On the Identification of High-order Lightly-damped Multivariable Systems [1]

Darin Fisher
Deborah Jue
Andrew Packard
Kameshwar Poolla[†]
Department of Mechanical Engineering
University of California, Berkeley

Abstract

This paper describes a practical off-line approach to system identification of very high-order, lightly-damped, multivariable systems. In particular, we address the practical and computational aspects of this problem. We also discuss the various choices that the modeler must make interactively. Our eventual objective is to enable "automated" modeling for such systems by minimizing the burden on the modeler.

We begin with sampled measurements of the system's frequency response to obtain an initial model. The frequency domain data could be supplied from finite-element simulations, or from traditional sine-sweep experiments. This model is further refined based on time-domain data as it becomes available. The method is designed to identify very high-order systems by solving several low-order identification problems using the iterative least squares algorithm of Santhanam-Koerner, coupled with the multi-band strategy of Bayard. The individual solutions are combined to yield a transfer-function matrix model of the overall system. A succinct state-space realization of this estimate may be obtained using standard model reduction methods.

This approach has made it possible to readily estimate systems with hundreds of modes, thousands of frequency data points, and hundreds of input/output channels using standard PC hardware.

1 Introduction

Current identification techniques for fitting non-parametric frequency domain data often fall short of being able to identify lightly-damped, broadband systems [2, 5, 10]. Such techniques typically model the plant as a rational function, in which the order of the numerator and denominator polynomials is chosen to agree with the order of the system being estimated. The numerical problems associated with high-order polynomials imposes a ceiling on the size and complexity of the problems that can be solved. As described in [1], this difficulty can be overcome by utilizing composite curve fitting, in which two (or more) models are identified in parallel, with each model chosen to fit the system dynamics accurately over a fixed frequency band.

The present paper provides an overview of composite curve fitting in the context of continuous-time, multi-input multi-output (MIMO) system identification. A minimal state-space realization is supplied as the eventual model. In addition, model uncertainty as captured by parameter variance information is also supplied. This makes the overall technique compatible with modern robust control methodologies [13, 9, 7]. Finally, an approach to time-domain refinement is presented.

The remainder of this paper is organized as follows. Background on the Santhanam-Koerner [15] curve fitting algorithm and on Bayard's multi-band fitting procedure is presented in Section 2. Computational details of treating MIMO plants using these methods are discussed in Section 3. Following this, Section 4 deals with model refinement based on fresh time-domain data. An illustrative example is offered in the final section.

Motivation for this work has come in part to develop specific tools for system identification of the NASA funded Segmented Reflector Telescope in place at the Control and Structures Research Laboratory (CSRL) at the California State University in Los Angeles. Resulting models will be used in the formulation of high-resolution control designs. It is hoped that these techniques will be useful in the greater context of system

[1] Supported in part by NASA grant #NAG5-6757.
[†] Corresponding author. Department of Mechanical Engineering, University of California, Berkeley, CA 94720, Tel. (510) 642-4642, Email: poolla@jagger.me.berkeley.edu

identification for lightly-damped flexible structures.

The m-input, p-output, continuous-time, linear time-invariant system to be estimated will be written H. The frequency response of H at frequency ω_k will be written H_k. Frequency response *data* will be denoted G_k^d. For a matrix $M \in \mathbf{C}^{p\times m}$, $\mathrm{vec}\,(M) \in \mathbf{C}^{pm}$ denotes the vector formed by stacking its columns.

We exclusively employ Euclidean norms for vectors and Frobenius norms for matrices.

2 Background

We begin with (matrix-valued) frequency response data $\{G_k^d\}_{k=1}^L$ at the ordered frequencies $\{\omega_k\}_{k=1}^L$. This could be supplied from finite element simulations or from traditional multi-tone testing [12, 3]. The problem then, is to find a rational transfer function matrix $\hat{H}(s)$ of pre-specified order that best fits the data. More precisely, we would like to solve the following nonlinear optimization problem:

$$\hat{H} = \arg\min_H \sum_{k=1}^{L} \left\| W_k (G_k^d - H_k) \right\|^2 \qquad (1)$$

where H is of the form

$$H = \frac{N(s)}{d(s)} = \frac{\sum_0^n N_i s^i}{\sum_0^n d_i s^i}, \quad \text{with} \quad d_n = 1$$

Here W_k are weights that reflect the relative importance of various frequencies. It is straightforward to incorporate left and/or right weights as needed.

