Nonlinear Autoregressive Modeling and Estimation in the Presence of Noise

Andrew C. Singer, Gregory W. Wornell, and Alan V. Oppenheim
Research Laboratory of Electronics
Massachusetts Institute of Technology

Revised September 1993

Abstract

Nonlinear autoregressive processes constitute a potentially important class of nonlinear signal models for a wide range of signal processing applications involving both natural and man-made phenomena. A state space characterization is used to develop algorithms for modeling and estimating signals as nonlinear autoregressive processes from noise-corrupted measurements. Special attention is given to chaotic processes, which form an important subclass of nonlinear autoregressive processes. The modeling algorithms are based on the method of total least-squares, and exploit the local structure of the signals in state space. The recursive estimation algorithms for addressing problems of filtering, prediction, and smoothing, are based on extended Kalman estimators, and jointly exploit aspects of both the temporal and state-space structure in these processes. The resulting algorithms are practical both in terms of computation and storage requirements, and their effectiveness is verified through simulations involving noisy nonlinear autoregressive data.

EDICS: 5.2.1

1   Introduction

Signal modeling is an important aspect of a variety of signal processing problems involving such objectives as signal detection, prediction, smoothing, enhancement, classification and separation. Many traditional signal processing algorithms rely heavily on signal models which are inherently linear. However, for large classes of physical signals, nonlinear models are more appropriate. Examples of such signals include turbulent flow, wind noise and acoustic reflections from sea ice. Furthermore, nonlinear models hold the promise of yielding

more efficient representations and more effective processing algorithms in applications such as speech and imagery for which linear models have traditionally proven useful though not optimal.

In this paper, we focus on a family of nonlinear signal models for signal processing applications which, due to their particular dynamic structure, are referred to as *nonlinear autoregressive* (NLAR) processes. These models, which generalize the well-known linear autoregressive processes, are potentially highly useful models for a range of natural and man-made phenomena and lend themselves to the development of some efficient, practical, and robust estimation algorithms for their application to real data.

There are numerous examples of signals that are well-described by the nonlinear autoregressive model framework. These include several classes of signals generated as outputs of nonlinear dynamic systems exhibiting chaotic behavior [1]. Such signals have been increasingly identified as potentially important models for many signal processing scenarios; see, *e.g.*, [2] [3]. Furthermore, chaotic signal sets have also been recently identified as potentially useful in a variety of broadband and secure communication applications, motivated, in part, by their ease of generation, attractive pseudorandom characteristics, compelling synchronization properties [4], and robustness with respect to channel distortion [3] [5]. Although a discussion of these applications is beyond the scope of this paper, their importance motivates our choice of chaotic signals as representative examples in several of the simulations we describe in the paper.

The primary contributions of this paper are a collection of practical and efficient modeling and estimation algorithms designed specifically for use with NLAR data corrupted by potentially large amounts of broadband measurement noise. Furthermore, we focus on a fairly typical scenario in which virtually no *a priori* information concerning the underlying NLAR signal is available. By minimizing prior assumptions about the NLAR signal model and by explicitly taking into account measurement noise, the algorithms we develop turn out to be rather general and particularly robust to both modeling and observation error.

In the first part of the paper, we develop a fast and effective total least squares algorithm which builds and dynamically refines a model for an NLAR process from a stream of arbitrarily noisy measurements of the process. In the remainder of the paper, we then use

this modeling algorithm in conjunction with an extended Kalman estimation framework to build efficient recursive estimation algorithms for extracting and extrapolating the NLAR process from its noise-corrupted measurements. In particular, we develop solutions to the problems of NLAR signal filtering, predicting, and smoothing.

NLAR models and their variants have appeared over the last several decades in a number of fields including statistics, economics, mathematics, physics, and, to a lesser extent, engineering. Furthermore, more recently there has been dramatic growth of interest in the specific case of chaotic signals in the physics community. However, for the most part, the existing literature has tended to focus on modeling and estimation algorithms for data with relatively high signal-to-noise ratio (SNR). Furthermore, many of these algorithms have required more detailed knowledge of the dynamic structure of the underlying signal than we assume in this work. For the interested reader, some of the references in this area include [6] − [17].

The organization of this paper is as follows. In Section 2, we define a class of NLAR models and develop some of their special properties and characteristics. In Section 3 we construct and evaluate the total least squares algorithms for modeling these processes. In Section 4, we combine the total least squares modeling procedure with an extended Kalman filter framework to develop recursive algorithms for predicting, filtering and smoothing NLAR signals. Finally, Section 5 contains some concluding comments and remarks.

2   Nonlinear Autoregressive Models

The class of processes that we consider are those described by discrete-time nth-order nonlinear dynamics of the form

$$y[k+1] = \mathbf{F}(y[k], y[k-1], \ldots, y[k-n+1]) + u[k], \tag{1}$$

where $\mathbf{F}(\cdot)$ is a nonlinear mapping, and $u[k]$ is zero-mean stationary white random process of variance σ_u^2. It is clear from (1) that the signal history provides an observation of the state of the underlying system. The canonical state space characterization of (1) is given

by

$$\mathbf{x}[k+1] = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \mathbf{x}[k] + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathbf{F}(\mathbf{x}[k]) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u[k], \tag{2a}$$

$$y[k] = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \mathbf{x}[k] \tag{2b}$$

with the state vector $\mathbf{x}[\cdot]$ given by[1]

$$\mathbf{x}[k] = \begin{bmatrix} y[k-n+1] & y[k-n+2] & \cdots & y[k] \end{bmatrix}^T. \tag{3}$$

Although cumbersome in appearance, the form (2) provides a natural framework for the development of NLAR modeling and estimation techniques, as we will see.

A particularly simple but conceptually useful example of an NLAR process as represented by (1) is given by

$$y[k+1] = 10\sin(-0.9y[k] + 0.5y[k-1]) + u[k] \tag{4}$$

This second-order process, which can be viewed as a traditional linear autoregressive process with a soft-limiter nonlinearity in the feedback path, will be used in a variety of the simulations described within this paper.

A number of special subclasses of NLAR processes warrant further discussion. Linear autoregressive processes are, of course, special cases of (1). In this case, (2) reduces to

$$\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k] + \mathbf{e}_0\, u[k] \tag{5a}$$

$$y[k] = \mathbf{e}_0^T\, \mathbf{x}[k] \tag{5b}$$

---

[1]The notation $[\cdot]^T$ is used in this paper to denote transposition.

4

where the state transition matrix $\mathbf{A}$ takes companion form, *i.e.*,

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \hline & \mathbf{a}^T \end{bmatrix} \tag{6}$$

with

$$\mathbf{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^T$$

denoting the vector of autoregressive parameters, and where $\mathbf{e}_0$ is the unit vector

$$\mathbf{e}_0 = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}^T. \tag{7}$$

A number of well-known linear prediction algorithms exist for modeling and processing signals in this subclass both in the case of noise-free data [18] and noise-corrupted data [19] [20]. Furthermore, as will become apparent, these algorithms may be viewed in some sense as special cases of the more general algorithms we develop in this paper. However, we emphasize that linear autoregressive processes are certainly not representative of the class of NLAR processes, and that the properties and behavior of NLAR processes are far richer than can be inferred from those of linear autoregressive processes.

Several families of chaotic signals constitute another distinct subclass of NLAR processes. In particular, many important discrete-time chaotic signals can be described in the state space form

$$\begin{align} \mathbf{s}[k+1] &= \mathbf{H}(\mathbf{s}[k]) \tag{8a} \\ y[k] &= \mathbf{G}(\mathbf{s}[k]), \tag{8b} \end{align}$$

where $\mathbf{H}$ and $\mathbf{G}$ are nonlinear mappings with some specific attributes and the $\mathbf{s}[k]$ are state vectors. In accordance with the embedding theorems of Takens [21] and Mañe [22], it is frequently the case that chaotic signals $y[k]$ generated according to (8) can be equivalently described in the form (2) with $u[k] = 0$ through the use of nonlinear coordinate transformations of the state vector $\mathbf{s} \to \mathbf{x}$.

A prominent example of a chaotic signal which will be used in a variety of the simulations presented in this paper is that associated with the Henon map [23]. This signal can be directly described in the NLAR form

$$y[k+1] = 1 - 1.4y^2[k] + 0.3y[k-1]. \tag{9}$$

Several families of oscillators and quasi-oscillators can also be represented in the form (1) with $u[k] = 0$. More generally, it is well known [6] that the asymptotic characteristics of signals generated according to dynamics of the form (1) with $u[k] = 0$ can be classified into one of a number of distinct categories depending on the initial conditions and the details of the mapping $\mathbf{F}(\cdot)$. Specifically, as $n \to \infty$ such signals may approach either a constant value, corresponding to a fixed point; a periodic oscillation, corresponding to a limit cycle; a quasiperiodic cycle, analogous to the superposition of two or more sinusoids having incommensurate frequencies; or a nonperiodic trajectory, corresponding to chaos. While a further discussion of these classifications is beyond the scope of this paper, it is worth noting that what makes the case of chaos rather interesting from a signal processing perspective is that the corresponding signals have broadband spectral characteristics.

In the remainder of this work, it will be convenient to view NLAR processes both in the signal domain and in n-dimensional state space through which the states $\mathbf{x}[k]$ evolve with iteration k. In general, the portions of the state space visited during the generation of $y[k]$ according to (1) depend strongly not only on the particular map $\mathbf{F}(\cdot)$ and on the initial conditions, but on the details of the random sequence $u[k]$ as well. For certain types of NLAR processes, the corresponding trajectories through state space have special characteristics. For example, the state space orbits of chaotic signals, specifically, are both bounded and exhibit strong sensitivity to initial conditions. This has a number of implications, among which is that the trajectories trace out a geometric attractor in state space whose dimension is not only lower than that of the state space itself, but non-integral (*i.e.*, fractal) as well [1]. While it turns out that ultimately these features strongly enhance the viability of NLAR modeling in the case of chaotic data, we do not specifically explore such issues in this paper. However, the algorithms we develop in the following sections are able to fully exploit such behavior when it is present.

