AIAA JOURNAL
Vol. 31, No. 5, May 1993

Approximation of Parameter Uncertainty in Nonlinear Optimization-Based Parameter Estimation Schemes

W. R. Witkowski* and J. J. Allen*
Sandia National Laboratories, Albuquerque, New Mexico 87185

Theoretical models are often developed using experimental data to estimate the model parameters, but an evaluation is not always made of the model's reliability or robustness. Errors can exist in the theoretical model for many reasons, including inadequate model formulation and poor parameter estimates. This paper investigates the approximation of second-order statistics (confidence regions) of parameters estimated using nonlinear optimization schemes. An optimal set of parameters is calculated to minimize the difference between theoretical model predictions and experimental data. This metric is formulated in terms of a weighted least squares objective function. A truss structure was analyzed to evaluate nonlinear optimization and confidence region approximation techniques. It was found that since the truss model was linear and parameter interaction was small, the confidence region approximations were reliable.

I. Introduction

A MATHEMATICAL model is only as good as its estimated parameters. The ability to accurately model physical processes mathematically is vital in all areas of scientific research. In the structural engineering field, the ability to predict a system's behavior leads not only to the efficient design of new structures but also to the optimal operation and control of existing ones. Unfortunately, an exact model is usually unavailable and an approximate one must be used. Model parameters are typically determined by fitting the model to experimental data. The procedure of formulating a mathematical model, estimating its parameters, and validating its structure is referred to as *model identification,* which can involve both state and parameter estimations.

Nonlinear optimization schemes have become very popular for use as parameter estimation methods.[1-4] This popularity stems from the fact that the analysis scheme is not dependent on a particular model form or on the type of experimental data available. The minimization of the summed squared error between model predictions and experimental data is commonly used as a criterion of optimality. Least squares theory states that at the optimum, linearized estimates of higher order statistics can be approximated for the estimated parameters, assuming that the measurement errors have zero mean, are independent, and have constant variances.

Unfortunately, these linearization schemes provide reliable confidence intervals only when the objective function, or response surface, can be well represented with a Taylor series at the optimum, and therefore these schemes apply only approximately to nonlinear models. When the model form does not fit the required criteria, the approximation schemes can provide confidence intervals that are severely underestimated, providing a false measure of parameter reliability. Several investigators have devised techniques to evaluate the objective function curvature[5,6] to help analyze the validity of the assumptions made.

This paper further analyzes the model identification results achieved by Allen and Martinez[1] for a truss structure. In this analysis, parameter estimates were found using the MSC/NASTRAN software package for finite element analyses and optimization.

Received Jan. 14, 1992; revision received Oct. 2, 1992; accepted for publication Oct. 2, 1992.

*Senior Member of Technical Staff, Structural Dynamics Department 1434.

The truss parameters that were estimated, using eigenvalue data, included Young's modulus of the members and the stiffness value of the truss support structure. Confidence intervals for the parameter estimates are calculated using different approximation schemes. The reliability of these approximations with respect to objective function curvature is also discussed.

II. Nonlinear Parameter Estimation

Nonlinear parameter estimation schemes have been successfully used to approximate model parameters.[1,4,7-9] In this approach, the difference between theoretical and experimental data is minimized. This is done by finding the optimal set of parameters that provides a model that fits the experimental data the "best." Therefore, an accurate mathematical model is of primary importance to provide reliable theoretical predictions. Otherwise, an ill-conditioned problem can be formulated or erroneous parameter estimates can be predicted.

