A Robust Nonlinear System Identification Algorithm using Orthogonal Polynomial Network

P. Singla, T. Henderson, J. L. Junkins and J. E. Hurtado

Department of Aerospace Engineering
Texas A&M University, College Station, TX-77843-3141

15th AAS/AIAA Space Flight Mechanics Conference

Copper Mountain, Colorado January 23-27, 2005
AAS Publications Office, P.O. Box 28130, San Diego, CA 92198
A Robust Nonlinear System Identification Algorithm using Orthogonal Polynomial Network.
Puneet Singla*, Troy Henderson†, John L. Junkins‡, and John E. Hurtado§

A robust system identification algorithm is presented which makes use of linear system identification algorithms, such as Eigensystem Realization Algorithm, Observer/Kalman Identification, etc, and an orthogonal polynomial-based artificial neural network. Adaptive learning laws are derived by a thorough Lyapunov analysis to adjust different parameters of the neural network based model. The learning algorithm proposed in this paper is inspired by recent developments in adaptive control. The algorithm presented here is validated by analysis and simulation of examples based mainly on space applications. A detailed comparative study is performed to show the performance of the proposed algorithm with respect to some existing identification algorithms, specifically the Eigensystem Realization Algorithm.

1 Introduction

In the last five decades, mathematical system identification theory has evolved into a powerful scientific tool of wide applicability. However, the most mature part of the theory deals with linear systems using well established techniques of linear algebra and the theory of ordinary differential or difference equations. In contrast to this, the nonlinear system identification problem is still treated mostly on a system by system basis. One main challenge in nonlinear system identification theory is the dimensionality of the system.¹

Active control of flexible structures subject to disturbances is performed using sensors to measure the distortion of the structure and actuators to restore its original or desired shape. But to derive a control law, a model of the system dynamics between the control variables and system output is desired. Further, real time estimates of the input-output model of the system is needed. The system output in this case is generally the measured displacements

*Graduate Research Assistant, Department of Aerospace Engineering, Texas A&M University, Student Member AAS and AIAA.
†Graduate Assistant Research, Department of Aerospace Engineering, Texas A&M University, Student Member AAS.
‡Distinguished Professor, holder of George J. Eppright Chair, Department of Aerospace Engineering, Texas A&M University, Fellow AAS.
§Assistant Professor, Department of Aerospace Engineering, Texas A&M University, Member AAS, je-hurtado@tamu.edu
of the structure at various points along the surface of the structure. In this paper, our main interest is to present a general nonlinear system identification technique that can be applied for large flexible space structures. This paper is being written with four main objectives. The first and most important objective is to present a robust nonlinear system identification method using neural networks based upon novel orthogonal polynomials as activation functions. The second objective of this paper is to present adaptive learning algorithms to adjust in real time the parameters of the neural network based model. The learning algorithm proposed in this paper is inspired by recent developments in adaptive control.\(^1\,\,^2\,\,^3\) The third objective of this paper is to compare the proposed algorithm with some existing identification algorithms like the Eigensystem Realization Algorithm (ERA)\(^4\) with applications involving modeling of large flexible space structures. The fourth and final objective of this paper is to set down a theoretical framework including all assumptions, that guarantee the stability of the algorithm. Because very few results exist in nonlinear system identification theory, special care is taken to clearly state all the assumptions and develop theoretical conditions for stability.

The structure of paper is as follow: first the system identification problem is introduced followed by a brief review of some existing system identification algorithm. We give special attention to the Eigensystem Realization Algorithm (ERA) because of its numerical robustness for linear and near linear systems. Then, a robust system identification algorithm is introduced using a recently developed Global-Local Orthogonal Polynomial (GLO-MAP) network.\(^5\) Finally, the proposed algorithm is validated by simulating test cases concerned with space applications.

## 2 Problem Statement and Background

Let us consider a nonlinear system described by the following differential and algebraic equations:

\[
\begin{align*}
\dot{x}(t) &= f(x(t), u(t)) \\
y(t) &= g(x(t), u(t))
\end{align*}
\]

where \(x \in \mathcal{R}^n\) and \(u \in \mathcal{R}^p\) represent state and control vectors respectively, and \(y \in \mathcal{R}^m\) represents a vector of system outputs at time \(t\). The discrete equivalent of this system can be described by the following nonlinear difference equations:

\[
\begin{align*}
x_k &= f_d(x_{k-1}, u_{k-1}) \\
y_k &= g_d(x_k, u_k)
\end{align*}
\]

Now, if the functions, \(f(\cdot), g(\cdot), f_d(\cdot)\) and \(g_d(\cdot)\) are unknown, then the system identification problem is formally stated as follows:

**Definition of the System Identification Problem.** Identify a mathematical model which when subject to the actual input vector, \(u\), an output estimate \(\hat{y}\) is produced which approximates the actual system output, \(y\), such that

\[
\|y - \hat{y}\| \leq \epsilon
\]
Here, $\|\cdot\| : \mathcal{R}^m \rightarrow \mathcal{R}$ represents a suitable norm on the system output space, $\mathcal{Y}$, and $\epsilon$ dictates the desired accuracy of the system identification problem. In other words, the system identification problem corresponds to finding a model whose outputs are as close as desired to the true system outputs when the same input is applied to both. Therefore, the system identification problem can also be regarded as the identification of a continuous map from system input space to system output space. Consequently the problem of approximating a continuous functional arises in the system identification problem. The output at any given time is considered as a function of the input signal, which is a function of time.

