FREQUENCY DOMAIN STATE-SPACE SYSTEM IDENTIFICATION

Chung-Wen Chen, Jer-Nan Juang and Gordon Lee

July 1992

NASA
National Aeronautics and Space Administration
Langley Research Center
Hampton, Virginia 23665-5225
Frequency Domain State-Space System Identification

Chung-Wen Chen*
North Carolina State University, Raleigh, NC 27695-7910

Jer-Nan Juang**
NASA Langley Research Center, Hampton, VA 23665

Gordon Lee+
North Carolina State University, Raleigh, NC 27695-7921

Abstract

This paper presents an algorithm for identifying state-space models from frequency response data of linear systems. A matrix-fraction description of the transfer function is employed to curve-fit the frequency response data, using the least-squares method. The parameters of the matrix-fraction representation are then used to construct the Markov parameters of the system. Finally, state-space models are obtained through the Eigensystem Realization Algorithm using the Markov parameters. The main advantage of this approach is that the curve-fitting and the Markov-parameter-construction are linear problems which avoid the difficulties of non-linear optimization of other approaches. Another advantage is that it avoids windowing distortions associated with other frequency domain methods.

Introduction

State-space models of dynamic systems are usually required for many current control design methods as these control approaches are developed based upon some state-space representation of the system. Recently, it has been found that state-space models can be effectively identified through the Observer/Kalman Filter System Identification method (OKID)[1-4] using time domain input-output data. However, there are cases in

* Research Associate, Mars Mission Research Center, Member AIAA.
** Principal Scientist, Spacecraft Dynamics Branch, Fellow AIAA.
+ Professor, Mars Mission Research Center, Member AIAA.
which frequency response data rather than time histories are available. This is often the case with the advent of sophisticated spectrum analyzers and associated automatic test equipment. Therefore, the technique of obtaining state-space models from frequency response data is of practical interest.

Classically, the Inverse Discrete Fourier Transform method (IDFT) is used to transform frequency response data to time domain data, that is, to transform the frequency response function (FRF) of the system to its pulse response. The pulse response of discrete-time systems is also known as the Markov parameters. The disadvantage of this approach is that the Markov parameter sequence thus obtained is distorted by time-aliasing effects.

Recently, a method called the State-Space Frequency Domain (SSFD) identification algorithm has been developed. This method can estimate Markov parameters from the FRF without windowing distortion and an arbitrary frequency weighting can be introduced to shape the estimation error. The method uses a rational matrix description (the ratio of a matrix polynomial and a monic scalar polynomial denominator) to curve-fit the frequency data and obtains the Markov parameters from this equation. In obtaining the state-space models from the Markov parameters, the Eigensystem Realization Algorithm (ERA) or its variant ERA/DC is used. The disadvantage of this method is that the curve-fitting problem must either be solved by non-linear optimization techniques or by linear approximate algorithms requiring several iterations.

This paper proposes a simple yet effective way of curve-fitting the FRF data and of constructing the Markov parameters. Instead of using a rational matrix function, this method uses a matrix-fraction for the curve-fitting. Thus the curve-fitting is reformulated as a linear problem which can be solved by the ordinary least-squares method in one step; that is, no iteration is required. The method can match the frequency response data perfectly if the FRF is accurate in ideal cases, and will seek an optimal match if noise
and/or distortion are involved in data. This new approach retains all the advantages associated with the SSFD while avoids the iterative, approximate curve-fitting procedures.

Section 2 gives some background and the notation used for this problem. Section 3 discusses the curve-fitting method while Section 4 describes a method to compute the Markov parameters from the parameters obtained from curve-fitting. The process of going from the Markov parameters to a state-space model is discussed in Section 5. Finally, simulated data from a model of the Mini-Mast structure and experimental data from a NASA testbed are used in Section 6 as illustrative examples. The simulated data discuss an ideal FRF case (without distortion and noise) whereas the experimental data present a practical case. The illustrative examples show that the method is effective in both cases.