3   NLAR Signal Modeling

In this section, we address the following modeling problem: given an observed data sequence

$$z[k] = y[k] + w[k] \qquad k = 0, 1, \ldots, K - 1 \qquad (10)$$

where $y[k]$ represents a signal of interest and $w[k]$ represents a zero-mean, stationary white noise process of variance σ_w^2, we wish to model $y[k]$ in the form (1). The complexity of this estimation problem depends strongly on the degree of *a priori* information available. In this work, we restrict our attention to the case in which the model order n is known, although we recognize that in practice n generally must also be imposed or estimated from the noisy data. We return to this issue in Section 3.3. Furthermore, we assume only that $\mathbf{F}(\cdot)$ is a continuously differentiable mapping. While further constraining $\mathbf{F}(\cdot)$ to be of a particular parameterized form frequently leads to a well-posed and highly tractable estimation problem, such constraints invariably limit the class of signals that can be represented using the NLAR framework. For this reason, in our treatment of the problem we specifically avoid parameterizations such as those used in polynomial approximation and radial basis function approaches [7] [17].

In our development, we formulate the problem of modeling $\mathbf{F}(\cdot)$ as one of functional interpolation from samples whose values and whose locations are both noisy. To be more specific, $(\mathbf{x}, \mathbf{F}(\mathbf{x}))$ defines a surface in $(n+1)$-dimensional space. Using (10) we rewrite (1) as

$$z[k + 1] - w[k + 1] - u[k] =$$
$$\mathbf{F}(z[k] - w[k], z[k - 1] - w[k - 1], \ldots, z[k - n + 1] - w[k - n + 1]). \qquad (11)$$

If $u[k]$ and $w[k]$ are both zero, then a set of $n+1$ successive observations $z[k+1]$, $z[k]$, $\ldots$, $z[k-n+1]$ provides a sample of the function $\mathbf{F}$, with $z[k]$, $\ldots$, $z[k-n+1]$ specifying the location of the sample in n-dimensional space and $z[k+1]$ the value of $\mathbf{F}(\cdot)$ at that sample point. More generally, $w[k]$ introduces noise into the location of the sample and $u[k]$ and $w[k+1]$ introduce noise into the observed value of the function at the sample point. Consequently, given

K observations $z[0]$, $z[1]$, ..., $z[K-1]$ we view the modeling problem as one of modeling the surface $\mathbf{F}(\cdot)$ from the observed samples of $\mathbf{F}(\cdot)$ which are noisy in both their location and their value. Our approach is developed in the next section and is based on a total least squares error criterion [24].

There are two essential components to the modeling problem in the noisy scenario. One component involves estimating the underlying NLAR signal from the noisy measurements (10). The second involves estimating the dynamics of the underlying NLAR signal from these estimates. Clearly, both components are closely coupled: separating the signal from its noise background requires some reasonable estimate of the dynamics $\mathbf{F}(\cdot)$, while accurate estimation of the dynamics requires some reasonable estimate of the underlying NLAR signal. This close coupling will be apparent in the algorithm we develop, which treats both components jointly.

3.1   Modeling Based on Total Least Squares

Our basic approach is to implicitly exploit the smoothness constraint on the mapping $\mathbf{F}(\cdot)$ by selecting a locally affine interpolant for $\mathbf{F}(\cdot)$ in every neighborhood of state space based on a total least squares criterion. As such, it will be possible to interpret this approach as a generalization of one used by Farmer and Sidorowich to model chaotic signals from noise-free measurements [6]. It can also be viewed as a generalization of traditional linear autoregressive modeling algorithms such as linear prediction [18], which uses a globally linear interpolant with a least squares criterion. Although one can imagine a variety of higher-order generalizations to this basic approach, in practice, the associated improvements in performance are typically marginal.

Following the previous discussion, we pose the modeling problem as follows. For a specified location $\mathbf{x}_*$ in state space, estimate $\mathbf{F}(\mathbf{x}_*)$ from the noisy data (10). A local modeling strategy utilizes the data in the vicinity of $\mathbf{x}_*$ to estimate $\mathbf{F}(\mathbf{x}_*)$. Specifically, the observed data $z[0]$, $z[1]$, ..., $z[K-1]$ corresponds to a set of vectors

$$\mathbf{z}[k] = \left[\begin{array}{cccc} z[k-n+1] & z[k-n+2] & \cdots & z[k] \end{array} \right]^T,$$

for $k = n - 1$, n, ..., $K - 1$ which constitute noisy measurements of $K - n + 1$ corresponding state vectors $\mathbf{x}[k]$ for $k = n - 1$, n, ..., $K - 1$. However, these same data also provide noisy measurements of the corresponding values $\mathbf{F}(\mathbf{x}[k])$ to which these states map, $viz.$, $z[k+1]$ for $k = n - 1$, 0, ..., $K - 2$. In other words, from the observations we can, in effect, extract a noisy table consisting of state vectors and the corresponding values to which they map. To estimate $\mathbf{F}(\mathbf{x}_*)$ we may select data vectors $\mathbf{z}[k]$ suitably close to $\mathbf{x}_*$ and then appropriately interpolate based on where their corresponding states map.

In order to specify a rule for the selection of data vectors $\mathbf{z}[k]$ near our arbitrary state of interest $\mathbf{x}_*$, two factors must be considered. First, it is necessary to choose a metric for defining distance between state vectors. Second, given this metric, a useful definition of neighborhood is required. For our purposes, the usual Euclidean norm is an adequate distance metric. With this metric, two of the most practical neighborhood selection strategies are to either select all data vectors within some prescribed radius of $\mathbf{x}_*$, or to select a prescribed number of the nearest data vectors to $\mathbf{x}_*$. When the observation noise variance σ_w^2 is known, the former strategy is often rather natural. In particular, with $E\left[\cdot\right]$ denoting expectation and $\|\cdot\|$ the Euclidean norm, for each k,

$$E\left[\|\mathbf{z}[k] - \mathbf{x}[k]\|^2\right] = n\sigma_w^2. \tag{12}$$

Consequently, it is reasonable to choose a radius that is a small multiple of the standard deviation $\sqrt{n}\sigma_w$. On the other hand, the nearest neighbor selection strategy is particularly convenient when σ_w^2 is unknown. Furthermore, although it is difficult to evaluate the optimality of such strategies in general, it is interesting to note that at least in a number of closely related problems of nonparametric estimation, such nearest neighbor selection strategies are asymptotically optimal in an appropriate statistical sense [25]. Other possible but more elaborate approaches to optimal neighborhood selection in such problems are described in [26] and [27]. In the simulations presented, only the nearest neighbor selection rule was used.

Choosing the number of nearest neighbors to include in such a selection strategy naturally involves a compromise between modeling error and estimation error. On the one hand, selecting fewer nearest neighbors leads to better localization and, consequently, re-

9

duced modeling error. On the other hand, using more nearest neighbors allows more data to be used in estimating the dynamics, thereby reducing the associated estimation error. As a result, in practice, selection of an appropriate number of nearest neighbors in the modeling requires experience and some trial and error. This issue will be revisited in Section 3.3.

We now address the development of a local model for the dynamics within each neighborhood. Let us denote by

$$\{\mathbf{z}[k], \ k \in \{k_1, k_2, \ldots, k_L\} = \mathcal{K}_*\},$$

the neighborhood in state space about the specified state $\mathbf{x}_*$ according to an L-nearest neighbor selection criterion. Because of the smoothness constraint on the dynamics, we base the interpolation on a locally linear (or, more precisely, affine) model of the form

$$\mathbf{F}(\mathbf{x}[k]) = b_* + \mathbf{a}_*^T \mathbf{x}[k], \qquad \text{for } k \in \mathcal{K}_*, \tag{13}$$

where b_* is a real constant and $\mathbf{a}_*$ is a real n-dimensional vector, both of which are determined by the observations in the neighborhood of $\mathbf{x}_*$.