An objective function Φ is chosen to suitably measure the error between the model and data. The choice of Φ must properly describe the experimental error and best utilize available measurements. The most commonly used formulation is the method of weighted least squares. The objective function

$$\Phi(\hat{y}, y; \Theta) = \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij}(\hat{y}_{ij} - y_{ij})^2 = \sum_{k=1}^{n*m} E_k^2 \qquad (1)$$

is chosen to be the weighted summed squared error e^2 between model predictions $\hat{y}_{ij}$ and experimental measurements y_{ij} where m is the number of experiments and n is the number of observations per experiment. The model parameters are contained within the parameter vector Θ. The weighting factors w_{ij} are used to scale variables of different magnitudes and to incorporate known information about measurement uncertainty. The following assumptions are used in this formulation: the measurement error is normally distributed with zero mean, is independent, and has a known variance that remains constant. The validity of these assumptions are verified via residual analysis.

For an unconstrained problem, an important requirement to estimate the optimal set of parameters Θ^* is that the objective function should have a minimum. Necessary and sufficient conditions for a local minimum to exist are given in Gill et al.[10] However, it should be noted that, for parameter estimate uniqueness, $\Phi(\hat{y}, y; \Theta^*)$ must have a global, not just a local, minimum. That is, it should be shown that $\Phi(\hat{y}, y; \Theta^*) = \Phi^*$ is convex in the parameter space of interest. If this is not the case, a non-unique solution may be formed at a local minimum.

The optimal parameter vector Θ^* is found when the convergence criterion,

$$\Phi(\Theta) - \Phi^* \leq \epsilon \tag{2}$$

where ϵ is an indifference criterion, is met. The indifference criterion maps out a region where values of Θ that satisfy Eq. (2) are stated to lay within the ϵ-indifference region. To evaluate the reliability of Θ^*, a confidence interval is evaluated. The confidence interval of a parameter estimate provides a quantitative assessment of the reliability and precision of the value. If a large confidence interval is calculated, the reliability of the model's prediction is low. However, if the parameters are identified with high precision (i.e., small confidence intervals), confidence in the model's predictions is increased. The confidence intervals are commonly estimated by expanding Φ around its minimum using a truncated Taylor series,

$$\Phi(\Theta) \approx \Phi^* + q^{*T}\,\delta\Theta + \tfrac{1}{2}\delta\Theta^T H^*\delta\Theta \tag{3}$$

where $\delta\Theta = \Theta - \Theta^*$ are the differences in the parameter estimates from the optimal values, and q^* and H^* are the gradient and Hessian of Φ evaluated at Θ^*, respectively. Since Θ^* is an unconstrained optimum of Φ,

$$\Phi(\Theta) \approx \Phi^* + \tfrac{1}{2}\delta\Theta^T H^*\delta\Theta \tag{4}$$

because $q^* = 0$ at the optimum. Using the convergence criterion shown in Eq. (2),

$$|\delta\Theta^T H^*\delta\Theta| \leq 2\epsilon \tag{5}$$

where ϵ is defined as the indifference criterion. Therefore, the best linearized confidence intervals are calculated by solving Eq. (5). Because the solution to Eq. (5) is computationally intensive and numerically unstable, several approximate schemes have been adopted with simplified calculations. Donaldson and Schnabel[11] evaluated five different methods to approximate confidence intervals: three variants of linearization schemes, a likelihood method, and a lack-of-fit technique were analyzed. The major distinction between the linearization schemes and the other two methods is that the former assumes that the nonlinear function can be adequately approximated by a linear function at the optimum. Bard[12] and Donaldson and Schnabel[11] stated that the most common approach used in practice is a linearization scheme in which the Hessian matrix is approximated using gradient information and the interaction between parameters is neglected. This approach is convenient because it is computationally inexpensive and numerically stable. Therefore, the confidence interval for parameter α is calculated using

$$\delta\Theta_\alpha \leq \left(\frac{2\epsilon}{A_{\alpha\alpha}}\right)^{1/2} \tag{6}$$

where A is an approximation to the Hessian matrix and $A_{\alpha\alpha}$ is the diagonal element of A corresponding to the parameter α. Bard[12] showed that the Hessian can be approximated using the Gauss method (V_a) by