**Background and Notation**

The objective of frequency domain state-space system identification is to identify state-space models from the given frequency response data — the frequency response functions (FRF). The state-space representation of a linear discrete-time system is

\[
\begin{align*}
    x(k+1) &= Ax(k) + Bu(k) \\
    y(k) &= Cx(k) + Du(k)
\end{align*}
\]

where \( x(k) \in R^{nx1} \) is the state vector, \( u(k) \in R^{rx1} \) the input vector, \( y(k) \in R^{mx1} \) the output vector; \( A, B, C, D \) are the system matrix, input matrix, output matrix and direct-influence matrix, respectively. Matrices \( A, B, C, D \) are referred to as state-space parameters or the state-space model. The relation between the state-space parameters and the FRF \( G(\omega_i) \) is

\[
G(\omega_i) = C(e^{i\omega_i T}I_n - A)^{-1}B + D
\]

where \( T \) is the sampling time of the discrete-time system in seconds and \( \omega_i \) are the frequencies in rad/sec. Given \( G(\omega_i) \), the problem of frequency state-space system
identification is to find a set of state-space parameters, denoted by \([\hat{A}, \hat{B}, \hat{C}, \hat{D}]\) (hereafter "\(^\wedge\)" denotes an estimated value), such that the estimated FRF
\[
\hat{G}(\omega) = \hat{C}(e^{j\omega T} \times I_n - \hat{\hat{A}})^{-1} \hat{\hat{B}} + \hat{\hat{D}}
\]
matches \(G(\omega)\) optimally under some optimality criterion. Note that \(G(\omega)\) is a matrix of a dimension \(p \times m\). If the \(L_2\)-norm of the error is to be minimized, then an appropriate error criterion is
\[
\min_{\hat{A}, \hat{B}, \hat{C}, \hat{D}} \sum_{i=1}^{\ell} w^2(\omega_i) \| G(\omega_i) - \hat{G}(\omega_i) \|_2
\]
where \(w(\omega_i)\) is a specified weighting function of frequency and \(\ell\) is the total number of the frequency data.

Optimizing Eq. (5) with respect to the state-space parameters directly is a nonlinear problem, which may be difficult to solve. To avoid the difficulties associated with the nonlinear optimization, one possible alternative is to optimize first with respect to the Markov parameters and then convert the Markov parameters to a state-space model, as optimizing Eq. (5) with respect to the Markov parameters is a linear problem. To formulate this alternative mathematically, it begins with expanding Eq. (3)
\[
G(\omega) = D + CBe^{-j\omega T} + CABe^{-j2\omega T} + CA^2Be^{-j3\omega T} + \cdots = \sum_{k=0}^{\infty} Y_k e^{-jkaT}
\]
where \(Y_k = CA^{-k}B\)

Substitution of Eq. (6) into Eq. (5) yields
\[
\min_{(a,b,c,d)} \sum_{i=1}^{\ell} w^2(\omega_i) \left\| G(\omega_i) - \sum_{k=0}^{\infty} \hat{Y}_k e^{-jkaT} \right\|_2
\]
The parameters \(Y_k = CA^{-k}B\) \((k = 1, \cdots, \infty)\) are the Markov parameters. However, the problem associated with this approach is that theoretically the number of Markov parameters is infinite. Though overall the Markov sequence is a decreasing sequence, assuming the system is stable, it may take a large number of terms to make \(CA^{k-i}B = 0\) for all \(k \geq k_i\), for some arbitrarily large \(k_i\), especially when the system is lightly damped. A large number of Markov parameters will make the optimization in Eq. (7) computationally too intensive and impractical for many applications.
To avoid the problem of excessive number of parameters in the optimization, an intermediate step should be taken. That is, curve-fit the FRF data using a finite-ordered matrix-fraction first and then construct the Markov parameters from this result. This approach is detailed in the next section.