Various system identification algorithms are described in the literature for input-output mapping.\textsuperscript{1,6–9,12,13} The main computational tool employed by algorithms are the Least Squares Estimation (LSE) frequently implemented using the Singular Value Decomposition (SVD), which results in numerical robustness under very weak assumptions on the persistency of excitation of the inputs. In the past few decades, Artificial Neural Networks (ANNs) have emerged as a powerful set of tools in the areas of pattern classification, time series analysis, signal processing, dynamical system modeling and control. The emergence of the ANN can be attributed to the fact that these network models are frequently able to learn behavior when traditional modeling is very difficult to generalize. Typically, a neural network consists of several computational nodes called perceptrons arranged in layers. The number of hidden nodes (perceptrons) essentially determines the degrees of freedom of the non-parametric model. However, the optimal number of hidden units, perceptrons, depends upon many factors, like the ability of the chosen basis functions to approximate the given system’s behavior, the number of data points, the signal to noise ratio, the complexity of the learning algorithms, etc. Narendra et al.\textsuperscript{6,8} have proposed different models that utilize two-layered neural networks with sigmoid functions as activation functions for system identification. In those papers, the output signal at any time is considered a function of finitely many samples of the input and output signals. The different ANN parameters are estimated using a back-propagation algorithm.\textsuperscript{10} A key issue arises because if one fixes the architecture and activation functions, a given ANN’s ability to approximate a given system’s behavior can only be deduced after the learning process is over. Adaptation of the network architecture, not simply adjusting weights, has emerged as the key to reliability and accuracy.

The major drawback of traditional ANN algorithms is that their performance decreases drastically as the dimension of the system output increases. To make this point more clear, consider a problem of active control of a flexible space structure. To derive a control law, a model of the system dynamics from the control variable, $u(t)$ to the system output, $y(t)$ is desired. Generally, the system output vector consists of surface distortion measurements at various spatial points, $O(10^3)$ which are measured by sensors like strain gauges, stereo vision systems, LIDAR, etc., placed at various points of interest. Therefore, if one seeks a dynamic continuous map between the system output and input vectors then the dimension of such a map will be as large as number of measurements, i.e., $O(10^3)$. However, the dimension of hidden states corresponding to the true system corresponds to the number of modes of interest which are generally of the order of $10 – 30$. So, a system identification algorithm is desired that can approximate the system output well, while keeping the dimension of dynamic map as low as possible. To deal with this problem, various model reduction techniques are often adopted\textsuperscript{11} for approximating high order dynamic models by simpler,
lower order models. The most popular method for model reduction is Proper Orthogonal Decomposition (POD), also known as the Principal Component Analysis (PCA). However, in model reduction, one would like to preserve properties of the original model, such as stability and physically important dynamical mode shapes. However, as POD uses a second-order statistics for model reduction, it sometimes de-emphasizes infrequent events which can be dynamically very important.

In the next section, a novel nonlinear system identification algorithm is introduced which makes use of the classical Eigensystem Realization Algorithm (ERA) and recently developed Global-Local Orthogonal Polynomial Mapping (GLO-MAP) network to deal with the issues of high dimension output vector in an efficient manner.

3 Novel System Identification Algorithm

In the previous section, issues concerning the inability of various system identification algorithms to handle dimensionality was discussed. Here, a novel system identification algorithm is presented which not only has the approximation ability of ANN but also has model reduction ability of algorithms like POD.

The basic idea of the proposed algorithm is to split the identification process into a linear and nonlinear identification processes. It should be noted that linear system identification process not only helps in designing an estimator to estimate hidden dynamic states from sensor noise corrupted measurement data, but also gives a desired order dynamical model for the hidden state vector. It implicitly defines a state space that is physically motivated to capture the best linear representation of the the system input-output behavior. Referring to Fig. 1, let the best linear model be written as

\[ \dot{x}_l = A_l x_l + B_l u \]
\[ y_l = C_l x_l + D_l u \]

Here, \( x_l \in \mathbb{R}^n \) is a hidden state vector corresponding to the best linear approximation of given input-output data. While \( A_l, B_l, C_l, D_l \) are not unique, the underlying input-output map is, and the \( A_l, B_l, C_l, D_l \) realization from ERA can be robustly computed, including the dimension \( n \) of the state space. The accuracy of the output vector, \( y_l \), in approximating the true system output data, \( y \), depends upon the nonlinearities involved. Further, an efficient estimator such as a Kalman filter can be designed to find the best estimates of the hidden dynamic state vector, \( x_h \), from given system output data, \( y \), using Eq. (7). Eq. (6) is then perturbed by a nonlinear term to learn the difference between the linear propagated state vector, \( x_l \), and the best estimate of state vector, \( x_b \).

\[ \dot{x} = A_l \dot{x} + \dot{B} u + g(\dot{x}) \]
\[ \dot{y} = C_l \dot{x} + D_l u \]

Here, \( g(\dot{x}) \) is a vector of unknown nonlinearities which can be learned by conventional ANN methods. The main steps of the algorithm proposed in this section are illustrated in Fig. 1. For the purpose of this paper, \( g(\cdot) \) represents the system nonlinearities not captured by the linear model. The convergence of the algorithm developed below is an important issue but will be considered later in this paper after discussing each step in detail.
3.1 Linear System Identification