Rewriting (1) via (10) and (13) as

$$z[k+1] - w[k+1] - u[k] = b_* + \mathbf{a}_*^T (\mathbf{z}[k] - \mathbf{w}[k]), \qquad \text{for } k \in \mathcal{K}_*, \tag{14}$$

where

$$\mathbf{w}[k] = [w[k-n+1], w[k-n+2], \ldots, w[k]]^T,$$

we obtain a set of L equations involving the available data $z[\cdot]$, the random and unknown noises $w[\cdot]$ and $u[\cdot]$ on both sides of the equation, and the parameters b_* and $\mathbf{a}_*$ to be determined. In matrix form, (14) is expressed as

$$\left[\begin{bmatrix} \mathbf{1} & (\mathbf{Z}_* - \mathbf{W}_*) \end{bmatrix} \quad (\mathbf{z}_* - \mathbf{v}_*) \right] \begin{bmatrix} \boldsymbol{\theta}_* \\ \hline -1 \end{bmatrix} = \mathbf{0}, \tag{15}$$

where

$$\mathbf{z}_* = \left[\begin{array}{cccc} z[k_1 + 1] & z[k_2 + 1] & \cdots & z[k_L + 1] \end{array} \right]^T$$

$$\mathbf{v}_* = \left[\begin{array}{cccc} w[k_1 + 1] + u[k_1] & w[k_2 + 1] + u[k_2] & \cdots & w[k_L + 1] + u[k_L] \end{array} \right]^T$$

and

$$\mathbf{Z}_* = \left[\begin{array}{cccc} \mathbf{z}[k_1] & \mathbf{z}[k_2] & \cdots & \mathbf{z}[k_L] \end{array} \right]^T$$

$$\mathbf{W}_* = \left[\begin{array}{cccc} \mathbf{w}[k_1] & \mathbf{w}[k_2] & \cdots & \mathbf{w}[k_L] \end{array} \right]^T,$$

and where $\boldsymbol{\theta}_*$ is a vector of the parameters, *i.e.*,

$$\boldsymbol{\theta}_* = \left[\begin{array}{cc} b_* & \mathbf{a}_*^T \end{array} \right]^T.$$

The appropriate model parameters $\boldsymbol{\theta}_*$ correspond to a solution of the vector equation (15). In general, however, this equation has infinitely many solutions since $\mathbf{v}_*$ and $\mathbf{W}_*$ are effectively unconstrained. However, although $\mathbf{v}_*$ and $\mathbf{W}_*$ are unknown, it is reasonable to assume that they are each comprised of uncorrelated elements. This results from the fact that while the elements in $\mathbf{Z}_*$ correspond to vectors that are close in state space, they will not typically be in close temporal proximity of each other. In particular, they are typically separated by distances longer than the correlation length of the noise $w[n]$, which, for simplicity, we have assumed to be white.

An approach well-matched to the solution of problems of the form (15) where $\mathbf{v}_*$ and $\mathbf{W}_*$ are white is based on the use of a total least squares criterion [24]. With this criterion, the optimum parameters are those which minimize the error terms $\mathbf{v}_*$ and $\mathbf{W}_*$, *i.e.*, the total least squares estimate $\boldsymbol{\theta}_{\mathrm{TLS}}$ is given, for a suitable weight matrix $\boldsymbol{\Lambda}$, by

$$\boldsymbol{\theta}_{\mathrm{TLS}} = \arg \min_{\boldsymbol{\theta}} \left\| \left[\begin{array}{cc} \mathbf{v}_* & \mathbf{W}_* \end{array} \right] \boldsymbol{\Lambda} \right\|_F^2, \tag{16}$$

where the matrix norm $\| \cdot \|_F^2$ refers to the sum of the squares of its elements[2]. When σ_w^2

---

[2]For a matrix M, the norm can be expressed according to $\|M\|_F^2 = \mathrm{tr}(M^T M)$.

and σ_u^2 are known, the appropriate weight matrix is

$$
\mathbf{\Lambda} = \begin{bmatrix} \sigma_w^2 + \sigma_u^2 & 0 & \dots & 0 \\ 0 & \sigma_w^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_w^2 \end{bmatrix}^{-1/2}.
\tag{17}
$$

However, when σ_w^2 and σ_u^2 are unknown, as is frequently the case in practice, $\mathbf{\Lambda}$ is often initially chosen to be the diagonal matrix $\mathrm{diag}(2, 1, \dots, 1)$ and then subsequently refined through experimentation.

Given a suitable weight matrix, the total least squares parameter estimates may be computed efficiently via singular value decomposition methods [24][28], as we now illustrate. For simplicity, we consider the case $\mathbf{\Lambda} = \mathbf{I}$; the extension to more general weight matrices is straightforward. Proceeding, the minimization in (16) requires that the submatrix

$$
\begin{bmatrix} \mathbf{Z}_* & \mathbf{z}_* \end{bmatrix}
$$

in (15) be perturbed in a minimum-norm sense such that the full matrix

$$
\begin{bmatrix} \mathbf{1} & \mathbf{Z}_* & \mathbf{z}_* \end{bmatrix}
\tag{18}
$$

has a null space. We assume that the matrix (18) has full column rank, since, if it did not, then it already would have a null space, and an exact solution to (15) could be found. Then, given the matrix

$$
\mathbf{A}_* = \begin{bmatrix} \mathbf{1} & \mathbf{Z}_* \end{bmatrix}
\tag{19}
$$

and the vector $\mathbf{z}_*$, we perform the following Householder transformation,

$$
\mathbf{Q}^T \begin{bmatrix} \mathbf{A}_* & \mathbf{z}_* \end{bmatrix} = \begin{bmatrix} r_{11} & \mathbf{r}_{12}^T & r_{13} \\ \mathbf{0} & \mathbf{R}_{22} & \mathbf{r}_{23} \end{bmatrix}.
\tag{20}
$$

12

From this factorization, the total least squares solution is readily constructed as

$$\boldsymbol{\theta}_{\text{TLS}} = \left(\mathbf{A}_*^T \mathbf{A}_* - \sigma^2 \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \right)^{-1} \mathbf{A}_*^T \mathbf{z}_*, \tag{21}$$

where σ is the smallest singular value of $\mathbf{R}_{22}$ in (20). This closed-form expression for the solution vector is given here for comparison with the similar form for the least-squares solution. However, as with the least-squares problem, more efficient techniques exist for solving this total least squares problem, and these techniques were used in the simulations. See, for example [24] and [28].

3.1.1 Case: Noise-free Observations

Some useful insights are gained from considering the special case in which $w[k] = 0$, corresponding to noise-free observations [29]. In this case, $\mathbf{z}[k] = \mathbf{x}[k]$, *i.e.*, the states $\mathbf{x}[k]$ can be constructed perfectly from the data and consequently only the left-hand side of (14) contains noise. In terms of the parameter estimation problem, substituting $\sigma_w^2 = 0$ into (17) indicates that the appropriate weight matrix $\boldsymbol{\Lambda}$ places infinite weight on the error matrix $\mathbf{W}_*$, thus forcing a solution to (16) in which $\mathbf{W}_*$ is identically zero, while $\|\mathbf{v}_*\|^2$ is minimized. Clearly, the total least squares solution specializes to a more familiar least squares solution in this noise-free scenario. In particular, we can write

$$\boldsymbol{\theta}_{\text{TLS}} = \boldsymbol{\theta}_{\text{LS}} = \arg\min_{\boldsymbol{\theta}} \|\mathbf{z}_* - \mathbf{A}_* \boldsymbol{\theta}\|^2.$$

with $\mathbf{A}_*$ as given in (19). The solution to this least-squares problem is [24]

$$\boldsymbol{\theta}_{\text{LS}} = (\mathbf{A}_*^T \mathbf{A}_*)^{-1} \mathbf{A}_*^T \mathbf{z}_*.$$

In this case, our algorithm specializes to one used by Farmer and Sidorowich for modeling chaotic signals [6]. Note that if we further let the number of nearest neighbors L in our model be the entire data set and constrain $b_* = 0$, the resulting least squares estimates correspond to linear prediction coefficients and can be computed very efficiently using Levinson's

recursion [18].

3.2 Experiments

In this section, we present preliminary results on the performance of the modeling algorithms outlined in Section 3.1 applied to simulated data. In particular, we apply the total least squares modeling algorithms to NLAR processes embedded in additive stationary white Gaussian noise to a prescribed SNR. Two processes were considered: the nonlinear feedback process given by (4), and the chaotic Henon process (9).

A convenient measure of modeling error for these simulations is given by the normalized RMS error

$$\sqrt{\frac{\sigma_e^2 - \sigma_u^2}{\sigma_y^2 - \sigma_u^2}}, \tag{22}$$

where σ_y^2 and σ_u^2 are the variances of $y[k]$ and $u[k]$, respectively, and where

$$\sigma_e^2 = E\left[\left(\mathbf{F}(\mathbf{x}) - \hat{\mathbf{F}}(\mathbf{x})\right)^2\right]. \tag{23}$$

In (23), $E[\cdot]$ denotes an ensemble average over all states $\mathbf{x}$ with respect to their relative frequency of occurrence. Due to the ergodic properties of the processes we consider, it suffices to perform such averaging over states generated by a typical sample function of the process of interest. The experiments were conducted as follows. A sample function from an NLAR process was generated, and, from this signal, a collection of state vectors was extracted. The NLAR process was then embedded in noise to generate the observed data. For each state vector $\mathbf{x}$ extracted from the original signal, the total least squares algorithm is applied to the noisy data to estimate $\mathbf{F}(\mathbf{x})$. The squared errors in each of these estimates are then averaged to produce an approximation to (23). Note that because of the autoregressive structure of the NLAR model, each $\mathbf{F}(\mathbf{x})$ represents a one-step prediction, and thus the modeling error can be interpreted, in some sense, as a form of nonlinear prediction error.

The normalization in (22) is chosen so that the error metric is useful in a variety of comparisons. In particular, a worst case signal estimate equal to the mean of the process would result in a modeling error of unity since $\sigma_e^2 = \sigma_y^2$. Similarly, a best case signal

estimate equal to $\mathbf{F}(\mathbf{x}[k])$ for state $\mathbf{x}[k]$ would result in a zero modeling error. Thus, the RMS modeling error provides a measure of distance from the best achievable performance.

In Fig. 1, we plot normalized RMS modeling error as a function of the SNR of the observations for a fixed data length of $K = 10\,000$ samples. Fig. 1(a) depicts the modeling performance for the feedback NLAR process using a neighborhood size of $L = 100$. Fig. 1(b) depicts the corresponding performance for the case of the Henon process, using a neighborhood size of $L = 100$. The results reflect the monotonic decrease in modeling error with SNR expected, and suggest that for the data lengths involved, reasonable modeling can be achieved provided the SNR of the data exceeds 10 dB.