2.1 SK Algorithm

The nonlinear optimization problem (1) can be effectively solved using the Santhanam-Koerner (SK) algorithm [15], the essential idea of which is to utilize iterative least-squares methods. The optimization problem (1) can be recast as

$$\min_H \left\| W \left(G^d - \frac{N(s)}{d(s)} \right) \right\| = \min_H \left\| \frac{W(G^d d(s) - N(s))}{d(s)} \right\|$$

This immediately suggests the SK algorithm: Initialize with $i = 1$ and an initial estimate N^1, d^1. Subsequent estimates are obtained by solving

$$\left(N^{k+1}, d^{k+1}\right) = \arg\min_{N,d} \left\| \frac{W(G^d d(s) - N(s))}{d^k(s)} \right\| \qquad (2)$$

If $N(s)$ and $d(s)$ are parameterized using linear basis functions expansions (see for example [17]), this is an attractive least-squares problem.

The SK algorithm is known not to be globally convergent. In particular, the resulting model may be *unstable*. Extensive experimental evidence suggests that it performs remarkably well. However, for very high order systems and especially for lightly damped frequency responses, the SK algorithm can yield spurious results.

2.2 Bayard's Multi-Band Fitting

The shortcomings of the SK algorithm as applied to very high-order lightly damped systems have motivated the multi-band fitting methods of Bayard [1, 4]. The essential idea is to divide the frequencies of interest into smaller sub-bands. Models that fit the data accurately on each band are determined using the SK algorithm, and these are combined to produce a model for the overall system. This is described more precisely below.

We restrict our attention to two bands in solving the optimization problem (1). Define complimentary weighting functions W^ℓ and W^h by

$$W_k^\ell = \begin{cases} 1 & \text{for } \omega \le \overline{\omega} \\ 0 & \text{else} \end{cases} \qquad (3)$$

and $W^h = 1 - W^\ell$. Initialize the model $W^h = 0$ to serve as a good fit for the frequency response data in the high-frequency band $\omega > \overline{\omega}$. Determine a good model H^ℓ in the low-frequency band using the SK algorithm:

$$H^\ell = \arg\min_H \|W^\ell W (G^d - H - H^h)\|$$

Next, determine a good high-frequency model by fitting the *residuals* using the SK algorithm:

$$H^h = \arg\min_H \|W^h W (G^d - H^\ell - H)\|$$

Iterating between these two steps will produce a model $(H^\ell + H^h)$ that serves to match the wide-band frequency response data. This procedure circumvents the numerical problems associated with the SK procedure alluded to earlier.

3 Computational Details

In this section we present computational details relevant to applying the methods discussed in the previous section to high-order, lightly-damped multivariable plants.

The computational details reveal that the modeling of very high-order lightly-damped systems is an interactive process. The modeler must make several parameter selections based on inspections of residuals. Our eventual objective is to enable an "automated" modeling procedure by reducing these choices to the bare minimum.

3.1 Multivariable Systems

A difficulty that commonly arises when dealing with transfer-function *matrices* involves model order (see for example [12] in the context of multivariable ARX system modeling). In contrast with state-space realizations, these descriptions are not naturally succinct (for multivariable systems). We deal with these issues by vectorizing the problem formulations to transform it into a multi-output single-input modeling problem. This will likely result in models of unnecessarily high order, as they fail to be minimal. This is not critical as a succinct state-space realization of our model may be obtained using standard model reduction methods (such as balanced truncation).

The potential problems of this approach are computational. The number of parameters being estimated in transfer-function matrix descriptions is excessive, and this can lead to high parameter variance. In addition, storage of regression matrices that arise in the least squares iterations of the SK algorithms may be an issue. An important and difficult open problem is to develop direct state-space methods to solve *frequency*-domain modeling problems such as (1), perhaps along the lines of [16, 6].

3.2 SK Algorithm

One of the difficulties associated with the SK procedure is due to numerical conditioning problems associated with basis function choice. For continuous-time problems, the associated least-squares problems involve Vandermonde matrices containing large powers of ω_k. This difficulty can be handled by using a bilinear transformation to map the frequencies $\{\omega_k\}$ into the unit circle, and solving the resulting discrete-time problem. The resulting model is then re-transformed into a continuous-time model by inverting the bilinear transformation. This procedure preserves rationality of transfer function matrix models.