A large class of linear system identification methods\textsuperscript{1,12,13} are listed in the literature to estimate hidden state variables along with dynamical model from given input-output data. However, the ERA and Observer/Kalman filter Identification (OKID) algorithms\textsuperscript{4,13} are the most popular and have been successfully used in various system identification problems for structural analysis. The first realization, i.e., ERA, is particularly robust and useful for structural dynamic systems where one is particularly interested in knowing particular mode frequencies and modal shapes which may be identified directly in this approach as well as reducing the order of the system to recover only those modes actually participating in the measured behavior of the system. In this paper, use of the ERA algorithm is recommended for the linear system identification module (see Fig. 1). However, any other linear system identification algorithm can be used instead of the ERA. In this section, the main steps of the ERA algorithm are briefly discussed and more details can be found in Ref. [4].

1. The first step of the ERA method is to form the Hankel matrix from the measurement outputs $Y(t_k)$, according to the following expression.

$$
H_{rs}(k-1) = \begin{bmatrix}
Y(k) & Y(k+t_1) & \cdots & Y(k+t_{s-1}) \\
Y(j_1+k) & Y(j_1+k+t_1) & \cdots & Y(j_1+k+t_{s-1}) \\
\vdots & \vdots & \ddots & \vdots \\
Y(j_r+k) & Y(j_r+k+t_1) & \cdots & Y(j_r+k+t_{s-1})
\end{bmatrix}
$$

Further, it can be easily shown that the Hankel matrix expression generalizes to the following factored expression:

$$
H_{rs}(k) = V_r A^k W_s
$$

where $V_r$ is the observability matrix given by following expression

$$
V_r = \begin{bmatrix}
C \\
CA^{j_1} \\
\vdots \\
CA^{j_r-1}
\end{bmatrix}
$$
whereas \( W_s \) is a controllability or disturbability matrix, given by the following equations depending upon the control input.

\[
\text{Impulse Response (IR): } W_s = [B \ A^t B \cdots A^{t-1} B] \\
\text{Initial State Response (ISR): } W_s = [B \ A^t X_0 \cdots A^{t-1} X_0]
\]

2. In the second step, a matrix \( H^\sharp \) is desired such that following is true:

\[
W_s H^\sharp V_r = I_n
\]

A general solution for \( H^\sharp \) is found by the Singular Value Decomposition (SVD) of \( H_{rs}(0) = PDQ^T \):

\[
H^\sharp = QD^{-1/2} \text{P}^T
\]

3. Finally, after some algebraic manipulations, the following relationship for \( Y(k+1) \) is obtained:

\[
Y(k+1) = E^T_p \text{P}^{1/2} D^{-1/2} \text{P}^T H_{rs}(1) QD^{-1/2} D^{1/2} Q^T E_m
\]

where, \( E_k^T = [I_k \ O_k \cdots \ O_k] \). Comparing this relationship with Eqs. (13) and (14), the following expressions for matrices \( A, B \) and \( C \) are obtained.

IR: \( A = D^{-1/2} P^T H_{rs}(1) QD^{-1/2} \quad B = D^{1/2} Q^T E_m \quad C = E_p^T \text{P}^{1/2} \)

ISR: \( A = D^{-1/2} P^T H_{rs}(1) QD^{-1/2} \quad X_0 = D^{1/2} Q^T E_m \quad C = E_p^T \text{P}^{1/2} \)

Now, let the estimated state matrix \( A \) be of order \( n \) and have a complete set of linearly independent eigenvectors \( (\psi_1, \psi_2, \cdots, \psi_n) \) with corresponding eigenvalues \( (\lambda_1, \lambda_2, \cdots, \lambda_n) \) which are not necessarily distinct. Define \( \Sigma \) as the diagonal matrix of eigenvalues and \( \Psi \) as the matrix whose columns are the eigenvectors. Then the minimum state-realization can be transformed to a minimum modal-realization. The diagonal matrix \( \Sigma \) contains the information about modal damping rates and damped natural frequencies, which are simply the real and imaginary parts of the eigenvalues, after transformation from discrete to continuous-time domain via \( \Lambda_c = ln(\Sigma)/\delta t \). The columns of the matrix \( \Psi^{-1} B \) define the initial modal amplitudes, or information that indicates how effective a particular input is at exciting each mode. The columns of the matrix \( C \) define the transformation from modal coordinates to the physical coordinates, i.e., system outputs.

### 3.2 State Variable Estimation

In this section, a linear estimation algorithm is described to find the best estimate of the hidden state variable, \( \hat{x}_0 \), given the output data and linear system identified in the last section. Among various estimation algorithm listed in the literature, the LSE method is the most widely used.