Linear Curve-fitting

The transfer function matrix of the system described by Eqs. (1) and (2) is

\[ G(z^{-1}) = C(zI_n - A)^{-1}B + D = \frac{\text{Adj}(zI_n - A)B + p(z)D}{p(z)} \]  

where \( p(z) \) is the characteristic polynomial of \( A \), and \( \text{adj}(\cdot) \) denotes the adjoint of a square matrix. Polynomial \( p(z) \), in general, is monic. The FRF is simply the transfer function matrix \( G(z^{-1}) \) calculated along the unit circle in the \( z \)-plane. Note that \( G(z^{-1}) \) does not change if the numerator and denominator of Eq. (8) are multiplied simultaneously by an arbitrary monic scalar polynomial. This implies that the expression of \( G(z^{-1}) \) as a rational matrix polynomial is not unique; therefore, one can over-specify the orders of the polynomials. In Ref. [6] the SSFD uses this rational matrix polynomial to curve-fit frequency data. By doing so, however, the estimation of the parameters becomes a non-linear problem. The SSFD therefore has to use an approximate, iterative method to solve the problem, which results in a value not generally optimal in any sense in the presence of noise and/or incorrect model order.

On the other hand, it is also known that the transfer function matrix can be expressed by a left matrix-fraction\(^{[3,4]}\) description as

\[ G(z^{-1}) = A^{-1}(z^{-1})B(z^{-1}) \]  

where both \( A(z^{-1}) \) and \( B(z^{-1}) \) are matrix polynomials:

\[
A(z^{-1}) = I_n + A_1z^{-1} + \cdots + A_pz^{-p} \quad (A_i \in R^{n \times n}) \tag{10}
\]
\[
B(z^{-1}) = B_0 + B_1z^{-1} + \cdots + B_pz^{-p} \quad (B_i \in R^{m \times r}) \tag{11}
\]
This factorization is also not unique. For convenience one can choose the orders of both polynomials to be equal (=\(p\)). Pre-multiplying Eq. (9) by \(A(z^{-1})\) one has
\[
A(z^{-1})G(z^{-1}) = B(z^{-1})
\]
which can be rearranged to become
\[
G(z^{-1}) = -A_1G(z^{-1})z^{-1} - \cdots - A_pG(z^{-1})z^{-p} + B_0 + B_1z^{-1} + \cdots + B_pz^{-p}.
\]
Note that with \(G(z^{-1})\) and \(z^{-1}\) known, Eq. (13) is a linear equation. Because \(G(z^{-1})\) is known at \(z = e^{i\omega_iT}\) \((i = 1, \cdots, \ell)\), there are a total of \(\ell\) equations available. Denoting \(e^{i\omega_iT}\) by \(z_i\) and stacking up the \(\ell\) equations, one has
\[
\begin{bmatrix}
G^T(z_1^{-1})z_1^{-1} & \cdots & G^T(z_1^{-1})z_1^{-p} & I_r & \cdots & z_1^{-1}I_r & \cdots & z_1^{-p}I_r \\
G^T(z_2^{-1})z_2^{-1} & \cdots & G^T(z_2^{-1})z_2^{-p} & I_r & \cdots & z_2^{-1}I_r & \cdots & z_2^{-p}I_r \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
G^T(z_{\ell}^{-1})z_{\ell}^{-1} & \cdots & G^T(z_{\ell}^{-1})z_{\ell}^{-p} & I_r & \cdots & z_{\ell}^{-1}I_r & \cdots & z_{\ell}^{-p}I_r
\end{bmatrix}
\begin{bmatrix}
-A_1^T \\
\vdots \\
-A_p^T \\
B_0 \\
\vdots \\
B_p^T
\end{bmatrix}
= \begin{bmatrix}
G^T(z_1^{-1}) \\
G^T(z_2^{-1}) \\
\vdots \\
G^T(z_{\ell}^{-1})
\end{bmatrix},
\]
or in short,
\[
\Theta \Theta = \Psi
\]
where
\[
\Theta^T = \begin{bmatrix}
-A_1 & \cdots & -A_p & B_0 & \cdots & B_p
\end{bmatrix}
\]
\[
\Psi^T = \begin{bmatrix}
G(z_1^{-1}) & G(z_2^{-1}) & \cdots & G(z_{\ell}^{-1})
\end{bmatrix}
\]
and \(\Phi\) is the large data matrix. Equation (15) is a normal equation, where a least-squares solution of \(\Theta\) can be found.