There is flexibility in this bilinear transformation. We can select any continuous frequency α, and have it mapped to the discrete frequency $\pi/2$. This flexibility permits us to "warp" the data so that the continuous frequency range of interest (centered around α) is favorably placed on the unit circle. The choice of the the "warp" frequency is left to the modeler, but a reasonable choice is the geometric mean of the minimum and maximum frequencies, i.e. $\alpha = \sqrt{\omega_1 \omega_L}$.

Where the SK algorithm applies $1/d^k(s)$ as an additional weight to solve for N^{k+1} and d^{k+1}, we have found that weighting by $1/\sqrt{d^k(s)}$ helps obtain a more uniform fit, as discussed in [14].

The actual number of SK iterations required to achieve a "reasonable" fit may vary from 1 to 20, or even more. We have observed that fitting "ideal" data with the exact number of states results in a non-converging iteration (in the 2-norm sense) in the least squares. However, fitting "real" data (i.e., noisy) with the exact number of states results in a converging iteration. We use this fact in determining the number of SK iterations and automatically stop iterating once the relative cost (in the 2-norm sense) falls below a specified tolerance.

In the event that the SK iteration produces an unstable model, the modeler is encouraged to select a lower model order. This can be automated or left to the modeler's discretion.

3.3 Bayard's Multi-Band Fitting

As pointed out in [1], the use of *indicator* weighting functions W^ℓ and W^h (3) is to be avoided because the DC bias is incorrectly identified by this procedure. Bayard suggests a modification that addresses this problem and also improves the "dove-tailing" of the composite model in the low-frequency band.

We suggest the use of three tiered weights. For the frequency band Ω_b, the weight we employ is

$$W_k^b = \begin{cases} \mathcal{W}_\ell & \text{if } \omega_k < \omega, \text{ for all } \omega \in \Omega_b \\ \mathcal{W}_h & \text{if } \omega_k > \omega, \text{ for all } \omega \in \Omega_b \\ 1 & \text{else} \end{cases} \qquad (4)$$

where $\mathcal{W}_\ell$ and $\mathcal{W}_h$ are generally less than or equal to 1. With this particular choice of weight, the degree to which the estimate in the current band is improved, at the cost of the estimate in the others, can be controlled. It should be noted that Bayard's weight is a special case of this in which $\mathcal{W}_\ell = 1$ and $\mathcal{W}_h = 0$. We have found that choosing $\mathcal{W}_\ell \approx 10^{-1}$ and $\mathcal{W}_h \approx 10^{-4}$ typically results in better overall performance. The entire procedure generally converges much sooner and requires far fewer SK iterations to estimate each band.

Since the choice of $\mathcal{W}_\ell$ and $\mathcal{W}_h$ will often depend on the nature of the problem being solved, we add these parameters along with model order m and α to the list of parameters that the modeler can select when fitting each band.

3.4 Assessing Model Accuracy

It is essential to assess the quality of the model obtained from our modeling procedure. For example, this information is vital to the design of robust controllers that achieve optimal disturbance rejection, or to refine the model based on subsequent time-domain data (see Section 4 below). The model quality can be captured classically by the covariance matrix Λ of the state-space parameters $\theta = [A\ B;\ C\ D]$. In what follows, we ignore cross-covariances and offer formulae for the covariance σ_{ij} of θ_{ij}.

As is well known [12], the Fisher information matrix

F provides an asymptotically unbiased estimate of Λ. A classical approximation to F is supplied for nonlinear least-squares problems by the gradient of the cost function. More precisely, define

$$\begin{aligned} J(\theta) &= \frac{1}{2}\left\|W(G^d - H(\theta))\right\|^2 \\ &= \frac{1}{2}\sum_1^L \left\|W_k(G_k^d - H(\theta)_k)\right\|^2 \end{aligned}$$

Then,

$$\sigma_{ij} = \left(\frac{\partial J(\theta)}{\partial \theta_{ij}}\right)^{-2} \tag{5}$$

The gradients $\frac{\partial J(\theta)}{\partial \theta_{ij}}$ evaluated at $\theta = [AB; CD]$ remain to be determined. For this, define

$$E = \sum_{k=1}^{L} E_k = \sum_{k=1}^{L} \left[CT_kB + D - G^d(j\omega_k)\right]$$

and

$$T_k = (j\omega_k I_n - A)^{-1}.$$

It can be shown that

$$\frac{\partial J}{\partial \theta} = \sum_{k=1}^{L} Re\left\{\frac{\partial J_k}{\partial \theta}\right\} \tag{6}$$

where

$$\frac{\partial J_k}{\partial \theta} = \begin{bmatrix} (CT_k)^* \\ I_p \end{bmatrix} E_k \begin{bmatrix} (T_kB)^* & I_m \end{bmatrix}. \tag{7}$$

4 Time-Domain Refinement

As the operating conditions of the system change, the system dynamics will incur small drifts. These drifts may be due to factors such as environmental effects, material fatigue, or micro-loading. In many situations, it becomes important to obtain fresh models that better reflect the current operating conditions. For example, in very-high performance disturbance rejection problems for flexible structures, it will become necessary to obtain increasingly accurate system models (e.g., at cross-over frequencies).