$$A^{-1} \approx H^* \approx \sigma^2\left[\sum_{i=1}^{n} w_i\left(\frac{\partial\hat{y}_i}{\partial\theta_j^*}\right)\left(\frac{\partial\hat{y}_i}{\partial\theta_k^*}\right)^T\right]^{-1} = V_a \tag{7}$$

where $\hat{y}_i$ is the theoretical prediction of the ith observation, σ^2 is the residual variance, and j and k loop over the number of parameters. This is considered a good approximation if the second derivatives of the model equations, which are multiplied by the estimation error ϵ, are small. Other approaches used to estimate the covariance matrix of Θ include 1) using finite differences to approximate the Hessian, $V_b = s^2H^{-1}$, and 2) using gradient and Hessian data, $V_c = s^2H^{-1}(J^TJ)H^{-1}$ where s^2 is the population residual variance.

If the residual variance σ^2 is unknown in Eq. (7) and no replicated data are available, it is typically estimated by dividing the sum of squares of the residuals by the number of degrees of freedom,

$$\sigma^2 = \frac{1}{n-l}\sum_{i=1}^{n} e_i^{*2} = \frac{1}{n-l}\,\Phi^* \approx s^2 \tag{8}$$

where l is the number of unknown parameters.

For a specified confidence level γ, linear least squares theory states that the joint confidence region is

$$\delta\Theta^T\, V_\Theta^{-1}\,\delta\Theta \leq l\, s^2\, F_{l,n-l,1-\gamma} \tag{9}$$

in which $F_{l,n-l,1-\gamma}$ is the F-statistic with l and $n-l$ degrees of freedom,[6] whereas if the parameters are assumed independent, the marginal confidence interval is

$$\delta\Theta_\alpha \leq (V_{\alpha\alpha})^{1/2}\, s\, t_{n-l,1-(\gamma/2)} \tag{10}$$

in which $t_{n-l,1-(\gamma/2)}$ is the t-statistic with $n-l$ degrees of freedom, and $V_{\alpha\alpha}$ is the diagonal element of the covariance matrix corresponding to parameter α.[6] The population standard deviation s is approximated by the sample deviation σ. The covariance matrix is defined as

$$V_\Theta = E(\sigma\Theta^*,\ \delta\Theta^{*T}) \tag{11}$$

where $E(\cdot)$ is the operator of mathematical expectation. As stated earlier, the covariance matrix is commonly approximated by

$$V_\Theta \approx H^{*-1} \tag{12}$$

The matrix V_Θ contains all of the correlation information between parameters. The diagonal elements estimate the parameter variances and the off-diagonals estimate their covariances that determine the parameter interaction. If the parameter interdependencies are large, reparameterization schemes are often used to minimize this effect.[13] The Gauss approximation is typically considered a good estimate unless the curvature at the optimum is large. Bates and Watts[6] suggested methods to measure the curvature, or the nonlinearity of the objective function, to assess the accuracy of the linearization approximations.

III. Truss Structure Description

Figure 1 shows the truss structure that was analyzed. The truss was designed and constructed to have linear characteristics. The truss structure is made of 1-in.-diam polycarbonate tubing. The truss has 1-ft. cubic bays, which are assembled with the tubes bonded to polycarbonate corner blocks. The truss is mounted on a 2000-lb inertial mass, which is supported on airbags. Finite element modeling of the truss structure was straightforward, since the geometry and load paths were well defined.

IV. Model Analysis

The NASTRAN model of the truss contained two unknown parameters, Young's modulus E and the airbag support stiffness K. Since the restrictions of NASTRAN prohibited optimization on material properties directly, properties that were specified on the element definition cards were used. Therefore, E and K were indirectly changed through the use of two normalized parameters, $\Theta_1 = \alpha$ and $\Theta_2 = \beta$. These parameters were estimated using a nonlinear optimization approach described earlier (NASTRAN uses a modified method of the feasible directions approach). A weighted least squares formulation of the objective function was used to compare the theoretical model predictions and experimental data. Normalized parameter values were used to eliminate any scaling difficulties that could cause problems in the optimization routine. Also, the objective functions were formulated such that normalized frequencies were used.