In Fig. 2, we plot the RMS modeling error as a function of data length at a fixed SNR of 15 dB for the feedback and Henon processes, using neighborhood sizes of $L = 100$. As expected, greater data lengths generally give rise to improved modeling performance, although there are some statistical deviations from monotonicity. It is interesting to note that with increasing data lengths, the improvement in modeling performance becomes increasingly gradual. This suggests that reducing modeling error below a certain threshold may be inherently impractical at any given SNR.

In the case of noise-free observations it is possible to compare our modeling method to the benchmark linear prediction modeling algorithm. In Fig. 3, we plot modeling error as a function of data length. In this case, our modeling method specializes to the locally affine least-squares algorithm discussed in Section 3.1.1, and the modeling error can be interpreted as a true nonlinear prediction error. Results are presented for both the feedback and Henon processes, with a neighborhood size of $L = 100$ for each. In both cases, it is apparent that the linear prediction algorithm is unable to provide adequate modeling due to the nonlinear structure and broadband spectra of the processes involved. However, the local least-squares modeling becomes increasingly accurate with increasing data length, reaching acceptable levels within a few thousand samples.

3.3 Parameter Selection

An important aspect of NLAR modeling is choosing the appropriate number of nearest neighbors to use in the local modeling. While this invariably involves some trial and er-
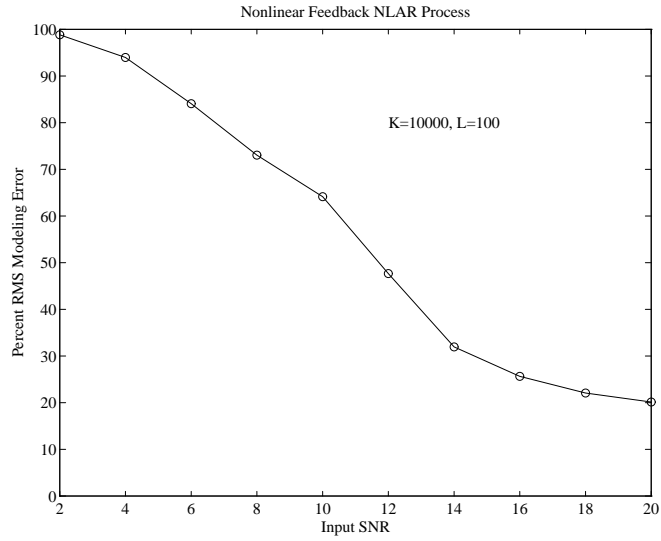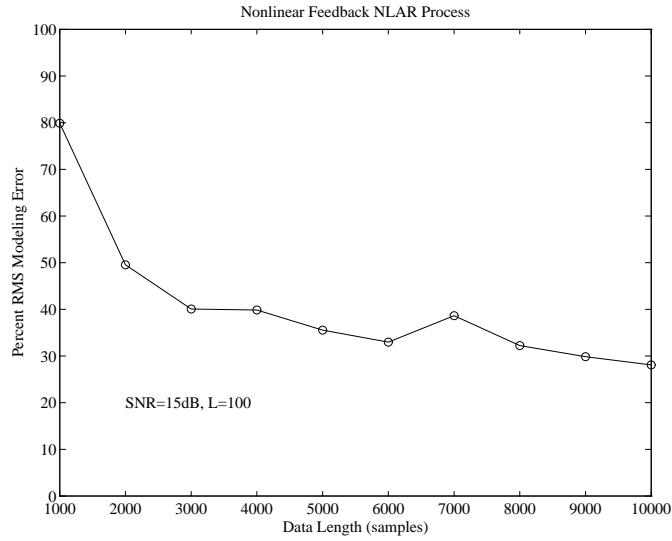
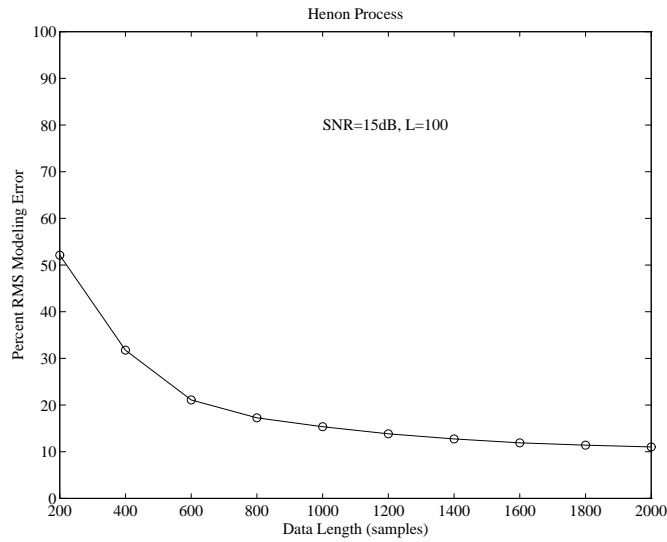(a) Case: Nonlinear feedback NLAR process.
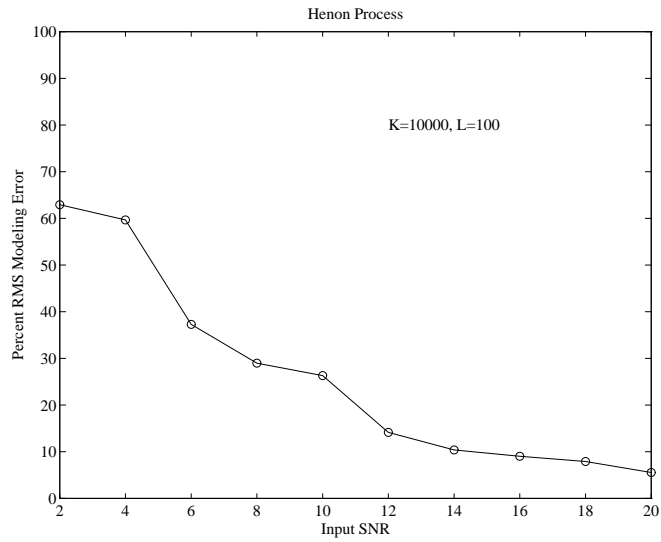
(b) Case: Henon process.

Figure 1: *NLAR modeling error as a function of SNR of the observations. The data length is fixed to $K = 10\,000$ samples. The actual data points are indicated by circles; connecting lines are provided as visual aides only.*
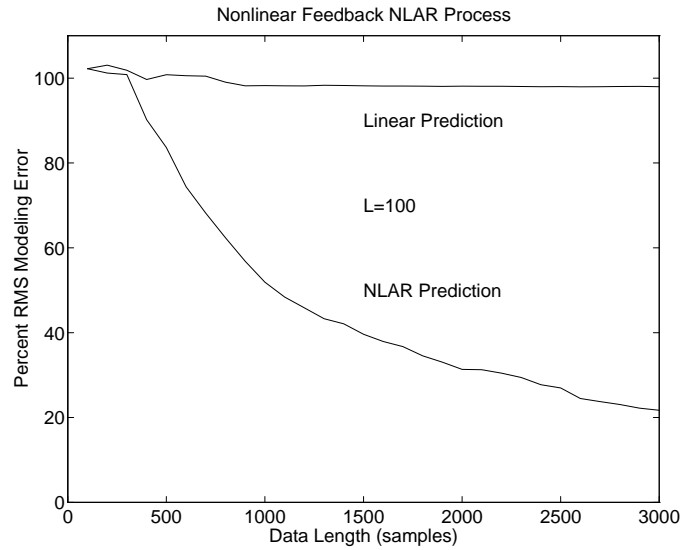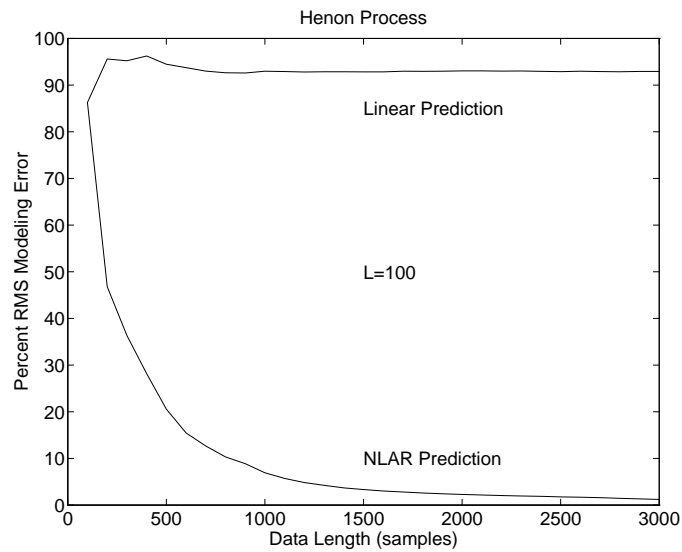
(a) Case: Nonlinear feedback NLAR process.

(b) Case: Henon process.

Figure 2: *NLAR modeling error as a function of data length of the observations. The SNR of the observations is fixed to 15 dB. The actual data points are indicated by circles; connecting lines are provided as visual aides only.*

(a) Case: Nonlinear feedback NLAR process.

(b) Case: Henon process.