As a result, the model obtained using frequency response data, as described in the earlier sections, needs to be periodically updated. Obtaining fresh complete frequency response measurements can be time-consuming and expensive, especially for multi-input lightly damped flexible structures. We therefore suggest using time-domain data directly, as it becomes available. This data can be employed to refine the model using maximum-likelihood (ML) parameter estimation methods.

More precisely, let $M^o = [A^o \;\; B^o;\;\; C^o \;\; D^o]$ be a (packed) state-space realization of our current estimated model, and let $\theta^o = \mathrm{vec}\,(M^o)$. Let Λ be the covariance matrix of our current estimate. Define $\theta = \mathrm{vec}\,(M)$, where the matrix M will serve as the (packed) state-space realization of our refined model. Let $H(\theta)$ denote the LTI system with vectorized (packed) system matrix θ. Suppose we have available time-domain input-output data u^d and y^d respectively. The associated likelihood function is

$$J(\theta) = \|y^d - H(\theta)u^d\|^2 + \frac{(\theta - \theta^o)'\Lambda^{-1}(\theta - \theta^o)}{2} \tag{8}$$

and the parameters of the refined model are obtained by minimizing $J(\theta)$. Note that the second term in the likelihood function serves to incorporate confidence information of our prior estimate. The likelihood function is minimized using standard nonlinear least squares methods [8, 11]. This requires gradient and/or Gauss-Newton descent direction computations. Given that we are refining a reasonably accurate initial model, Gauss-Newton descent methods are preferable given their super-linear convergence behavior. We can also obtain a new estimate of parameter variance using the Hessian of $J(\theta)$ to approximate the Fisher information matrix.

The loss of identifiability demands careful treatment. This is not material for gradient methods, but is essential for Gauss-Newton methods. An effective technique here is to select a local canonical form (such as a modal realization where the "A" matrix consists of 2×2 blocks along the main diagonal), and to confine the nonlinear least squares to perturbations that preserve the canonical form.

The associated calculations are reasonable. Space precludes us from supplying all relevant formulae. These may be found in [18].

5 Example

In this section, we present an example of frequency response fitting using the multi-band method discussed above. The example is of a single-input single-output system with 240 randomly generated, lightly-damped modes. Due to space limitations, we refrain from showing multivariable results. The frequency response serves as the data to our estimation algorithm and is plotted in Figure 1.

The resonant modes of the system were chosen so as to form pairs of nearly overlapping modes, representative of symmetry found in many structures.

The upper plot of Figure 2 compares the result of fitting the data with one band, where the lower line is

a plot of the residual (estimation error). The error norm is 7.240e-3, which represents the 2-norm of the percentage error, divided by the number of frequency points. This estimate was obtained using a pre-warp term of 63.3 rad/sec and 66 states. Higher model orders resulted in an unstable system or a higher error norm.

The values used in this example for α are listed in Table 1. Note that the pre-warp terms have not been optimized for this example; we used the default values, which were obtained from the geometric means of the frequency endpoints for each band.

	m	α (rad/sec)	$\mathcal{W}_\ell$	$\mathcal{W}_h$
band 1	50	7.4	.2	.0001
band 2	44	84.0	.2	.0001
band 3	30	714.8	.2	.0001

Table 1: Estimation parameters.

Let the set Ω_b contain the frequency points of the b-th frequency band. We can then define the b-th weight W^b as in (4). Table 1 lists the values of $\mathcal{W}_\ell$ and $\mathcal{W}_h$ used for each band. Note that this particular choice of weighting functions has the effect of *de*-emphasizing the fit at high-frequencies, to prefer the fit at lower frequencies. Biasing the weighting functions in this way has shown to limit DC drift introduced as the higher bands are fit.

As discussed previously, the number of SK-iterations required for each band can vary considerably, and indeed in this example the earlier cycles required far more SK-iterations than the later. From our experience, the number of SK-iterations required seem to converge along with the norm of the estimation error.