Three different sets of data combinations were used to estimate the structure's parameters and their uncertainties. The first data set included seven elastic modes (modes 2–8), the second introduced one "bounce" mode (mode 1) in place of the last elastic mode (mode 8), and the last set contained only one bounce mode (mode 1) and one elastic mode (mode 2). Table 1 shows the experimental observations and the resulting natural frequencies of the identified system for each of the data sets (cases 1–3). All of the data sets provided sufficient information to accurately predict the modal frequencies with a maximum error less than 2.3%. Resulting parameter estimates are shown in Table 2 where $\delta\alpha$ and $\delta\beta$ are deviations from the mean value of α and β, respectively. Marginal confidence intervals (using a 95 percentile statistic for the t-statistic) are given for each parameter based on Eq. (10).

As expected, when using the first data set, the normalized support stiffness β was not identified with great confidence (0.923 ± 0.265). This is because the parameter was not highly sensitive to the elastic frequencies. However, when mode 1 was added to the data set (case 2), β was accurately identified. Case 3 was analyzed to see if one bounce and one elastic mode contained sufficient information, which they did, to identify the parameters; albeit, since the data set involved only two data points and two parameters were estimated, an infinite variance was obtained. However, even the predicted frequencies for the data not used specifically in the regression compared extremely well with experimental values.

As can be seen, the parameter estimates were nearly identical for each of the three data sets. The choice of the covariance approximation did not make a significant difference either.

The marginal confidence intervals shown in Table 2 are based only on the individual parameter, not on a joint probability region as in Eq. (9). Therefore, using Eq. (10) to estimate a joint confidence region, using a rectangular approximation, typically underestimates the theoretical confidence region,[3,5] especially if there is any coupling between parameters. The joint confidence region for case 2 (using the V_a approximation) is defined by

$$0.346\delta\alpha^2 - 0.126\delta\alpha\beta + 1.67\delta\beta^2 \leq 1.70e - 3 \tag{13}$$

where $\delta\alpha$ and $\delta\beta$ are deviations from the nominal value of α and β, respectively, and 95 percentile F-statistic is used. Figure 2 shows a graphical representation of both the marginal confidence intervals (represented by straight dashed lines) and the joint confidence region (represented by the ellipsoidal solid line). The shadowed region is the estimated joint confidence region approximated using the marginal confidence intervals. Note how this approximation severely underestimates the confidence in α.

A singular value decomposition (SVD)[14] of the Hessian matrix used to approximate the covariance matrix was used to further assess the influence of the parameter interaction on the confidence interval approximation. Since $\boldsymbol{H}$ is the symmetric and positive definite, the matrix factorization

$$\boldsymbol{H} = \boldsymbol{Q}_1 \Sigma \boldsymbol{Q}_2^T$$

decomposes $\boldsymbol{H}$ into $\boldsymbol{Q}_1 = \boldsymbol{Q}_2$ that contain the eigenvectors of $\boldsymbol{H}\boldsymbol{H}^T$ and the diagonal matrix Σ that contains the singular values σ_i. The singular values are the square roots of the nonzero eigenvalues of $\boldsymbol{H}^T\boldsymbol{H}$. This information defines a hyperellipsoid, corresponding to an unscaled confidence region, whose axes positions are described by the eigenvectors of $\boldsymbol{Q}_1$, and relative axes lengths are determined by $1/\sqrt{\sigma_i}$. For case 2, the eigenvectors are

$$\boldsymbol{Q}_1 = \begin{bmatrix} 0.999 & -0.046 \\ 0.046 & 0.999 \end{bmatrix} \tag{14}$$

and the singular values are

$$\Sigma_1 = \begin{bmatrix} 2.94 & 0 \\ 0 & 0.550 \end{bmatrix} \tag{15}$$

It is seen that there was not significant interaction between the parameters since $\boldsymbol{Q}_1$ is essentially diagonal.