Given the measurement vector, \( \hat{y} \), at any time, \( t \), the algebraic relationship identified in the previous section can be used to find the best estimate of the unknown hidden state vector:

\[
\hat{y} = C_t x_t + D_t u + \nu
\]
where \( \nu \) denotes the measurement noise vector. The following expression can be derived for the state vector estimates using the least square criteria described in Ref.:

\[
\hat{x}_b = (C_l^T C_l)^{-1} (\tilde{y} - D_l u)
\]  

The main assumption in using the above expression is complete observability of the state vector \( x_l \) from Eq. (20). In other words, to estimate \( x_l \) using Eq. (20), the matrix \( C_l \) should have its rank equal to the dimension of \( x_l \) i.e. \( n \). Eq. (21) can be more robustly implemented using the SVD to compute the pseudo inverse of \( C_l \):

\[
\hat{x}_b = C_l^\dagger (\tilde{y} - D_l u)
\]  

### 3.3 Nonlinear System Identification Algorithm

In this section, a nonlinear system identification algorithm is described to update the linear dynamic model using the state estimates as measurements. This algorithm is based upon the recently developed GLO-MAP network and uses Lyapunov’s stability theorem to determine the update laws for different parameters of the GLO-MAP network.

Consider the perturbed linear dynamic model, where \( g(x) \) is a vector of nonlinear terms.

\[
\dot{x} = A_l x + Bu + g(x)
\]  
\[
y = C_l x + D_l u
\]

Here, \( A_l \in \mathbb{R}^{n \times n} \) is a known hurwitz matrix and \( B \in \mathbb{R}^{n \times p} \) is a control effectiveness matrix. It should be noted that the matrix \( A_l \) is chosen in such a way that it captures the modal frequencies of the interest and can be obtained by the ERA or any other linear system identification algorithm as described in section 3.1.

The time history estimates of the hidden state vector \( x \) can be obtained by using the procedure described in section 3.2, so the system identification problem can be re-defined as:

**System Identification Problem.** Given the time history estimates of the state vector \( x(t) \) and control variable, \( u(t) \), find estimates of the unknown nonlinearity vector \( g(\cdot) \) and control effectiveness matrix \( B \).

Further, if \( g(\cdot) \) is assumed to be a continuous function in \( x \) then according to Weierstrass approximation theorem, \( g(\cdot) \) can be approximated arbitrarily close by any set of complete functions, including a polynomial series.

\[
g(x) = C^T \Phi(x) + \epsilon
\]

where, \( \Phi(\cdot) \) is an infinite dimensional vector of polynomial functions and \( C \) is a matrix of Fourier coefficients corresponding to polynomial functions. However, according to the following theorem, the continuous function, \( g(\cdot) \) can be approximated by an orthogonal polynomial with a countable number of terms instead of infinite terms.
Theorem 1. Every nontrivial inner-product space has an orthonormal polynomial basis and further if \( \{ \phi_i \} \) is such an orthonormal basis then at most a countable number of Fourier coefficients, \( \langle g, \phi_i \rangle \) are non-zero.

Proof. Let us define a set \( S_n = \{ i \in I : | \langle g, \phi_i \rangle | > 1/n \} \). Here, \( I \) denotes an uncountable index set and should not be confused with set of integers. Note, to prove this theorem, one just need to show that \( S_n \) is a finite set. Now, if \( f = \sum_{j \in S_n} \langle g, \phi_j \rangle \phi_j \) is the orthogonal projection of \( g \) onto the subspace, \( \mathcal{U} = \text{span}[\phi_j : j \in S_n] \) then by Pythagorean Law:

\[
\|g\|^2 = \|(g - f) + f\|^2 = \|g - f\|^2 + \|f\|^2 \geq \sum_{j \in S_n} \| \langle g, \phi_j \rangle \phi_j \|^2.
\]

As \( \phi_i \) is an element of the orthonormal basis, i.e., \( \|\phi_i\| = 1 \), the above expression reduces to

\[
\|g\|^2 \geq \sum_{j \in S_n} | \langle g, \phi_j \rangle |^2 \geq \sum_{j \in S_n} 1/n^2 = \text{card}(S_n)/n^2.
\]

Now, as \( \|g\| < \infty \) hence \( \text{card}(S_n) < \infty \), i.e., \( S_n \) is a finite set.

According to this theorem, \( \Phi(.) \) can be chosen as a finite dimensional vector of orthogonal polynomials. Therefore, \( C \in \mathcal{R}^{n \times n} \) is a matrix of Fourier coefficients corresponding to these orthogonal polynomial functions. Now, substituting Eq. (25) in Eq. (23) yields:

\[
\dot{x}(t) = A_I x(t) + Bu(t) + C^T \Phi(x) + \epsilon
\]  

But the Fourier coefficient matrix \( C \) is unknown so we write an estimate equation

\[
\dot{x}(t) = A_I \hat{x}(t) + \hat{B}u(t) + \hat{C}^T \hat{\Phi}(x)
\]

Let us define \( e(t) = x(t) - \hat{x}(t) \), which leads to the following expression:

\[
\dot{e}(t) = A_I e + (B - \hat{B})u(t) + (C - \hat{C})^T \Phi(x) + \epsilon
\]

Now, to find adaptation laws for the unknown parameters, we consider the following Lyapunov function:

\[
V = \frac{1}{2} e^T Pe + \frac{1}{2} Tr(\hat{B} \Gamma_1 \hat{B}^T) + \frac{1}{2} Tr(\hat{C}^T \Gamma_2 \hat{C})
\]

where, \( P \) is a positive definite symmetric matrix. Now taking the time derivative of \( V \) leads to the following equation:

\[
\dot{V} = \frac{1}{2} e^T (PA_I + A_I^T P) e + e^T P \left( \hat{B}u(t) + \hat{C}^T \Phi(x) + \epsilon \right) + Tr \left( \hat{B} \Gamma_1 \hat{B}^T \right)
\]

\[
+ Tr \left( \hat{C}^T \Gamma_2 \hat{C} \right)
\]

\[
= -\frac{1}{2} e^T Q e + Tr \left( B \left[ \Gamma_1 \hat{B}^T + u \epsilon^T P \right] \right) + Tr \left( C^T \left[ \Gamma_2 \hat{C} + \Phi(x) \epsilon^T P \right] \right)
\]

\[
+ e^T P \epsilon
\]
Note, here $Q \in \mathbb{R}^{n \times n}$ is a positive definite matrix which satisfies the following algebraic Ricatti equation
\begin{equation}
PA_t + A_t^T \mathbf{P} = -Q \tag{33}
\end{equation}