Our fit using a multi-band approach is shown in the lower plot of Figure 2; again the lower line is a plot of the residual (estimation error). The error norm for a 124 state estimate is 1.8245e-3. If we perform a balanced truncation on the estimated system down to 66 states, we obtain an error norm of 2.1556e-3, which is still 3.3 times better than the one-band estimate.

A comparison of the residuals of the time response due to a pseudo-random input for the one-band and multi-band estimates is shown in Figure 3. For these plots, we took the difference between the time-response of the actual plant and the time-response of each estimate. The error norm (i.e., the 2-norm divided by the data length) for each of the time response plots is: 0.4055 for the one-band fit, and .0976 for the truncated three-band fit.

From this example, it is clear that multi-band fitting and the SK algorithm can be effectively used to estimate lightly-damped systems of extremely high-order.

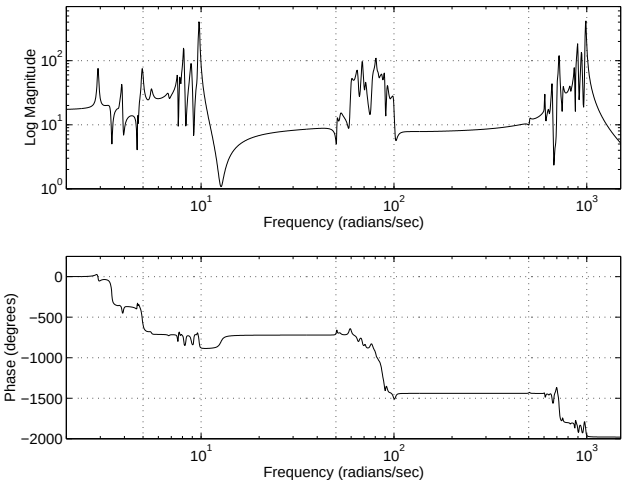

Figure 1: Frequency response data.

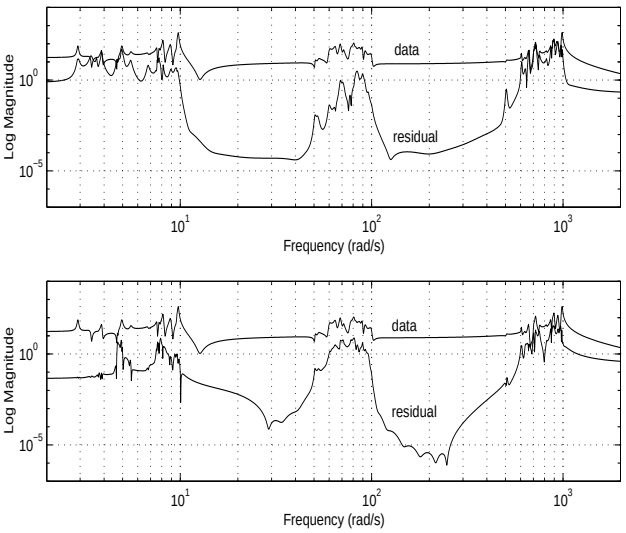

Figure 2: Estimated transfer functions plotted overtop data, One-Band Fit (top), Truncated Three-Band Fit (bottom).

6 Conclusions

In this paper we have combined various elements of existing methodologies and presented a procedure for the purpose of identifying high-order lightly damped multivariable systems. This approach generally requires the user to select values for several key parameters; unfortunately the resulting fit can be highly sensitive to these choices. It has been our experience that the choice of cut-off frequencies and corresponding pre-warp frequencies are the most critical. Our procedure

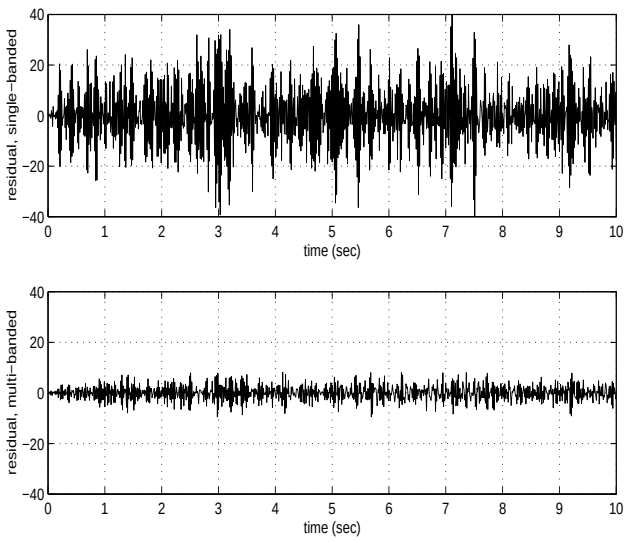

Figure 3: Residual Time Response to a Pseudo-Random Input For One-Band(top) and Three-Band Truncated(bottom) Estimated Systems.

is less sensitive to choices of band weights or iteration schedule (i.e., the sequence of iterations performed, varying band number and model order). Sub-optimal choices here result only in an increased number of computations to arrive at the "final" parameter estimates.