Figure 3: *NLAR modeling error (i.e., nonlinear prediction error) as a function of data length of the observations. The observations are noise free. The performance of linear prediction is illustrated for comparison purposes only.*
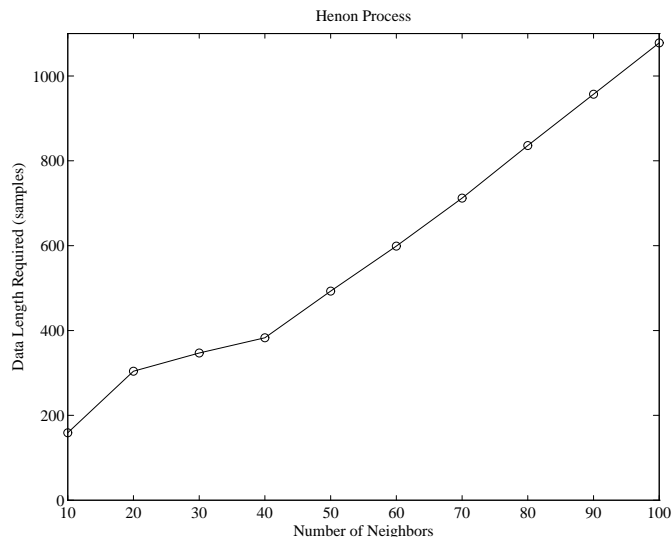
Figure 4: *The data lengths required to achieve a fixed 10% RMS modeling error are shown as a function of the number of neighbors, L. The actual data points are indicated by circles; connecting lines are provided as visual aides only.*

ror, some insight is provided by considering the following approximation to the relationship between data length, neighborhood size, and estimation error. For the purposes of illustration, we describe the result for the noise-free case. Since the dynamics are nonlinear, modeling error is largely a function of how well each local region is modeled as linear—in particular, the greater the extent of the local region, the greater the modeling error. Hence, by fixing the volume of the local regions, we effectively fix the modeling error. However, the volume of each local region is roughly proportional to L/N, where, again, L is the number of nearest neighbors and N is the data length, and where the constant of proportionality depends on the particular nonlinear dynamics being modeled. When L is fixed, we see that modeling error decreases inversely with N. Alternatively, for a fixed modeling error, the required data length N grows in proportion to L. This general behavior is confirmed by empirical studies. In Fig. 4, for example, the data lengths required to achieve a fixed 10% RMS modeling error with the Henon process are shown as a function of the number of neighbors, L, used in the local modeling procedure.

Another important aspect of NLAR modeling is selection of the model order, n. While in this paper we have assumed the model order to be known, in practice it is desirable to

estimate the appropriate model order. Traditional techniques for model order estimation include the minimum description length (MDL) criterion [30] [31] and the Akaike information criterion (AIC) [32] [18] . For AR processes, the MDL and AIC criteria specialize, respectively, to

$$\text{MDL}(n) = \ln e^2(n) + \frac{2n}{N_e}, \tag{24}$$

and

$$\text{AIC}(n) = \ln e^2(n) + \frac{n \ln N}{N_e}, \tag{25}$$

where $e^2(n)$ is the nth-order AR prediction error, N is the number of data samples, and N_e is the effective number of samples due to data windowing.

For NLAR modeling, it is reasonable to consider using these methods in conjunction with the local modeling procedure developed in this paper. In practice, replacing $e^2(n)$ with the NLAR prediction or modeling error appears to work well, trading off complexity of increased model order against decreasing prediction error. In particular, in very preliminary empirical studies with the Henon process, both the MDL and AIC exhibit unique minima at $n = 4$. For specifically chaotic processes, many other algorithms for determining appropriate model orders can also be applied. For example, standard algorithms for determining fractal dimension and minimum embedding dimension [33][34][22] [2] can be used in principle. In any case, further study of model order selection is an interesting and important area of investigation for future research, but beyond the scope of this paper.

4   NLAR Signal Estimation

In Section 3, we developed a total least squares algorithm for estimating the model of an NLAR signal embedded in white measurement noise. Signal estimation turned out to be a key component of the modeling algorithm. Indeed, the process of obtaining parameter estimates via (16) simultaneously led to estimates of the noises $\mathbf{v}_*$ and $\mathbf{W}_*$, which in principle can also be used to provide estimates of the underlying signal.

In this section, however, we explicitly address the problem of signal estimation, in particular, the problems of signal filtering, prediction and smoothing. Our approach is based on an extended Kalman filter formulation, and we explicitly utilize the total least squares mod-

eling algorithms developed in Section 3. The algorithms we develop are computationally efficient and highly practical in a number of application scenarios.

We remark in advance that few properties of the joint modeling-estimation algorithm can be developed analytically either for NLAR processes in general or for chaotic processes in particular. Indeed, even with perfectly known dynamics, the analysis of extended Kalman algorithms by themselves is extremely complicated. When jointly estimated dynamics are involved, the difficulties are compounded. While a thorough study of the convergence and stability characteristics of the joint modeling-estimation algorithm in this context is an important area of research, it is beyond the scope of the current paper. Nevertheless, we will discuss those aspects of the algorithm which have proved important in our experience with the algorithm.

Throughout the remainder of the paper, it will be convenient to rephrase our problem of estimating signal samples $y[k]$ into one of estimating state samples $\mathbf{x}[k]$. For this reason, we express the signal generation and measurement model in the following state space form

$$
\begin{array}{rcll}
\mathbf{x}[k+1] & = & \tilde{\mathbf{F}}(\mathbf{x}[k]) + \mathbf{e}_0\, u[k] & \text{(26a)} \\
z[k] & = & \mathbf{e}_0^T\, \mathbf{x}[k] + w[k] & \text{(26b)}
\end{array}
$$

where $\mathbf{e}_0$ is the unit vector (7), and where $\tilde{\mathbf{F}}$ is derived from $\mathbf{F}$ in accordance with (2). Furthermore, because $y[k]$ and $\mathbf{x}[k]$ are related through (3), we obtain signal estimates $\hat{y}[k]$ from state estimates $\hat{\mathbf{x}}[k]$ via the projection

$$
\hat{y}[k] = \mathbf{e}_0^T\, \hat{\mathbf{x}}[k].
$$

The estimation algorithms we derive are all based on a common linearization strategy. Specifically, for states $\mathbf{x}$ in a neighborhood of some arbitrary point $\mathbf{x}_*$ in state space, we linearize (26a) according to

$$
\tilde{\mathbf{F}}(\mathbf{x}) \approx \nabla\tilde{\mathbf{F}}(\mathbf{x}_*)\,(\mathbf{x} - \mathbf{x}_*) + \tilde{\mathbf{F}}(\mathbf{x}_*), \tag{27}
$$

where $\nabla\tilde{\mathbf{F}}(\mathbf{x})$ denotes a Jacobian matrix whose (i,j)th component is the partial derivative of

the ith component of $\tilde{\mathbf{F}}(\mathbf{x})$ with respect to the jth component of $\mathbf{x}$. Given this linearization, we may then apply extended Kalman filtering in a relatively straightforward manner[3].

In order to compute the linearization (27) as required during signal estimation, we naturally turn to the modeling algorithms derived in Section 3 and apply them to the appropriate segment of the measurement data. In particular, for $\mathbf{x}$ near $\mathbf{x}_*$, we approximate $\tilde{\mathbf{F}}(\mathbf{x})$ by

$$\tilde{\mathbf{F}}(\mathbf{x}) \approx \left[\begin{array}{c|c} \mathbf{0} & \mathbf{I} \\ \hline & \mathbf{a}_*^T \end{array}\right] \mathbf{x} + \left[\begin{array}{c} \mathbf{0} \\ \hline b_* \end{array}\right] \tag{28}$$

where $\mathbf{a}_*$ and b_* are total least squares parameter estimates corresponding the neighborhood of state $\mathbf{x}_*$. By comparison with (27) we see that the corresponding approximation to the Jacobian $\nabla\tilde{\mathbf{F}}(\mathbf{x}_*)$ is given by the companion form matrix in (28). Hence, the signal modeling algorithms will play a complementary and crucial role in the signal estimation algorithms we develop.

4.1   Filtering

In this section, we consider the problem of signal filtering, *i.e.*, estimating an underlying NLAR signal using causal processing of the noisy data. The corresponding algorithms are most useful in scenarios where the data must be processed sequentially and in real time.

We use the notation $\hat{\mathbf{x}}[i|j]$ to denote an estimate of $\mathbf{x}[i]$ based on data $z[0]$, $z[1]$, ..., $z[j]$, and $\mathbf{P}[i|j]$ will denote the anticipated mean-square error between $\hat{\mathbf{x}}[i|j]$ and $\mathbf{x}[i]$. In this section, we specifically seek to produce the estimates

$$\hat{\mathbf{x}}[n-1|n-1],\ \hat{\mathbf{x}}[n|n],\ \ldots,\ \hat{\mathbf{x}}[K-1|K-1].$$

Applying a discrete-time extended Kalman filter (EKF) [35] to (26) based on the dynamic linearization (27), we recursively obtain $\hat{\mathbf{x}}[k|k]$ from $\hat{\mathbf{x}}[k-1|k-1]$ and the newly available data sample $z[k]$ in two steps. In the first step, we produce a state estimate based

---

[3]As an aside, we remark that this linearization is also the basis for some closely related algorithms developed by Richard [13] for processing chaotic signals. However, in our case, we not only apply this idea to the broader class of NLAR processes, but also avoid the restrictive assumptions used in [13] that the nonlinear dynamics be either known *a priori* or available through some noise-free reference signal.