The fact that the problem was quite linear and the parameter interaction was low made the confidence interval approx-

Table 1 Truss weighted least squares parameter estimation results: theoretical vs experimental frequencies (asterisks indicate regressor variables for each case)

Mode	Modal test Frequencies, Hz	Case 1 (modes 2–8) Freq., Hz	Error, %	Case 2 (modes 1–7) Freq., Hz	Error, %	Case 3 (modes 1 and 2) Freq., Hz	Error, %
1	3.53	3.60	1.98	3.57*	1.13	3.55*	0.57
2	10.28	10.30*	0.19	10.32*	0.39	10.28*	0.00
3	10.84	10.84*	0.00	10.85*	0.09	10.80	0.37
4	19.86	19.46*	2.01	19.50*	1.81	19.42	2.21
5	21.28	21.35*	0.33	21.40*	0.56	21.31	0.14
6	45.70	46.18*	2.16	46.28*	2.16	46.10	0.88
7	60.45	59.91*	0.89	60.04*	0.68	59.81	1.06
8	62.46	63.09*	1.01	63.23	1.23	62.97	0.82

Table 2 Truss weighted least squares parameter estimation results: parameter estimates and marginal confidence intervals using three different approximation schemes for the covariance matrix V

		$\bar{\alpha}$	$\pm\delta\alpha$	$\bar{\beta}$	$\pm\delta\beta$
Case 1, $s^2 = 1.46E\text{-}4$	V_a	0.781	0.0188	0.923	0.265
	V_b	0.781	0.0190	0.923	0.267
	V_c	0.781	0.0190	0.923	0.269
Case 2, $s^2 = 1.47E\text{-}4$	V_a	0.785	0.0184	0.909	0.0403
	V_b	0.785	0.0188	0.909	0.0420
	V_c	0.785	0.0182	0.909	0.0438
Case 3, $s^2 = \infty$	V_a	0.778	∞	0.899	∞
	V_b	0.778	∞	0.899	∞
	V_c	0.778	∞	0.899	∞

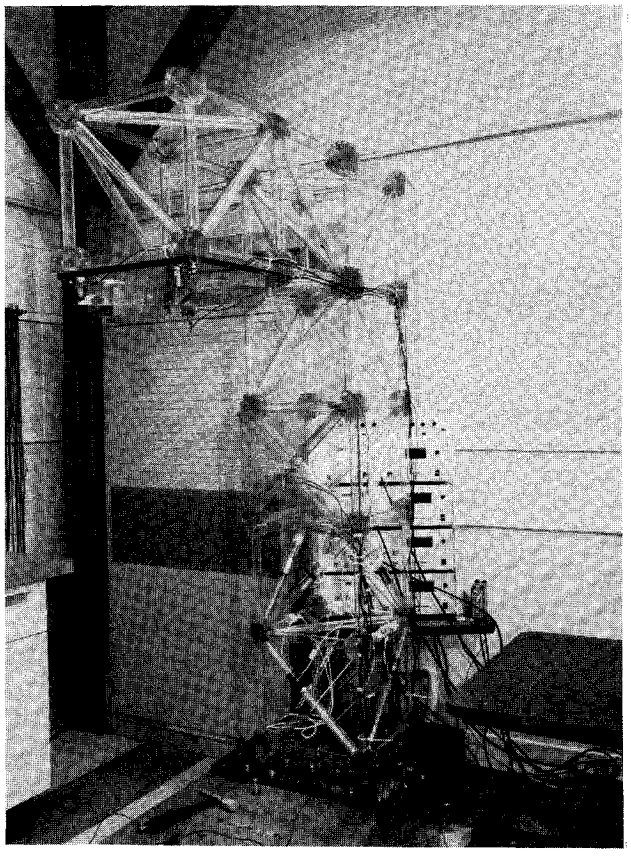

Fig. 1 Truss structure.

