Journal of Sound and Vibration (2000) **231**(4), 1071–1078
doi:10.1006/jsvi.1999.2721, available online at http://www.idealibrary.com on IDEAL®

ON A STATISTICAL OPTIMIZATION METHOD USED IN FINITE ELEMENT MODEL UPDATING

HUA HONGXING

Shanghai Jiao Tong University, Vibration Shock and Noise Lab, 200030 *Shanghai, People's Republic of China*

AND

H. SOL AND DE W. P. WILDE

Free University of Brussels, Brussels 1050, *Belgium*

(*Received* 21 *July* 1997, *and in final form* 1 *September* 1999)

Bayesian estimator is a commonly used statistical optimization technique for finite element model updating. This paper presents a modified Bayesian estimator and discusses its unbiasedness, efficiency, learning ability and robustness. The main differences from other estimators, for instance, the least-squares method, are shown. The new Bayesian estimator can also be used as a multi-objective, multi-design variable optimization method. An example is presented to demonstrate its features.

1. INTRODUCTION

Finite element models, in general are of uncertainties and model errors. This is especially true in the case where complex structures are modelled. If the experimental results of the real physical structure are available, it is recommended to update the finite element model with experimental data. The result of updating by means of experimental data is a finite element model that is more reliable for further predictions. Modal analysis data or measured response functions are very well suited to this purpose. It is, however, impossible and unnecessary that the numerical results should be identical with the experimental results, because (1) experimental results are noise corrupted; (2) finite element models are always an approximation of the physical reality; (3) damping is often neglected in calculation, etc.

The principle of finite element model updating is to compare the computed results with the measured results from the real structure. A number of parameters in the finite element model are selected and tuned in such a way that the computed output matches the measured output. The correlation between the experimental output and the numerical output is obtained by minimization of a cost function which contains the differences between both outputs. From a mathematical point of

0022-460X/00/141071 + 08 $35.00/0

view, the difficulty with updating is that the relation between the output column $\{y\}$ (dimension $n \times 1$) and the parameter column $\{x\}$ (dimension $m \times 1$) is nearly always non-linear. This means that updating the parameter values from an initial value to a final estimate has to be done in an iterative way. The value of the computed output column $\{y\}$ for new parameter values can be evaluated with a Taylor expansion at the initial values $\{y_0\}$ and $\{x_0\}$. The Taylor expansion is usually cut-off after the linear term:

$$\{y\} = \{y_0\} + [S](\{x\} - \{x_0\}). \tag{1a}$$

The $n \times m$ matrix $[S]$ that appears in the linear term is called the sensitivity matrix. This matrix contains the partial derivatives of the output components for the different parameters. For sensitivity analysis-based model updating, if the $n \times m$ sensitivity matrix is rank full, there are two cases which should be dealt with differently: in the case of $m \leqslant n$, the reasonable objective is that the parameters to be identified should converge to their true values in the probabilistic sense: if $n < m$, there are numerous solutions, among which there is an optimal solution in the least-squares sense. The success of the updating is highly dependent on the numerical condition of the sensitivity matrix. The numerical stability and hence the robustness of the updating can be improved by applying the Bayesian estimator algorithm presented in this paper.

Reference [1] might be the first paper on statistical model updating method. Reference [2] also discussed statistical model updating procedure. This paper first derives the statistical optimization algorithm in a different way. Then the estimation unbiasedness and robustness of the algorithm are discussed. Finally, the estimation efficiency is illustrated with an example.

2. MATHEMATICAL MODEL

Model updating implies an existing model. Suppose the correct parameter values are in the vicinity of $\{x_0\}$; the truncated Tailor expansion of $\{y\}$ is written again as

$$\{y\} = \{y_0\} + [S](\{x\} - \{x_0\}), \tag{1b}$$

where $\{x\}$ is the updated parameter column of dimension $m \times 1$, $\{x_0\}$ is its initial parameter column, $\{y\}$ is the $n \times 1$ column which consists of either eigenvalues, or eigenvectors, or frequency response functions, etc., the $n \times 1$ column $\{y_0\}$ is computed using $\{x_0\}$; $[S] = [\partial y/\partial x]$ is the sensitivity matrix of dimension $n \times m$. Set

$$\{\Delta x\} = \{x\} - \{x_0\}, \tag{2}$$

$$\{\Delta y\} = \{y_e\} - \{y\}, \tag{3}$$

where the $n \times 1$ column $\{y_e\}$ is composed of measured eigenvalues, or eigenvectors, or frequency response functions. It can be found that, if $\{x\}$ approaches correct

values, $\{y\}$ will approach exact values and $\{\Delta y\}$ will approach $\{\varepsilon\}$, the measurement noise. That is,

$$\{y_e\} = \{y\} + \{\varepsilon\}. \tag{4}$$

It is assumed that $\{\Delta x\}$ and $\{\varepsilon\}$ are normally distributed with zero means and are uncorrelated with each other.

$$E[\{\varepsilon\}\{\varepsilon\}^{\mathrm{T}}] = [V_y], \qquad E[\{\Delta x\}\{\Delta x\}^{\mathrm{T}}] = [V_x], \tag{5a}$$

$$E\{\Delta x\} = E(\Delta y\} = 0, \qquad E[\{\varepsilon\}\{\Delta x\}^{\mathrm{T}}] = 0, \tag{5b}$$

where $E[\ ]$ indicates the expectation with respect to $[\ ]$; $n \times n$ matrix $[V_y]$ and $m \times m$ matrix $[V_x]$ are the positive-definite and symmetrical covariance matrices of $\{\varepsilon\}$ and $[\Delta x]$ respectively. The joint probability density function is [3].