on a one-step prediction together with its associated prediction error covariance, *i.e.*,

$$\hat{\mathbf{x}}[k|k-1] \;=\; \tilde{\mathbf{F}}(\hat{\mathbf{x}}[k-1|k-1]) \tag{29a}$$

$$\mathbf{P}[k|k-1] \;=\; \nabla\tilde{\mathbf{F}}(\hat{\mathbf{x}}[k-1|k-1])\,\mathbf{P}[k-1|k-1]\,\nabla\tilde{\mathbf{F}}(\hat{\mathbf{x}}[k-1|k-1])^T \;+$$

$$\mathrm{diag}(0,\ldots,\sigma_u^2). \tag{29b}$$

In the second step, we then update the state estimate and its covariance using the new data sample $z[k]$ via

$$\hat{\mathbf{x}}[k|k] \;=\; \hat{\mathbf{x}}[k|k-1] + \mathbf{g}[k]\,(z[k] - \mathbf{e}_0^T\,\hat{\mathbf{x}}[k|k-1]) \tag{29c}$$

$$\mathbf{P}[k|k] \;=\; \left[\mathbf{I} - \mathbf{g}[k]\,\mathbf{e}_0^T\right]\,\mathbf{P}[k|k-1], \tag{29d}$$

where the Kalman gain vector is given by

$$\mathbf{g}[k] = \mathbf{P}[k|k-1]\,\mathbf{e}_0 \;\Big/\; \left(\mathbf{e}_0^T\,\mathbf{P}[k|k-1]\,\mathbf{e}_0 + \sigma_w^2\right). \tag{29e}$$

To initialize the algorithm, it suffices to choose the initial one-step prediction state estimates to be the first block of available data, in which case the initial error variance in each element is given by the observation noise variance, *viz.*,

$$\hat{\mathbf{x}}[n-1|n-2] \;=\; \left[\begin{array}{cccc} z[0] & z[1] & \cdots & z[n-1] \end{array}\right]^T \tag{29f}$$

$$\mathbf{P}[n-1|n-2] \;=\; \sigma_w^2\,\mathbf{I}. \tag{29g}$$

Collectively eqns. (29), when used in conjunction with the total least squares modeling algorithm, constitute the EKF algorithm for NLAR processes. Note that this algorithm makes explicit use of both σ_w^2 and σ_u^2. While these are rarely known *a priori,* a number of methods can be used to obtain reliable estimates of these parameters empirically.

It is important to emphasize that the error covariance matrices computed in the EKF algorithm (29) are not the actual error matrices. Indeed, while the Kalman filter structure is an optimal (minimum mean-square error) estimation structure for linear systems,

its extension to nonlinear systems is generally not optimal. As a consequence, the error covariance matrices are at most crude estimates of the actual error.

4.2 Prediction

Next, we address signal prediction—the extrapolation of an NLAR signal beyond some currently available set of measurements. These algorithms are particularly useful in applications where there is a fixed measurement or processing delay which forces an estimate of a particular sample of the process to be produced before the corresponding noisy measurement is available. They are also useful in applications that require extrapolation of the data outside the available noisy data record. As we shall see, such signal prediction can be accomplished using a straightforward extension of the signal filtering algorithm of Section 4.1, consistent with traditional Kalman filter theory.

In the m-step prediction scenario, we seek to produce, for some $m \geq 1$,

$$\hat{\mathbf{x}}[n-1+m|n-1], \, \hat{\mathbf{x}}[n+m|n], \, \ldots, \, \hat{\mathbf{x}}[K-1+m|K-1],$$

The extended Kalman predictor (EKP) appropriate to this problem [35] [36] amounts to a two-part process. First, for each k we compute $\hat{\mathbf{x}}[k|k]$ using the EKF algorithm (29) which exploits all the available data. Then, since we have no further data from which to extract information, we propagate $\hat{\mathbf{x}}[k|k]$ forward based on the current knowledge of the dynamics. Specifically, we may recursively compute the prediction according to

$$\hat{\mathbf{x}}[k+m|k] = \tilde{\mathbf{F}}(\hat{\mathbf{x}}[k+m-1|k]) \tag{30a}$$

for $m \geq 2$. The initial estimate $\hat{\mathbf{x}}[k+1|k]$ corresponding to $m = 1$ is obtained directly from $\hat{\mathbf{x}}[k|k]$ using the step (29a) in the EKF algorithm.

The error covariance matrix associated with this estimate can be computed recursively according to

$$\begin{aligned}
\mathbf{P}[k+m|k] &= \nabla \tilde{\mathbf{F}}(\hat{\mathbf{x}}[k+m-1|k]) \left[\mathbf{P}[k+m-1|k] + \mathrm{diag}(0,\ldots,0,\sigma_u^2) \right] \\
&\quad \nabla \tilde{\mathbf{F}}(\hat{\mathbf{x}}[k+m-1|k])^T, \tag{30b}
\end{aligned}$$

24

where the initial estimate $\mathbf{P}[k+1|k]$ corresponding to $m = 1$ is obtained from (29b). Again, in general, one would expect (30b) to be an at least crude estimate of the actual m-step prediction error. Collectively (29) and (30) constitute the EKP prediction algorithm for NLAR processes.

4.3 Smoothing

Finally, in this section we develop algorithms for fixed-interval signal smoothing, *i.e.*, for estimating an underlying signal given the complete record of noisy measurements. Because these algorithms involve non-causal processing, they are generally used off-line and provide improved signal recovery performance when such processing constraints can be tolerated.

The specific estimates we seek in this case are

$$\hat{\mathbf{x}}[n-1|K-1], \, \hat{\mathbf{x}}[n|K-1], \, \ldots, \, \hat{\mathbf{x}}[K-1|K-1].$$

An algorithm for constructing such estimates is based on the use of a fixed-interval extended Kalman smoother (EKS) in Rauch-Tung-Striebel form [35] [36], and involves two stages of processing. In the first stage, frequently referred to as the forward filtering stage, the causal EKF algorithm (29) described in Section 4.1 is applied to the data to produce the estimates $\hat{\mathbf{x}}[k|k]$ for $k = n-1, \, n, \, \ldots, \, K-1$. In the second stage, each of these estimates is refined to produce the corresponding smoothed estimate $\hat{\mathbf{x}}[k|K-1]$, as we now describe.

The reverse recursion for obtaining the $\hat{\mathbf{x}}[k|K-1]$ from both $\hat{\mathbf{x}}[k+1|K-1]$ and the previously stored estimates from the EKF algorithm is given by

$$\hat{\mathbf{x}}[k|K-1] = \hat{\mathbf{x}}[k|k] + \mathbf{A}[k] \left(\hat{\mathbf{x}}[k+1|K-1] - \hat{\mathbf{x}}[k+1|k] \right) \tag{31a}$$

where

$$\mathbf{A}[k] = \mathbf{P}[k|k] \, \nabla \tilde{\mathbf{F}}(\hat{\mathbf{x}}[k|k]) \, \mathbf{P}^{-1}[k|k-1]. \tag{31b}$$

Likewise, the error covariance matrix $\mathbf{P}[k|K-1]$ is obtained from $\mathbf{P}[k+1|K-1]$ and the previously stored error matrices from the EKF algorithm according to

$$\mathbf{P}[k|K-1] = \mathbf{P}[k|k] + \mathbf{A}[k] \left(\mathbf{P}[k+1|K-1] - \mathbf{P}[k+1|k] \right) \mathbf{A}[k]^T. \tag{31c}$$

Due to the nature of the recursion, this is generally referred to as the backward filtering stage. Furthermore, it's initialization from the estimates from the EKF algorithm is straightforward since both $\hat{\mathbf{x}}[K-1|K-1]$ and $\mathbf{P}[N|N]$ are directly available.

Collectively, eqns. (29) and (31) constitute our EKS algorithm for NLAR processes. As with the case of the EKF algorithm, it is impossible to attach a strong interpretation to the error covariance matrices. However, these estimation algorithms rely on such error information and nevertheless perform remarkably well in practice. This suggests that these matrices provide an at least crude estimate of the quality of the estimates being produced.

4.4   Experiments

In this section, we illustrate the performance of the signal estimation algorithms applied to noisy measurements of the two NLAR processes corresponding to (4) and (9). In the filtering and smoothing cases, an appropriate performance metric is the SNR "gain" in the signal estimate, which we denote by

$$\Delta\mathrm{SNR} = \mathrm{SNR}_o - \mathrm{SNR}_i$$

where the SNR_o is the output SNR

$$\mathrm{SNR}_o = 10\log_{10}\frac{\sigma_y^2}{E\left[(\hat{y}[k] - y[k])^2\right]}$$

and SNR_i is the SNR of the observations, *i.e.*,

$$\mathrm{SNR}_i = 10\log_{10}\frac{\sigma_y^2}{\sigma_w^2}.$$

In the case of prediction, the notion of SNR gain is less natural, so we instead use $-\mathrm{SNR}_o$, the mean-square prediction error, as a measure of performance. In each of the experiments, a data length of $K = 10\,000$ samples was used, and $L = 500$ nearest neighbors were used in the total least squares modeling algorithm.

Fig. 5 depicts the SNR gain achieved by filtering both the feedback and Henon processes. We note that over the full 20 dB variation in input SNR, the SNR gain is relatively constant.

Furthermore, the SNR gain is rather modest, which suggests that relatively little signal enhancement is attained through filtering alone.

In Fig. 6, mean-square m-step prediction error is plotted as a function of the SNR of the measurements for the cases $m = 1$ and $m = 2$. In each case, the prediction error (in dB) decreases roughly linearly with input SNR. We also note that for both the feedback and Henon processes, m-step prediction performance degrades rapidly with m. This is principally a consequence of the locally unstable dynamics (and, hence, positive Lyapunov exponents [1]) associated with the nonlinear dynamics involved.

Fig. 7 depicts the SNR gain of the NLAR signal estimates obtained from the smoothing algorithm as a function of the SNR of the data. In general, we observe increasing SNR gain with input SNR. Furthermore, by comparing these results with those of Fig. 5, we note that the additional backward filtering in the smoothing algorithm leads to a substantial improvement in overall signal estimation performance. In the case of the Henon process, this extra stage of processing yields nearly 6 dB extra performance at 20 dB SNR.