The least-squares solution of \(\Theta\) could be complex numbers. To avoid this, one can force \(\Theta\) to be real by solving either the real part or the imaginary part of Eq. (15):
\[
real(\Phi)\Theta = real(\Psi)
\]
\[
imag(\Phi)\Theta = imag(\Psi).
\]
Or one can combine and solve both equations as:
\[
\begin{bmatrix}
real(\Phi) \\
imag(\Phi)
\end{bmatrix}\Theta = \begin{bmatrix}
real(\Psi) \\
imag(\Psi)
\end{bmatrix}.
\]

Remark 1: Using a matrix polynomial as the left divisor in Eq. (9) instead of using a scalar polynomial denominator as in Eq. (8) has a remarkable advantage of making the
parameter estimation become a linear problem. With a scalar polynomial denominator, Eq. (8) can also be formulated to have a form similar to Eq. (13); however, in this case some parameters have to be diagonal matrices with identical diagonal elements. With this constraint, the ordinary least-squares method can no longer be applied.

**Remark 2:** In an ideal case where the FRF is accurate, if the order of $A(z^{-1})$ and $B(z^{-1})$ is over-specified, the rank of $\text{real}(\Phi)$ will be less than the row number of $\Theta$. In other words, the number of unknowns is more than the number of equations; therefore, the answer is not unique. However, a minimum norm solution still can be found using the least-squares method\[9\]. In this case, the match is exact; i.e., $\text{real}(\Phi)\Theta$ is exactly equal to $\text{real}(\Psi)$. The cases of using the imaginary part or of using combined real and imaginary parts are similar.

**Remark 3:** In a sense, $A(z^{-1})$ and $B(z^{-1})$ in Eq. (9) can be interpreted as the observer Markov parameters\[2,3,4\]. The order $p$ of $A(z^{-1})$ and $B(z^{-1})$ can be arbitrary as long as it is set equal to or greater than $n/m$,\[4\] where $n$ is the system dimension and $m$ is the number of outputs. Therefore, if more outputs are available, a smaller order can be assigned to the matrix polynomials.

**Remark 4:** The frequency weighting $w(\omega_i)$ is set to unity in the above derivation. If this is not done, a weighted least-squares algorithm should be used.

### Estimation of Markov Parameters

After obtaining a solution to Eq. (14), it is now necessary to construct the system Markov parameters. Equation (12) can be written as

$$\left(\sum_{i=0}^{\infty} A_i z^{-i}\right) \left(\sum_{i=0}^{\infty} Y_i z^{-i}\right) = \sum_{i=0}^{\infty} B_i z^{-i}. \quad (19)$$

From this relation, the following equations are derived by equating terms of like powers and recalling Eqs. (10) and (11):\[4\]

$$Y_o = B_o \quad (20)$$
\[ Y_k = B_k - \sum_{j=1}^{k} A_j Y_{k-j} \quad \text{for} \quad k = 1, \ldots, p \]  \tag{21}

\[ Y_k = -\sum_{j=1}^{k} A_j Y_{k-j} \quad \text{for} \quad k = p + 1, \ldots, \infty \]  \tag{22}

Using the estimated \( \hat{A}_i \) and \( \hat{B}_i \) instead of the true \( A_i \) and \( B_i \) in Eqs. (20) through (22), thus yields a set of estimated Markov parameters with as many terms as desired.

Note that there are only \( p \) independent Markov parameters excluding \( Y_0 \); all the other parameters are linear combinations (with matrix coefficients) of this independent set. Therefore, the Markov parameters represent a signal containing a maximum of \( m \times p \) states where \( m \) is the number of outputs.

### Identification of State-Space Models

Once the Markov parameters are constructed, the Eigensystem Realization Algorithm (ERA) or the Eigensystem Realization Algorithm using Data Correlation (ERA/DC) can be used to obtain a state-space model. The ERA (or ERA/DC) has been proved valuable for modal parameter identification from the pulse response. The ERA uses singular value decomposition to decompose a data matrix (referred to as the general Hankel matrix) and to compute a state-space model from the decomposed matrices. The system order is determined by examining the magnitudes of the singular values, where the small values are ascribed to the effects of noise and are truncated. The number of the retained singular values is taken as the order of the system.