Now, if following adaptation laws are chosen for $\hat{B}$ and $\hat{C}$,
\begin{align*}
\dot{\hat{B}}^T &= -\hat{B}^T = -\Gamma_1^{-1} \mathbf{u} \mathbf{e}^T \mathbf{P} \\
\dot{\hat{C}} &= -\hat{C} = -\Gamma_2^{-1} \Phi(x) \mathbf{e}^T \mathbf{P} 
\end{align*}
\begin{equation}
\tag{34}
\text{and (35)}
\end{equation}

then $\dot{V}$ reduces to:
\begin{align*}
\dot{V} &= -\frac{1}{2} \mathbf{e}^T \mathbf{Q} \mathbf{e} + \epsilon^T \mathbf{P} \mathbf{e} \\
\Rightarrow \dot{V} &\leq -\frac{1}{2} \lambda_{\text{min}}(Q) \| \mathbf{e} \|^2 + \| \epsilon \| \| \mathbf{P} \| \| \mathbf{e} \| \tag{37}
\end{align*}

Therefore, $\dot{V}$ is negative semi-definite if $\| \mathbf{e} \| > 2 \frac{\| \epsilon \| \| \mathbf{P} \|}{\lambda_{\text{min}}(Q)} = \epsilon_{lb}$. This means that if system tracking residual is less than $\epsilon_{lb}$ then different learning parameters can no longer be guaranteed convergent in this boundary layer region. However, convergence difficulties can be avoided by insisting upon small the identification error and restricting the magnitude of various parameters. In case of $\dot{V} < 0$, the convergence of tracking residual $\mathbf{e}$ follows from the assumption that $\mathbf{e} \in L_\infty$, i.e., both $\mathbf{x}$ and $\hat{\mathbf{x}}$ are bounded signals. Further, from the integral of $\dot{V}$, it can be easily shown that $\mathbf{e} \in L_2 \cap L_\infty$ and therefore, from Barbalat’s Lemma\(^2\) $\mathbf{e} \to 0$ as $t \to \infty$, which in turn leads to $\hat{\mathbf{B}} \to 0$ and $\hat{\mathbf{C}} \to 0$ based on Eqs. (34) and (35).

It should be noticed that according to Weirstrass’s approximation theorem as $N \to \infty$, that the approximation error $\epsilon \to 0$. However, in practice this is not possible as these adaptation laws are based upon the assumption that all the parameters of the network can be optimized simultaneously. However, the global nature of the continuous map, $g(.)$, can lead to globally-optimal network parameters which adequately minimize the approximation error on one set of input data but are not robust when tested on some new input data set. An alternative to global learning is local learning using local weight functions. The local learning algorithms involve estimation of network parameters using the observations in the local neighborhood of the operating point. Generally, the sizing of the local neighborhood is dictated by the support or domain of the weight functions. In reference,\(^5\) an approximation method is presented that enables a piecewise continuous approximation in a $n$-dimensional space using orthogonal polynomials and specially designed weight functions for overlapping the approximations in contiguous overlapping local regions to obtain the desired order of global continuity. In the next section, those results will be extended for the dynamical system identification case so that the approximation error $\epsilon$ can be significantly reduced. The adaptive nature of this approximation approach can essentially guarantee a small $\epsilon$, if low noise measurement density in space and time is available.

### 3.4 Adaption Law Derivation using GLO-MAP Network

Before developing the adaption laws for different parameters of the GLO-MAP network, some features of the GLO-MAP network are summarized. The main idea of GLO-MAP
network is a weighting function technique that generates a global family of overlapping preliminary approximations whose centroids of validity lie on the vertices of an \( n \)-dimensional grid, with vertices separated by a uniform step \( h \). These preliminary approximations are constructed so they represent the behavior in local hypercubes with a volume \( (2h)^n \) centered on a typical vertex in the grid. A novel averaging process is developed in Ref. 17 to determine a piecewise continuous global family of local least squares approximations, while having the freedom to vary the nature (e.g., degrees of freedom) of the local approximations. The continuity conditions are enforced by using a unique set of weighting functions in the averaging process. The weight functions are designed to guarantee the global continuity conditions while retaining near complete freedom on the selection of the generating local approximations. However, if the preliminary local approximations are chosen as linear combinations of a set of basis functions constructed such that they are orthogonal with respect to the weight functions, then many advantages are realized, as demonstrated in the Ref. 5.