Finally, in Fig. 8, we present a signal smoothing example. Fig. 8(a) illustrates the state pairs $(y[k], y[k-1])$ associated with a 1 000-sample segment of a $K = 10\,000$-sample Henon process. Fig. 8(b) shows the corresponding pairs $(z[k], z[k-1])$ of the signal embedded in white Gaussian noise to the level of 15 dB. Fig. 8(c) illustrates the signal restoration obtained by applying the smoothing algorithm to the noisy data, which corresponds to an SNR gain of 7.9 dB in the Henon signal. For the purposes of comparison, Fig. 8(d) depicts the signal restoration obtained by applying a conventional linear Wiener filter to the observations. In this case, virtually no SNR gain is achieved despite the fact explicit knowledge of the dynamics was used in the design of the linear Wiener filter. This results from the fact that, although deterministic, the Henon process is spectrally broadband.

5   Concluding Remarks

We have discussed a general framework for a class of nonlinear signal models that are inherently well-suited for a wide range of physical and man-made phenomena. Furthermore, we have derived some robust, effective, and computationally efficient modeling and estimation algorithms for applying these models to data. The modeling algorithms are based on locally
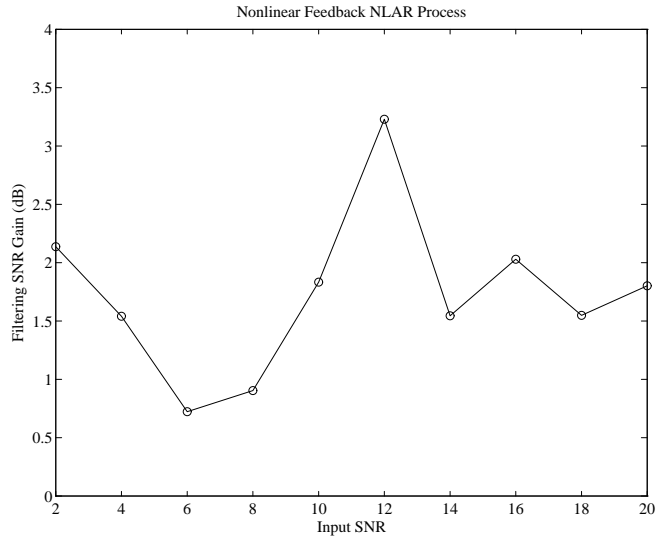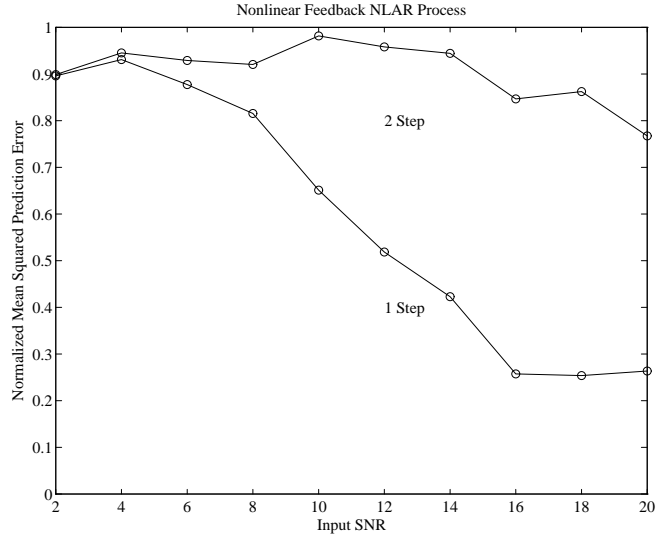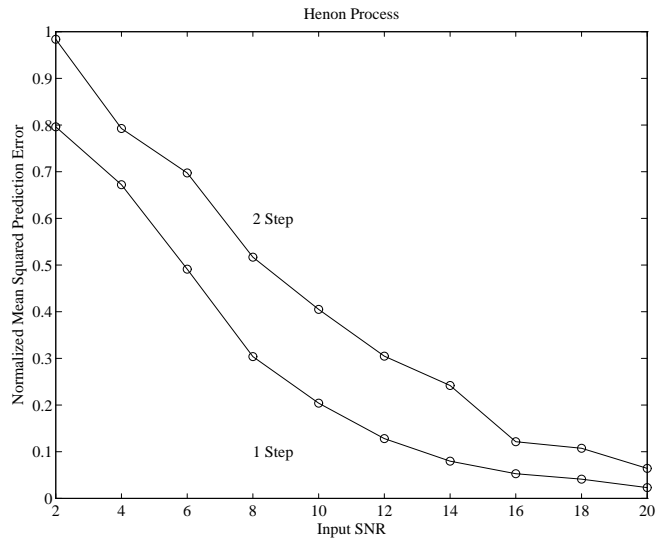
(a) Case: Nonlinear feedback NLAR process.
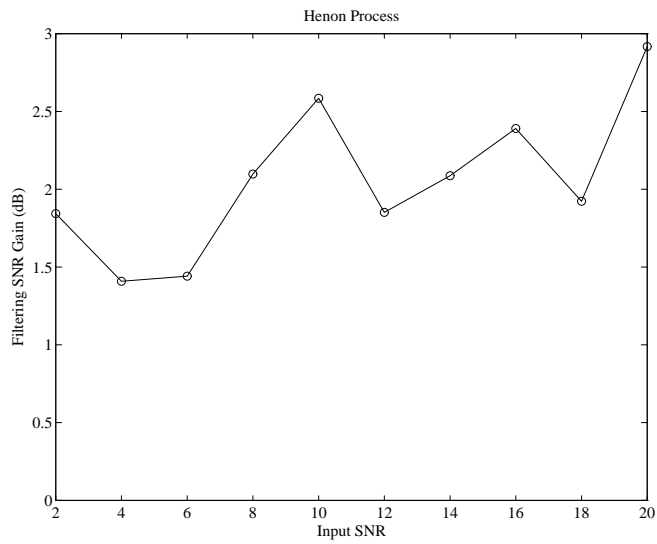
(b) Case: Henon process.

Figure 5: *SNR gain in the NLAR signal estimates obtained from the filtering algorithm as a function of the SNR of the data. The actual data points are indicated by circles; connecting lines are provided as visual aides only.*
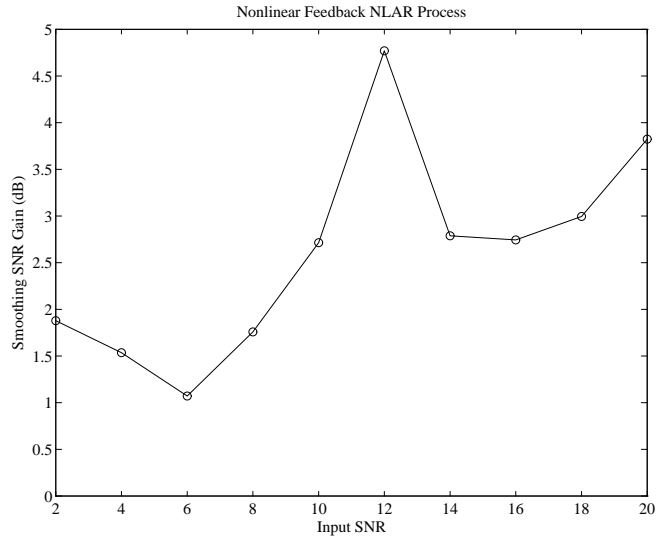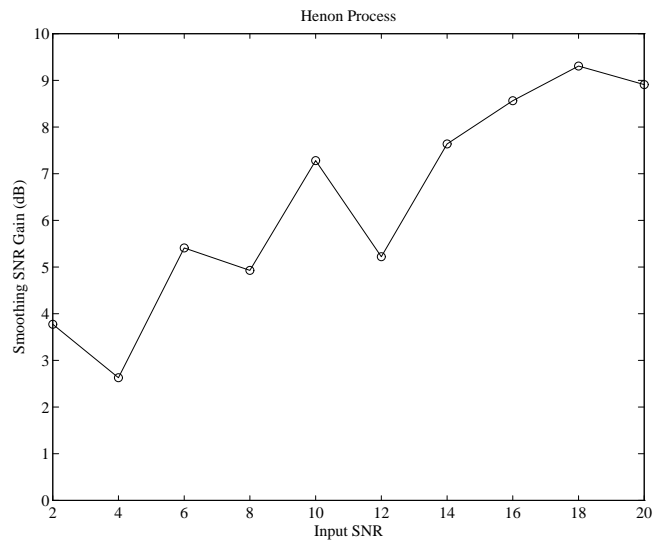
(a) Case: Nonlinear feedback NLAR process.

(b) Case: Henon process.