Because the Markov parameters contain at most \( m \times p \) states, the order assigned to the system cannot exceed the number \( m \times p \). Real systems, in theory, have an infinite dimension; therefore, the larger the number \( p \) is chosen, the better the results will be. However, a large \( p \) causes intensive calculation. Thus there is a trade-off between accuracy and computation. In practice, by examining the peaks in the FRF, an approximate number of dominant modes can be estimated to assist the selection of a proper value for \( p \).
Illustrative Examples

Two examples are given — one simulated example and one experimental example. The simulated example uses a model of a 60-ft-long cantilevered truss structure, whereas the experimental example is a 55-ft-long truss with a 15-ft diameter antenna.

Example 1

The first example uses an analytical model of the Mini-Mast structure described and reported in Ref. 10. The model has five fundamental modes. The first two modes are closely spaced, representing the first bending modes in two orthogonal planes. The third mode is the torsion model while the last two modes are also closely spaced, representing the second bending modes. The discrete-time state-space parameters of this system using a sampling rate of 30 Hz are listed as follows:

\[
A = \text{diag}\left(\begin{bmatrix} 0.9859 & 0.1500 \\ -0.1500 & 0.9859 \end{bmatrix}, \begin{bmatrix} 0.9859 & 0.1501 \\ -0.1501 & 0.9859 \end{bmatrix}, \begin{bmatrix} -0.7257 & 0.6736 \\ -0.7257 & 0.6736 \end{bmatrix}, \begin{bmatrix} 0.9064 & 0.3943 \\ -0.9064 & 0.3943 \end{bmatrix}, \begin{bmatrix} 0.3943 & 0.9064 \\ -0.3943 & 0.9064 \end{bmatrix}\right)
\]

\[
B = \begin{bmatrix} -0.0407 & -0.0454 \\ -0.5384 & -0.6001 \\ 0.0746 & -0.0669 \\ 0.9867 & -0.8850 \\ 0.0164 & 0.0373 \\ 0.0376 & 0.0860 \\ -0.0460 & -0.0421 \\ -0.0711 & -0.0650 \\ 0.0653 & -0.0655 \\ 0.0997 & -0.1000 \end{bmatrix} \times 10^{-6}
\]

\[
C^T = \begin{bmatrix} 0.8570 & 1.8701 \\ -0.0000 & 0.0000 \\ 1.5700 & -1.2390 \\ -0.0000 & -0.0000 \\ 1.4030 & 1.4254 \\ 0.0000 & 0.0000 \\ 0.9016 & 1.7852 \\ -0.0000 & 0.0002 \\ 1.3509 & -1.4748 \\ -0.0000 & 0.0000 \end{bmatrix}
\]

\[
D = \begin{bmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \end{bmatrix}
\]
With these parameters, one can calculate the frequency response function (FRF) according to Eq. (3). Two hundred frequency data points equally spaced in a frequency range from 0 to 16.67 Hz are calculated. Assuming the orders of $A(z^{-1})$ and $B(z^{-1})$ are ten (i.e., $p=10$) and solving Eq. (14), one obtains an estimated $\hat{\Theta}$. This set of parameters match the FRF through Eq. (13) exactly, as shown in Fig. 1. The $(1,1)$ elements of the true FRF and of the estimated FRF (calculated using $\hat{\Theta}$) are plotted in the same figure. Note the two curves match exactly and thus can not be distinguished. The error between the two curves is within the numerical precision of the computer. Comparisons of the other FRF elements are similar.

Figure 2 compares the true Markov parameters and the estimated Markov parameters, calculated based on $\hat{\Theta}$. Only the $(1,1)$ elements are shown as the other elements have similar results. Four hundred points of the Markov parameters are calculated. The two curves also coincide within an error of order $10^{-16}$. In this case the ERA (or ERA/DC) is capable of recovering the true state space model (under some equivalence transformation), and the reconstructed FRF also matches the true FRF perfectly.