While we have determined some heuristic approaches to selection of the cut-off frequencies and pre-warp terms, we feel that an analytical or iterative-calculation based algorithm would be of most benefit. We have identified several problems which require further study in order to meet our goal of developing a completely automated solution. These include:

- Optimal cut-off frequency selection
- Determination of lower and upper error bounds for a set of cut-off frequencies and model orders
- Optimal pre-warp term selection
- Optimal band weights selection
- Computing algorithm which sifts through the above optimizations and then iterates through the bands and model orders

References

[1] David. S. Bayard, "High-Order Wide-Band Frequency Domain Frequency Domain Identification using Composite Curve Fitting," *American Control Conference*, pp. 3181-5, 1992.

[2] David. S. Bayard, "An Algorithm for State-Space Frequency Domain Identification Without Windowing Distortions," *Proceedings of the 31st Conference on Decision and Control*, pp. 1707-12, 1992.

[3] D. S. Bayard, F. Y. Hadaegh, Y. Yam, R. E. SCHEID, E. Mettler and M. H. Milman, "Automated On-orbit Frequency Domain Identification for Large Space Structures," *Automatica*, vol. 27, No. 6, pp. 931-46, 1991.

[4] David S. Bayard "Multivariable Frequency Domain Identification via 2-Norm Minimization," *American Control Conference*, pp. 1253-1257, 1992.

[5] Jean-Noel Aubrun, Alain C. Carrier, Kenneth R. Lorell, "Active Control for Segmented Mirror Optical Systems," *Control and Dynamic Systems*, Vol. 57, 1993.

[6] Young Man Cho, Guanghan Xu, T. Kailath, "Fast Identification of State-Space models via Exploitation of Displacement Structure," *IEEE Transactions on Automatic Control*,vol. 39, No. 10, pp. 2004-2017, 1994.

[7] M. Dahleh and J. B. Pearson, "ℓ^1-Optimal controllers for MIMO discrete-time Systems," *IEEE Trans. Automatic Control,* vol. AC-32, pp. 314-323, 1987.

[8] J.E. Dennis, Jr. and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice-Hall, 1983.

[9] J. Doyle, K. Glover, P. Khargonekar, and B. Francis, "State-space solutions to standard H_2 and H_∞ control problems," *IEEE Trans. Automat. Contr.*, vol. AC-34, no. 8, pp. 831–847, 1989.

[10] Jonathan H. Friedman, "Identification, Modeling, and Control of Flexible Structures," Ph.D. Dissertation, University of Michigan, 1996.

[11] P. E. Gill, W. Murray, and M. H. Wright, *Practical optimization*, Academic Press, 1981.

[12] L. Ljung, *System Identification, Theory for the User*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.

[13] A. Packard and J.C. Doyle, "The complex structured singular value," *Automatica*, vol. 29, no. 9, pp:71–110, 1993.

[14] Sundeep Rangan, "Stochastic H_∞ Identification: An Iteratively Weighted Least Squares Approach," M.S. Thesis, University of California, 1994.

[15] C. Santhanum, J. Koerner, "Transfer function synthesis as a ratio of two complex polynomials," *IEEE Transactions on Automatic Control*, vol. 8, pp. 56-58, 1963.

[16] M. Verhaegen, "Identification of the Deterministic Part of MIMO State Space Models given in Innovations Form from Input-Output Data", *Automatica*,vol 20, No. 1, pp. 61-74, 1994.

[17] Bo Wahlberg, "System Identification using Laguerre Models," *IEEE Transactions on Automatic Control*,vol. 36,No. 5, pp. 551-561, 1994.

[18] G. Wolodkin, S. Rangan, and K. Poolla, "An LFT Approach to Parameter Estimation," *Proc. of the Automatic Control Conference*, San Francisco, pp. 2088-2092, June 1997.