yield a mean-square prediction error that decreases monotonically to the variance of the driving noise. In Fig. 2, the mean-square prediction error for 100 Monte Carlo trials of both linear prediction and locally linear prediction are shown. We see that empirically, only the locally linear technique achieves a mean-square prediction error near the variance of the driving noise, *i.e.*, the minimum attainable mean-square prediction error.

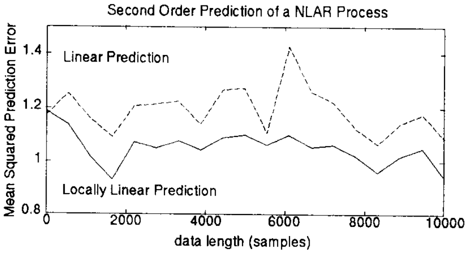

Figure 2: Prediction of 2^{nd} Order NLAR Process

A rich class of NLAR processes exist even when the drive term, $u[k]$, is identically zero. In fact, an increasingly important class of such NLAR processes are chaotic processes. As an example, the Henon Map is a NLAR process given by

$$y[k+1] = 1 - 1.4y[k]^2 + .3y[k-1], \tag{15}$$

which is known to exhibit chaotic behavior for certain initial conditions, *e.g.*, $y[0] = y[1] = 0$. Because there is no drive term, and the processes are therefore deterministic, the prediction performance is limited only by the numerical instability of the map. For this example, four methods of prediction were employed: linear, locally constant, 25 neighbor locally linear, and 10 neighbor locally linear. Prior to prediction, the mean of the process was removed and the variance was normalized to ensure fair comparisons. From Fig. 3, we see that linear prediction yielded almost no prediction gain, while the mean-square error for the local models decreases steadily with the data length. In the absence of the driving noise, as the data length increases, the average distance from $\mathbf{x}[k]$ to the selected pairs decreases. Thus, the approximation of (12) becomes more accurate. The 10 neighbor locally linear technique outperforms the 25 neighbor technique because the average distance to the selected pairs is smaller. More generally, NLAR prediction is useful in modeling a variety of chaotic data. In fact, Farmer and Sidorowich have used essentially such an approach for the prediction of chaotic time series [2].

Our final example involves prediction of speech. Traditionally, speech is modeled as a time-varying AR process

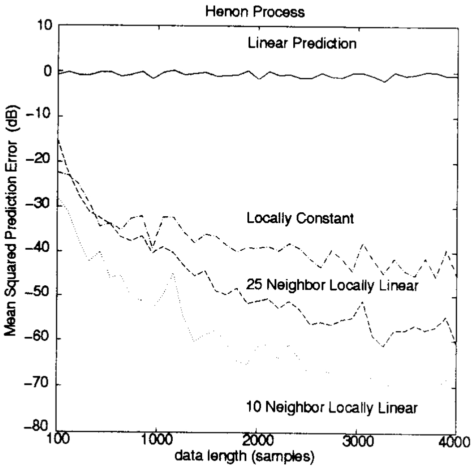

Figure 3: Prediction of Henon Process

to account for the nonstationarity of the speech production mechanism. Fig. 4 shows the mean-square prediction error achieved by time-varying AR and state-varying AR models. In each case, 50 previous samples were used to determine the parameters of a 12^{th} order model. For the state-varying case, the 50 nearest neighbors in state space were used, while for the time-varying model, the 50 most recent samples in time were used. That there are regions in the signal where the state-varying approach significantly outperforms the LTV AR model, suggests that NLAR models may be potentially useful for speech analysis. In fact, one can interpret the preliminary work of Townsend on the nonlinear prediction of speech in the NLAR framework [6].

5 CONCLUDING REMARKS

For a nonlinear time-invariant generalization of AR processes we have developed a signal modeling algorithm based on a codebook prediction paradigm. The development of algorithms for model order selection represents an important direction for future research. Another direction involves improving the computational efficiency of these algorithms perhaps by incorporating codebook structuring notions from vector quantization. Finally, by combining the LTV and NLTI generalizations of AR processes, a broader class of signals can be modeled. This too represents a potentially rich area of investigation.

REFERENCES

[1] M. Casdagli. Nonlinear prediction of chaotic time series. *Physica D*, 35:335–356, 1989.

[2] J.D. Farmer and J.J. Sidorowich. Predicting chaotic time series. *The American Physical Society*, 59(8), 1987.

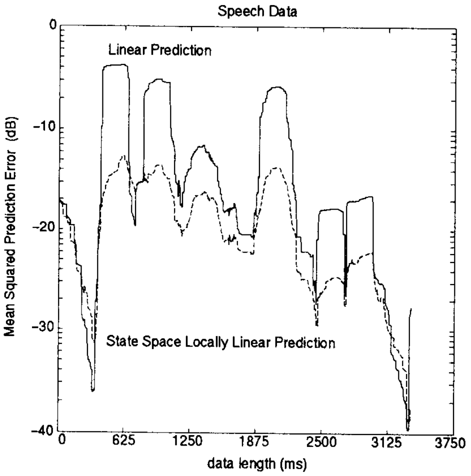

Figure 4: Prediction of Speech

[3] Stephen M. Kay and Stanley L. Marple Jr. Spectrum analysis — a modern perspective. *Proc. IEEE*, 69:1380–1418, 1981.

[4] J. Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4), April 1975.

[5] Andrew C. Singer. Codebook signal prediction. Master's thesis, M. I. T., Cambridge, MA, December 1991.

[6] Brent Townsend. Nonlinear prediction of speech signals. submitted to IEEE Trans. on Acoustics, Speech and Signal Processing, 1990.

[7] S. Yakowitz. Nearest-neighbour methods for time series analysis. *Journal of Time Series Analysis*, 8(2), 1987.