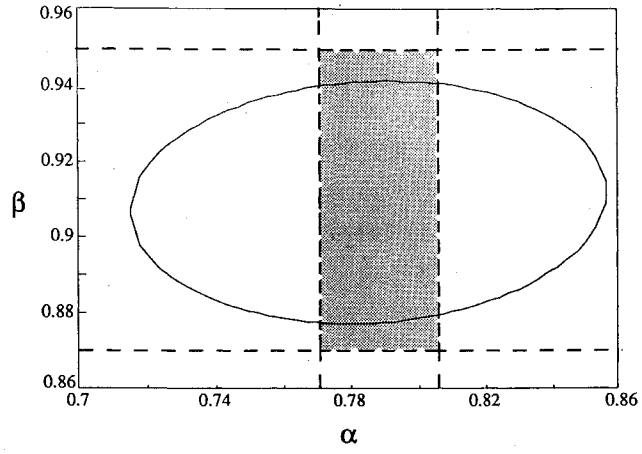

Fig. 2 Marginal (dashed lines) and joint (ellipse) confidence regions for case 2.

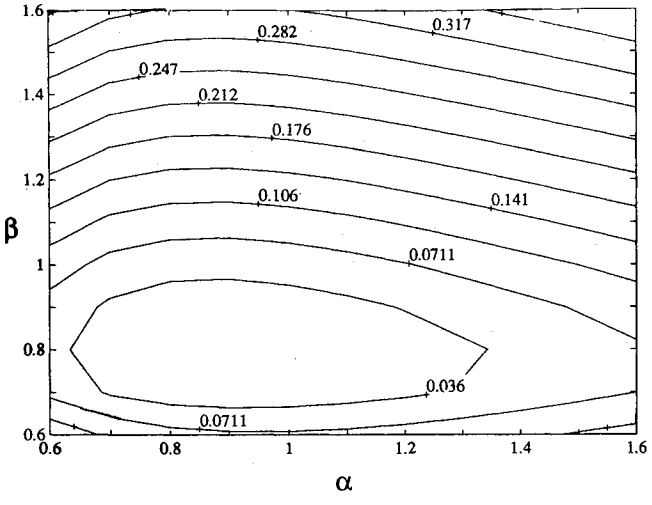

Fig. 3 Sum of squares contours for case 2.

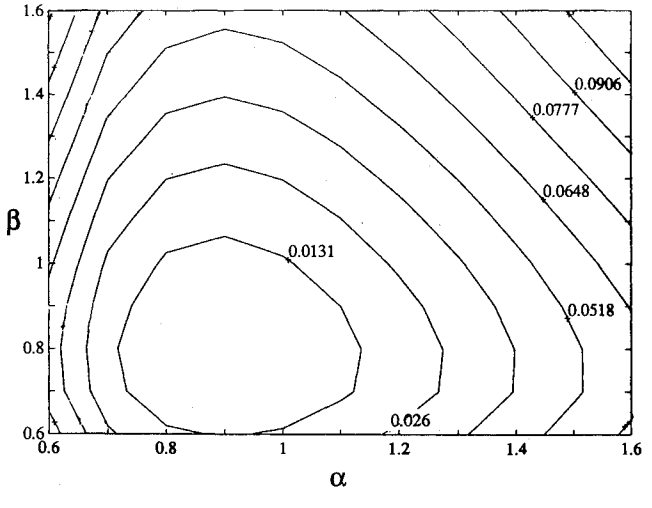

Fig. 4 Sum of squares contours for case 3.

imation quite good. Figures 3 and 4 show the sum of squares contours [using Eq. (1)] for cases 2 and 3 over the parameter space. The sum of squares were also elliptical for case 1; however, it was very elongated since the sensitivity of β was smaller than that for α. It is evident that the nonlinearity is small since the contours are nearly circular. This indicates that the linear approximations using the joint probabilities will give reasonable estimates for the confidence intervals in this problem.