Figure 3 shows a comparison between the true Markov parameters and those parameters obtained using the inverse discrete Fourier transform (IDFT) method. There are 201 frequency samples available, including data at $\omega T = 0$ and $\omega T = \pi$; therefore, 400 points of the Markov parameters can be calculated. Due to the finite number of frequency samples and long-duration pulse response of the system, the Markov parameter thus obtained involves distortion as shown in Figure 3, even though the FRF data are perfect. Additional frequency response data (by decreasing the frequency increment to increase the resolution) will reduce the distortion. However, totally eliminating the distortion requires a large number of FRF data points for this lightly damped system. The Markov parameters obtained using the method proposed in this paper do not have distortion, and the number of data required is small.
Example 2

The second example uses experimental data taken from the CSI Evolutionary Model (CEM). The CEM shown in Fig. 4 is a NASA testbed to study controls and structures interaction problem.[11] The system has eight inputs and eight collocated outputs for control. The inputs are air thrusters and the outputs are accelerometers. The locations of the input-output pairs are depicted in Fig. 4. In this example, the structure was excited using random input signals to four thrusters located at positions 1, 2, 6, 7. The input and output signals were filtered using low-pass digital filters with the range set to 78% of the Nyquist frequency (12.8Hz) to concentrate the energy in the low frequency range below 10 Hz. A total of 2048 data points at a sampling rate of 25.6 Hz from each sensor are used for identification. In this example, sixteen FRFs from four input and output pairs located at positions 1, 2, 6, 7 are simultaneously used to identify a state space system model to represent the CEM.

The order of the matrix polynomial is set to $p=25$, which is sufficient to match as many as 50 modes (a system of dimension 100). A state-space model is obtained using ERA/DC with the system order assigned to 100. The reconstructed frequency response data (dash lines) are compared with the experimental data (solid lines) in Figs. 5 and 6. Figure 5 is the frequency response of output 1 with respect to input 1, representing a case of a strong signal, while Fig. 6 is the frequency response of output 2 with respect to input 1, representing a case of a weak signal. The signal is weak because sensor 2 is orthogonal to input 1. Similar results are obtained for other input/output pairs which are not shown in this paper. The results show that the matching is better for the strong signal cases. This is expected because the strong signal has a larger signal-to-noise ratio than the weaker signal. The results for other input-output pairs are similar and hence omitted.
Conclusion

A novel method of identifying state-space models from frequency response data is developed. The method uses only a linear technique in the frequency domain without iteration to obtain the Markov parameters from frequency response data. It is capable of avoiding windowing distortions inherent to other frequency domain algorithms. In this method, the frequency response data given is assumed reliable. For accurate frequency response data of an ideal linear system the method can identify the system perfectly; for experimental data of real systems the method provides a least-squares fit to the frequency response data. The method was tested on both numerical and experimental data sets with successful results.

Acknowledgment

The experimental data in Example 2 were generated by Dr. Kenny B. Elliott of NASA Langley Research Center and Mr. Joseph W. Gregory of Lockheed Engineering and Sciences Company. The frequency response functions (FRF) of the experimental data were provided by Mr. Gregory. Dr. Elliott and Mr. Gregory are currently continuing with the analysis of this data as well as other data sets associated with different experiments.

References


Fig. 1. The least-squares matching of the frequency response function (the (1,1) element) of Mini-Mast (the estimated data coincides with the true data).
Fig. 2. Comparison of the true and the estimated Markov parameters (pulse response histories) of the Mini-Mast.

Fig. 3. Comparison of the true Markov parameters (pulse response history) of the Mini-Mast and the IDFT-recovered Markov parameters (pulse response history).
Fig. 4 CSI Evolutionary Phase Zero Model
Fig. 5. Comparison of the test FRF (solid line) and the reconstructed FRF (dash line) obtained using the identified system matrices.
Fig. 6. Comparison of the test FRF (solid line) and the reconstructed FRF (dash line) obtained using the identified system matrices.
This paper presents an algorithm for identifying state-space models from frequency response data of linear systems. A matrix-fraction description of the transfer function is employed to curve-fit the frequency response data, using the least-squares method. The parameters of the matrix-fraction representation are then used to construct the Markov parameters of the system. Finally, state-space models are obtained through the Eigensystem Realization Algorithm using the Markov parameters. The main advantage of this approach is that the curve-fitting and the Markov-parameter-construction are linear problems which avoid the difficulties of nonlinear optimization of other approaches. Another advantage is that it avoids windowing distortions associated with other frequency domain methods.