In Fig. 2, several qualitative observations regarding the weighting function approach are illustrated. One critical attractive property of the weight functions is that they add to unity everywhere in the overlapping unit region, i.e., they form a partition of unity. Notice further that the weight functions have a qualitative bell shape, but fairing into a square base, the zero contour being the boundary opposite (e.g., 2-3-4) to the vertex (e.g., point 1) where the weight has a unit value. Furthermore, notice that along any boundary, only the two weight functions associated with the two approximations centered at the end points of that boundary are non-zero along that boundary, while the other two weight functions are zero (the partial derivatives of the other two weight functions are also zero along this boundaries). These continuity arguments on the averaged approximation of the function can be extended readily to corresponding properties on their partial derivatives: The averaged approximation osculate in value and partial derivatives with the four preliminary approximations at their corresponding vertices, and the function and both partial derivatives along any boundary are a weighted average of the corresponding two functions associated with the end point of that boundary and their partial derivatives are likewise an average of the partial derivatives of the functions at the end point of that boundary. Collectively, these observations lead to rigorous piecewise continuity of the averaged approximations, while leaving the user free to choose any preliminary local approximations desired or needed. In Ref. 5 these qualitative observations are developed systematically and extended rigorously to approximation with arbitrary order continuity in an \( n \) dimensional space. In general, the final approximation in any hypercube is obtained by averaging \( 2^n \) overlapping approximations centered at the vertices of that hypercube.

To illustrate this approach let us first assume \( n = 2 \) and \( g(x_1, x_2) : \mathcal{R}^2 \rightarrow \mathcal{R}^2 \) is a continuous function which can be approximated by GLO-MAP according to following Eq.:

\[
g(x_1, x_2) = w_{0,0}(x_1^{I_1}, x_2^{I_2}) \mathbf{g}_{I_1, I_2}(x_1, x_2) + w_{0,1}(x_1^{I_1}, x_2^{I_2+1}) \mathbf{g}_{I_1, I_2+1}(x_1, x_2) + \cdots \\
w_{1,0}(x_1^{I_1+1}, x_2^{I_2}) \mathbf{g}_{I_1+1, I_2}(x_1, x_2) + w_{1,1}(x_1^{I_1+1}, x_2^{I_2+1}) \mathbf{g}_{I_1+1, I_2+1}(x_1, x_2) \tag{38}
\]
Preliminary Approximations:

- $F_{11}(x, y)$

  \[ w_{11}(x, y) = x^2(3 - 2x)y^2(3 - 2y) \]

  \[ w_{01}(x, y) = w_{11}(1 - x, y) \]

- $F_{01}(x, y)$

  \[ w_{00}(x, y) = w_{11}(1 - x, 1 - y) \]

- $F_{10}(x, y)$

  \[ w_{10}(x, y) = w_{11}(x, 1 - y) \]

Final Approximation:

\[ \bar{F}(x, y) = \sum_{i=0}^{1} \sum_{j=0}^{1} w_{ij}(x, y) F_{ij}(x, y) \]

valid over \(0 \leq x \leq 1, 0 \leq y \leq 1\)

Weight functions are a partition of unity:

\[ \sum_{i=0}^{1} \sum_{j=0}^{1} w_{ij}(x, y) = 1 \]

Figure 2 Qualitative Representation of the Averaging Process in Two Dimensions
function: Now to find adaptation laws for unknown parameters, let us consider following Lyapunov

Once again, the Fourier coefficient matrix

Now, using the approximation for

So, Eq. (39) reduces to:

Further, the local approximations, \( f_{I_1, I_2}(x_1, x_2) \), can be approximated by a set of orthogonal basis functions, \( \Phi \) as follows:

Now, making use of Eq. (41) the matrix \( F \) in Eq. (39) can be rewritten as:

So, Eq. (39) reduces to:

Now, using the approximation for \( g(\cdot) \) given by Eq. (43), Eq. (23) reduces to:

Once again, the Fourier coefficient matrix \( C \) and control effectiveness matrix \( B \) are unknown and one can write:

Let us define \( e(t) = x(t) - \hat{x}(t) \) and time derivative of \( e(t) \) can be written as:

Now to find adaptation laws for unknown parameters, let us consider following Lyapunov function:

\[
V = \frac{1}{2} e^T P e + \frac{1}{2} Tr(\hat{B}^T \hat{B}) + \frac{1}{2} Tr(\hat{C}^T \hat{C})
\]
where, \( P \) is a positive definite symmetric matrix. Now taking time derivative of \( V \) leads to following Eq.:

\[
\dot{V} = \frac{1}{2} e^T (PA_t + A_t^T P) e + e^T P \left( \dot{B}u(t) + \dot{C}\Psi(\cdot) + \epsilon \right) + Tr \left( \dot{B}\Gamma_1 \dot{B}^T \right) \\
+ Tr \left( \dot{C}\Gamma_2 \dot{C}^T \right) \\
= -\frac{1}{2} e^T Q e + Tr \left( \dot{B} \left[ \Gamma_1 \dot{B}^T + u e^T P \right] \right) + Tr \left( \dot{C} \left[ \Gamma_2 \dot{C}^T + \Psi(\cdot) e^T P \right] \right) + e^T P \epsilon
\]

(49)

Therefore, if following adaptation laws are chosen for \( \dot{B} \) and \( \dot{C} \),

\[
\dot{B}^T = -\Gamma_1^{-1} u e^T P \\
\dot{C}^T = -\Gamma_2^{-1} \Psi(x) e^T P
\]

(50)

(51)

then \( \dot{V} \) reduces to:

\[
\dot{V} = -\frac{1}{2} e^T Q e + e^T P e
\]

(52)

\[\Rightarrow \dot{V} \leq -\frac{1}{2} \lambda_{\min}(Q) \|e\|^2 + \|\epsilon\| \|P\| \|e\|
\]

(53)

Therefore, \( \dot{V} \) is negative definite if \( \|e\| > 2 \frac{\|\epsilon\| \|P\|}{\lambda_{\min}(Q)} \). The asymptotic stability of the tracking residual \( e \) follows from the same arguments as outlined in the last section.