An analysis of the residuals, the difference between predicted and fitted values, is commonly used to see qualitatively how the model and data met the assumptions of the measurement errors having a zero mean, being independent, and having constant variance. Checking the residuals, values were scattered randomly around a zero mean without any bias, thereby verifying all of these assumptions.

V. Conclusions

This paper demonstrates that the nonlinear optimization approach to parameter estimation is a flexible and effective method. Although computationally intensive, for large problems, this method lends itself to a wide variety of analytical model formulations and can provide an assessment of the uncertainty of the mean parameter estimates. Other factors, such as measurement error distributions, instrumentation reliability, and problem constraints, can be incorporated into the estimation procedure.

An analysis of a truss structure was performed to evaluate the estimation scheme. The truss structure was easily modeled via a finite element model using MSC/NASTRAN. Parameter estimation was used to identify two crucial parameters, yielding a model that gave excellent results. Since the model was linear and accurate, and the parameter interaction was small, the linearization schemes used to approximate the confidence intervals worked well.

Acknowledgments

This work performed at Sandia National Laboratories was supported by the U.S. Department of Energy under Contract DE-AC04-76DP00789. The authors would like to acknowledge the efforts of J. P. Lauffer of Sandia National Laboratories for obtaining the modal test data used in this study.

References

[1]Allen, J. J., and Martinez, D. R., "Techniques for Implementing Structural Model Identification Using Test Data," Sandia National Lab. Rept. SAND90-1185, Albuquerque, NM, 1990.

[2]Biegler, L. T., Damiano, J. J., and Blau, G. E., "Nonlinear Parameter Estimation: A Case Study Comparison," *American Institute of Chemical Engineering Journal,* Vol. 32, No. 1, 1986, pp. 29–45.

[3]Himmelblau, D. M., *Process Analysis by Statistical Methods,* Wiley, New York, 1970.

[4]Witkowski, W. R., "Model Identification and Parameter Estimation of Crystallization Processes," Ph.D. Dissertation, Univ. of Texas at Austin, Austin, TX, 1990.

[5]Bates, D. M., and Watts, D. W., "Relative Curvature Measures of Nonlinearity," *Journal of the Royal Statistical Society B,* Vol. 42, No. 1, 1980, pp. 1-25.

[6]Bates, D. M., and Watts, D. W., *Nonlinear Regression Analysis & Its Applications,* Wiley, New York, 1988.

[7]Hasselman, T. K., and Chrostowski, J. D., "Parameter Estimation for Large Space Structures," American Society of Civil Engineers, Spring Conf., CO, 1985.

[8]Bronowicki, A. J., Lukich, M. S., and Kuritz, S. P., "Application of Physical Parameter Identification to Finite Element Models," First NASA/DOD CSI Technology Conf., VA, 1986.

[9]Ojalvo, I. U., Ting, T., and Pilon, D., "PAREDYM—A Parameter Refinement Computer Code for Structural Dynamic Models," *International Journal of Analytical and Experimental Model Analysis,* Vol. 5, No. 1, 1989, pp. 43–49.

[10]Gill, P. E., Murray, W., and Wright, M. H., *Practical Optimization,* Academic Press, London, 1981.

[11]Donaldson, J. R., and Schnabel, R. B., "Computational Experience With Confidence Regions and Confidence Intervals for Nonlinear Least Squares," *Technometrics,* Vol. 29, No. 1, 1987, pp. 67-82.

[12]Bard, Y., *Nonlinear Parameter Estimation,* Academic Press, New York, 1974.

[13]Ratkowsky, D. A., *Nonlinear Regression Modeling,* Marcel Dekker, New York, 1983.

[14]Strang, G., *Linear Algebra and Its Applications,* Harcourt Brace Jovanovich, San Diego, CA, 1988.