Figure 6: *Mean squared prediction error in the m-step NLAR signal predictions obtained from the prediction algorithm as a function of the SNR of the data. The actual data points are indicated by circles; connecting lines are provided as visual aides only.*
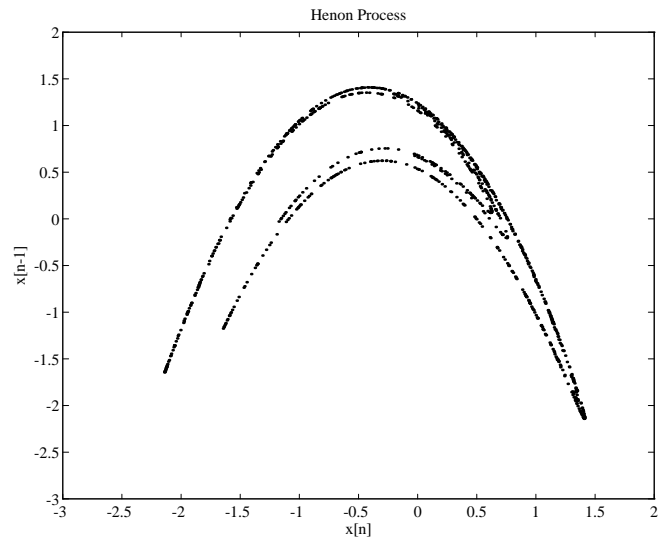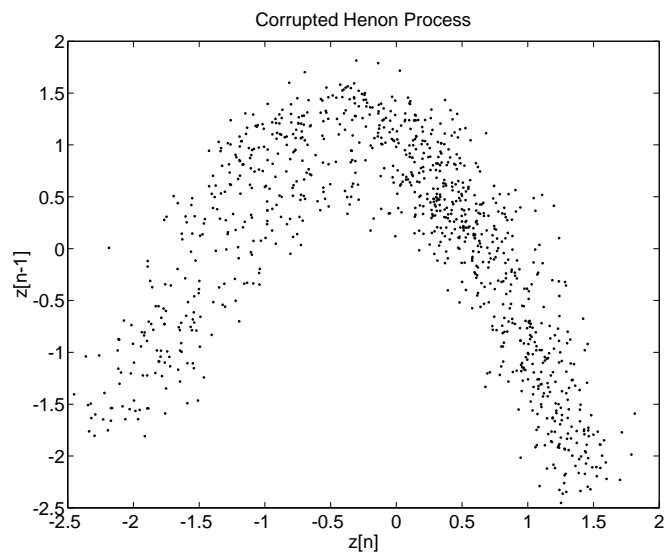
(a) Case: Nonlinear feedback NLAR process.

(b) Case: Henon process.

Figure 7: *SNR gain in the NLAR signal estimates obtained from the smoothing algorithm as a function of the SNR of the data. The actual data points are indicated by circles; connecting lines are provided as visual aides only.*
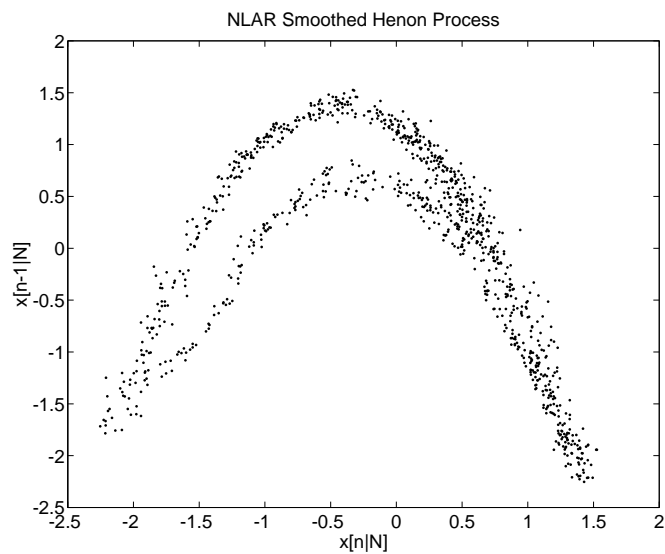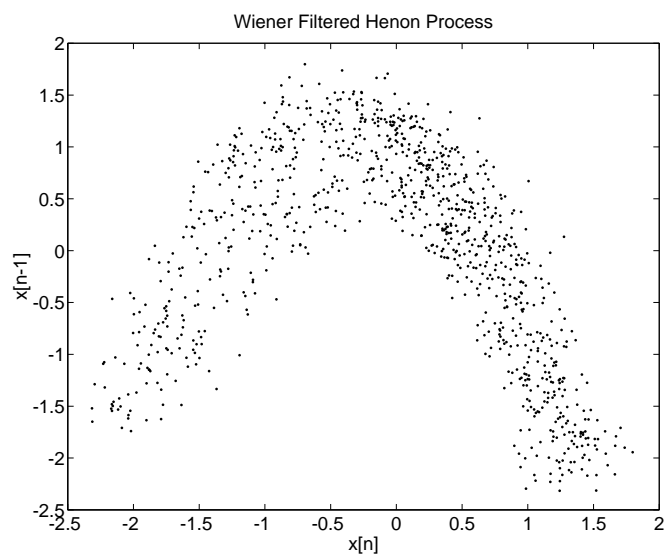
(a) Original Henon signal.

(b) Signal embedded in white noise to an SNR of 15 dB.

Figure 8: *Henon process smoothing example.*

31

NLAR Smoothed Henon Process

(c) Signal restoration via NLAR smoothing algorithm. SNR gain is 7.9 dB.

Wiener Filtered Henon Process

(d) Signal restoration via linear Wiener filter.

Figure 8: *Continued.*

affine modeling of the dynamics in state space from noisy-corrupted measurements using a total least squares criterion. The estimation algorithms combine total least squares modeling with extended Kalman filtering, prediction, and smoothing algorithms. The viability of the resulting algorithms, which effectively involve local processing both temporally and in state space, is demonstrated on synthesized chaotic and other NLAR data corrupted by additive stationary white Gaussian noise.

References

[1] F. C. Moon, *Chaotic Vibrations*. New York, NY: John Wiley and Sons, 1987.

[2] H. D. I. Abarbanel, "Chaotic signals and physical systems," in *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 1992.

[3] A. V. Oppenheim, G. W. Wornell, S. H. Isabelle, and K. M. Cuomo, "Signal processing in the context of chaotic signals," in *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 1992.

[4] L. M. Pecora and T. L. Carroll, "Synchronization in chaotic systems," *Phys. Rev. Lett.*, vol. 64, pp. 821–824, Feb. 1990.

[5] S. H. Isabelle, A. V. Oppenheim, and G. W. Wornell, "Effects of convolution on chaotic signals," in *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 1992.

[6] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Phys. Rev. Lett.*, vol. 59, p. 845, 1987.

[7] M. Casdagli, "Nonlinear prediction of chaotic time series," *Physica D*, vol. 35, pp. 335–356, 1989.

[8] S. Haykin and H. Leung, "Model reconstruction of chaotic dynamics: First preliminary radar results," in *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 1992.

[9] C. S. Hsu, *Cell-to-Cell Mapping*. New York, NY: Springer-Verlag, 1987.

[10] J. D. Farmer and J. J. Sidorowich, "Optimal shadowing and noise reduction," *Physica D*, 1990. Submitted for publication.

[11] S. M. Hammel, "A noise reduction method for chaotic systems," *Physics Letters A*, vol. 148, pp. 421–428, Sept. 1990.

[12] P. F. Marteau and H. D. I. Abarbanel, "Noise reduction in chaotic time series using scaled probabilistic methods," Institute for Nonlinear Science preprint, University of California, San Diego, CA, 1990.

[13] M. D. Richard, "Probabilistic state estimation with discrete-time chaotic systems," RLE Tech. Rep. No. 571, M. I. T., Cambridge, MA, Mar. 1992.

[14] C. Myers, S. Kay, and M. Richard, "Signal separation for nonlinear dynamical systems," in *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 1992.

[15] C. Myers, A. Singer, B. Shin, and E. Church, "Modeling chaotic systems with hidden Markov models," in *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 1992.

[16] H. C. Papadopoulos and G. W. Wornell, "Exact maximum likelihood estimation of a class of chaotic signals," *IEEE Trans. Inform. Theory*, Feb. 1993. Submitted for publication.

[17] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, "State space reconstruction in the presence of noise," *Physica D*, vol. 51, pp. 52–98, 1991.

[18] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, 1975.

[19] J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 26, pp. 197–210, 1978.

[20] E. Weinstein, M. Feder, and A. Oppenheim, "Signal enhancement using single and multiple-sensor measurements," RLE Tech. Rep. No. 560, M. I. T., Cambridge, MA, Dec. 1990.

[21] F. Takens, "Detecting strange attractors in turbulence," in *Lecture Notes in Mathematics 898* (D. Rand and L. S. Young, eds.), p. 366, New York, NY: Springer-Verlag, 1981.

[22] R. Mañe, "On the dimension of the compact invariant sets of certain nonlinear maps," in *Lecture Notes in Mathematics 898* (D. Rand and L. S. Young, eds.), New York, NY: Springer-Verlag, 1981.

[23] M. Henon, "A two-dimensional map with a strange attractor," *Commun. Math. Phys.*, vol. 50, pp. 69–78, 1976.

[24] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins University Press, 1989.

[25] S. Yakowitz, "Nearest-neighbour methods for time series analysis," *Journal of Time Series Analysis*, vol. 8, no. 2, 1987.

[26] L. Györfi, W. Härdle, P. Sarda, and P. Vieu, *Nonparametric Curve Estimation from Time Series*. Berlin: Springer-Verlag, 1989.

[27] Y. K. Lee and D. H. Johnson, "Nonparametric prediction of mixing, non-gaussian time series," in *Proc. Digital Signal Processing Workshop*, 1992.

[28] S. Van Huffel and J. Vanderwalle, *The Total Least Squares Problem, Computational Aspects and Analysis*. Philadelphia, PA: SIAM, 1991.

[29] A. C. Singer, G. W. Wornell, and A. V. Oppenheim, "Codebook prediction: A nonlinear signal modeling paradigm," in *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 1992.

[30] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[31] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat*, vol. 6, pp. 61–464, Mar. 1978.

[32] H. Akaike, "A new look at the statistical model identification problem," *IEEE Trans. on Auto. Control*, vol. AC-19, pp. 716–723, Dec. 1974.

[33] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Phys. Rev. Lett.*, vol. 50, pp. 346–349, 1983.

[34] P. Grassberger, "Estimating the fractal dimensions and entropies of strange attractors," in *Chaos* (A. V. Holden, ed.), Princeton University Press, 1986.

[35] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[36] A. Gelb, ed., *Applied Optimal Estimation*. Cambridge, MA: MIT Press, 1974.