The generalization of Eq. (39) is:

\[
g(X_1, \ldots, X_N) = \sum_{i_1=0}^1 \sum_{i_2=0}^1 \cdots \sum_{i_N=0}^1 \left( w_{i_1, \ldots, i_N} (I_1 + i_1 X_1, \ldots, I_N + i_N X_N) \right)
\]

\[
g_{I_1 + i_1, \ldots, I_N + i_N} (X_1, \ldots, X_N)
\]

(54)

However, the expression for the adaptation laws for the Fourier coefficients and control effectiveness matrix remains same except that now matrix \( C \) in Eq. (51) consists of the coefficients of \( 2^n \) neighboring approximations depending upon the value of \( x \).

Finally, we mention that the state vector, \( x \) is generally unknown so the best available estimates of state vector, \( \hat{x} \) are used as state measurements. Note these estimates can be obtained from linear system identified by ERA algorithm and using the least square algorithm as discussed in previous sections.

4 Convergence of Identification Algorithm

The convergence of the preceding nonlinear system identification algorithm can be proved under reasonable set of assumptions, which are captured in following theorem.

**Theorem 2.** If the linear system described by Eqs. (6) and (7) is fully observable and tuning parameters \( P \) and \( Q \) are chosen in such a way that \( \dot{V} \) described by Eq. (52) is negative definite then \( \|\hat{y} - y\| \geq \|C (\epsilon_1 + \epsilon_{ib})\| \). Where, \( \epsilon_1 \) represents the hidden state estimation accuracy and \( \epsilon_{ib} = 2 \frac{\|\epsilon\| \|P\|}{\lambda_{\min}(Q)} \).
Proof. The observability of the identified linear dynamic system guarantees that the hidden state $x$ can be detected from the given output $y$. In other words, the least square estimate $\hat{x}_b$ (given by Eq. (21)) can be obtained in such a way that

$$\|x - \hat{x}_b\| = e_1$$  \hspace{1cm} (55)

Now, we just need to show that $\hat{x}$ asymptotically converges to $\hat{x}_b$ and this follows from the negative definiteness of $V$. However, we have shown that $V$ is negative definite if $\|e\| > 2\|x_0\|\|P\| = e_{lb}$. This means that state identification error is always lower bound by $e_{lb}$ which implies

$$\|y - \hat{y}\| = \|C_1(x - \hat{x})\|$$
$$\quad = \|C_1(x - \hat{x}_b + \hat{x}_b - \hat{x})\|$$
$$\quad \geq \|C_1(e_1 + e_{lb})\|$$

The above theorem gives us the lower bound for the system identification error which can be reduced to a desired tolerance by judicious choice of various tuning parameters.

5 Numerical Simulation

The proposed nonlinear system identification algorithm is tested on a variety of test cases mainly concerned with large space structures. In this section, some results from these studies will be presented.

5.1 Dynamic System Identification of Large Space Antenna

Space Based Radar (SBR) systems envisioned for the future may be a constellation of spacecraft that provide persistent real-time radar images of the Earth environment through the identification and tracking of moving targets, high-resolution synthetic aperture radar imaging, and collection of high-resolution terrain information. The accuracy of the information obtained from the SBR system depend upon many parameters like the geometric shape of the antenna, permittivities of the media through which radar wave is traveling, etc. and our ability to compensate in real-time implicitly depends on the accuracy of system identification. Therefore the characteristics of the scattered wave received by the SBR antenna for a given frequency depend on the surface and geometric parameters of the radar. To apply necessary corrections for scattering of radar waves, the precise knowledge of the SBR antenna becomes a necessity. However, the transient excitation of the flexible dynamics mode necessitated by the need to slew the antenna makes the shape estimation problem a bit difficult. While a variety of surface models can be employed to model the instantaneous shape, we consider the case that the surface is measured at discrete points and a dynamical model for shape estimation is desired. The objective of this section is to apply the system identification methodology, developed in this paper, to estimate the real time SBR antenna shape using only the discrete time measurements of the antenna surface.
For simulation purposes the SBR antenna geometry is modeled in NASTRAN. The antenna model consists of total 7 panels as shown in Fig. 3. Each panel is assumed to be 100m long in length and 200 × 250m² in area. It is assumed that shape deflections measurements are available at uniformly distributed 25 points per panel so giving rise to total 175 shape measurements at each time instant. NASTRAN is used to generate mass, \( M \), and stiffness, \( K \), matrices for the antenna structure and coordinate transformation matrix, \( T \), to transform the modal coordinates to physical coordinates i.e. deflections along each axis.

\[
\text{Modal Equations: } M \ddot{\eta} + K \eta = 0 \tag{56}
\]

\[
\text{Transformation to Phy. Coord.: } \mathbf{y} = T \eta \tag{57}
\]

where, \( \eta \) and \( \mathbf{y} \) represent modal and physical coordinates respectively. The order of the FEM model was 175 × 3 = 525. However, order reduction order reduction methods are used to reduce the dimension of the model state space to 10-30. Later, these equations augmented with artificial damping and nonlinearities, are simulated in MATLAB environment to generate the measurement data for 50 seconds at 10Hz frequency. It should be mentioned that only first 10 modes are excited to generate measurement data. For the purpose of this paper, radial basis functions are used to simulate artificial nonlinearity with random magnitude and center.

Now, according to the procedure listed in section 3, first, the ERA algorithm is used to generate linear dynamic model for the SBR antenna model followed by estimation algorithm to estimate modal coordinates. As expected, the ERA system gives us 20th order linear dynamic model consistent with 10 modes. Finally, the nonlinear system identification algorithm is used to refine the linear model.

![NASTRAN SBR Antenna Model Consists of 7 panels](image1)

![Close-Up of One Panel](image2)

**Figure 3 NASTRAN Model of SBR Antenna**

Fig. 4(i) shows the true and the ERA model simulated measurements for various points on the antenna surface and Fig. 4(ii) shows the corresponding output error plots. Further, Fig. 4(iii) shows the plots of various true and the ERA model simulated hidden states and
Fig. 4(iv) shows the plot for estimated hidden state error. From these plots, it is clear that although ERA is able to capture all 10 modes of interest, there is significant error in estimating the true nonlinear output and even the modal coordinates.

To model the nonlinear correction of the dynamic model, the procedure listed in section 3.4 is used. Only 2 elements are used to grid the estimated modal data according to Eq. (54). The orthogonal polynomial functions used to model the nonlinear function, \( g(.) \) are listed in Table 1.

\[
g_{t_1}(r) = C^T \Phi(r)
\]

where, \( r = \|x\|_2 \) and \( \Phi \in \mathbb{R}^4 \) so that \( C \) is a 20 \( \times \) 4 matrix of Fourier coefficients. Initially, \( C \) is assumed to be a zero matrix and is adapted by using Eq. (51). Fig. 5 shows the adaptation plot of various Fourier coefficients corresponding to one such function. Fig. 6(i)
<table>
<thead>
<tr>
<th>degree</th>
<th>Basis Functions, $\phi_j(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>$x$</td>
</tr>
<tr>
<td>2</td>
<td>$(-2 + 15x^2)/13$</td>
</tr>
<tr>
<td>3</td>
<td>$(−9x + 28x^3)/19$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$n$</td>
<td>$\phi_n(x) = \frac{1}{c_n} x^n - \sum_{j=0}^{n-1} \frac{\langle x^n, \phi_j(x) \rangle}{\langle \phi_j(x), \phi_j(x) \rangle} \phi_j(x)$</td>
</tr>
</tbody>
</table>

Table 1 One Dimensional Basis Functions Orthogonal with respect to the weight Function $w(x) = 1 - x^2(3 - 2|x|)$

shows the true and identified nonlinear model simulated measurements for various points on the antenna surface and Fig. 6(ii) shows the corresponding output error plots. Further, Fig. 6(iii) shows the plots of various true and identified nonlinear model simulated hidden states and Fig. 6(iv) shows the plot for estimated hidden state error. From these plots, it is clear that the use of nonlinear system identification algorithm reduces the estimation error by at least an order of magnitude. However, we should mention that approximation accuracy will depend upon the order of polynomials used as well as the excited flexible dynamics mode.

6 Concluding Remarks

A general methodology for non-linear system identification is presented in this paper. The method splits the nonlinear system identification process into two parts: 1) Linear system identification using ERA 2) Nonlinear system identification using GLO-MAP. GLO-MAP determines an approximation of the nonlinear correction term in Eq. (8) A particularly attractive choice to model the nonlinear term is shown to be polynomial basis functions that are orthogonal with respect to the weight functions of the averaging process of the GLO-MAP algorithm. The adaption laws for different parameters of GLO-MAP network are derived by using Lyapunov’s analysis. The proposed algorithm, being the combination of ERA and GLO-MAP invokes the approximation ability of ANN and simultaneously has the model reduction ability of algorithms like POD. The use of ERA to reduce the order of the system and determine physically meaningful state space realization is an attractive feature. The convergence of the algorithm is supported by a thorough analysis and demonstrated in the numerical study. The broad generality of the method, together with simulation results provides a strong basis for optimism for the importance of these ideas. However, more testing would be required to reach stronger conclusions about the utility of this algorithm, particularly in the presence of sensor noise.
7 Acknowledgement

The author(s) wish to acknowledge Ms. Magdalini Lagoudas and Andy Mcevoy for their help in creating the NASTRAN model for the SBR antenna.

This research is supported by Lockheed Martin and Harris under the Defense Advanced Projects Agency's Innovative Space-Based Radar Antenna Technology (ISAT) program; the collaboration of Mr. Thomas Depkovich of Lockheed Martin is warmly appreciated.

The research in part is also supported by the Texas Institute for Intelligent Bio-Nano Materials and Structures for Aerospace Vehicles, funded by NASA Cooperative Agreement No. NCC-1-02038; the technical liaison of Dr. Tom Gates is warmly acknowledged.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies, nor do they represent a physical realization of any specific ISAT hardware design.

REFERENCES


Figure 6 Non-Linear System Identification Results After Completing All Steps