$$p(\Delta x, \varepsilon) = p(x)\,p(y|x) = \frac{1}{(2\pi)^{(n+m)/2}\sqrt{V_x V_y}} \exp -\frac{1}{2}(\Delta x^{\mathrm{T}} V^{-1}_x \Delta x + \varepsilon^{\mathrm{T}} V_y^{-1} \varepsilon). \tag{5c}$$

The matrix and vector signs are omitted in the above expression. The unconditional maximum likelihood estimation of $\{x\}$ is to choose that value of $\{x\}$ which maximizes the joint probability density function evaluated at any particular observation of $\{y\}$ [4]. Given the covariance matrices ($[V_y]$ is determined by experiment and $[V_x]$ is more or less subjectively guessed at the start), maximization of the joint probability density function is equal to minimization of the following cost function:

$$\min \Delta x^{\mathrm{T}} V^{-1}_x \Delta x + \varepsilon^{\mathrm{T}} V^{-1}_y \varepsilon. \tag{6}$$

Substitution of equations (1)–(4) into equation (6) leads to the following estimator:

$$\{\hat{x}\} = \{x_0\} + [K](\{y_e\} - \{y_0\}), \tag{7a}$$

$$[K] = ([S]^{\mathrm{T}}[V_y^{-1}][S] + [V_x^{-1}])^{-1}[S]^{\mathrm{T}}[V_y^{-1}]. \tag{7b}$$

Using the following matrix formula [5]:

$$[A]^1_{11}[A]_{12}([A]_{22} + [A]_{21}[A]^{-1}_{11}[A]_{12})^{-1}$$

$$= ([A]_{11} + [A]_{12}[A]^{-1}_{22}[A]_{21})^{-1}[A]_{12}[A]^{-1}_{22},$$

in which $[A_{11}]$ is an $(m \times m)$ matrix, $[A_{12}]$ an $(m \times n)$ matrix. $[A_{21}]$ an $(n \times m)$ matrix, and $[A_{22}]$ an $(n \times n)$ matrix, equation (7b) can also be written as

$$[K] = [V_x][S]^{\mathrm{T}}([S]\,[V_x][S]^{\mathrm{T}} + [V_y])^{-1}. \tag{7c}$$

3. THE ESTIMATION ERROR COVARIANCE MATRIX

The estimation error covariance can be derived as follows:

$$\begin{aligned}[V_x^*] &= E[\{\hat{x} - x\}\{\hat{x} - x\}^{\mathrm{T}}] \\ &= E[\{[V_x][S]^{\mathrm{T}}([S][V_x][S]^{\mathrm{T}} + [V_y])^{-1}(y_e - y_0) - \{\Delta x\}\} \\ &\quad \times \{[V_x][S]^{\mathrm{T}}([S][V_x][S]^{\mathrm{T}} + [V_y])^{-1}(y_e - y_0) - \{\Delta x\}\}^{\mathrm{T}}], \\ &= E[\{[V_x][S]^{\mathrm{T}}([S][V_x][S]^{\mathrm{T}} + [V_y])^{-1}([S]\{\Delta x\} + \{\varepsilon\}) - \{\Delta x\}\} \\ &\quad \times \{[V_x][S]^{\mathrm{T}}([S][V_x][S]^{\mathrm{T}} + [V_y])^{-1}([S]\{\Delta x\} + \{\varepsilon\}) - \{\Delta x\}\}^{\mathrm{T}}].\end{aligned}$$

By noting that

$$E[([S]\{\Delta x\} + \{\varepsilon\})([S]\{\Delta x\} + \{\varepsilon\})^{\mathrm{T}}] = ([S][V_x][S]^{\mathrm{T}} + [V_y]), \tag{8}$$

the following equation is obtained:

$$[V_x^*] = [V_x] - [V_x][S]^{\mathrm{T}}([S][V_x][S]^{\mathrm{T}} + [V_y])^{-1}[S][V_x]$$

or

$$[V_x^*] = ([I] - [K][S])[V_x], \tag{9}$$

where $[I]$ is the identity matrix.

4. UNBIASEDNESS, EFFICIENCY AND ROBUSTNESS

4.1. UNBIASEDNESS

$$\begin{aligned}E[\{\Delta\hat{x}\}] &= E[\{\hat{x}\} - \{x_0\}] \\ &= E[([S]^{\mathrm{T}}[V_y^{-1}][S] + [V_x^{-1}])^{-1}[S]^{\mathrm{T}}[V_y^{-1}](\{y_e\} - \{y_0\})] \\ &= E[([S]^{\mathrm{T}}[V_y^{-1}][S] + [V_x^{-1}])^{-1}[S]^{\mathrm{T}}[V_y^{-1}]([S]\{\Delta x\} + \{\varepsilon\})] \\ &= ([S]^{\mathrm{T}}[V_y^{-1}][S] + [V_x^{-1}])^{-1}[S]^{\mathrm{T}}[V_y^{-1}][S]E[\{\Delta x\}].\end{aligned} \tag{10}$$

It can be found that if $\{\Delta x\}$ is normally distributed with zero mean, the estimation will not be biased.

4.2. EFFICIENCY

If the nth estimation is more efficient than the $(n-1)$th estimation, the nth estimation error covariance should be smaller than the $(n-1)$th estimation error covariance. From equation (9), it can be seen that

$$[V_x^*][V_x]^{-1} = ([I] - [K][S]). \tag{11}$$

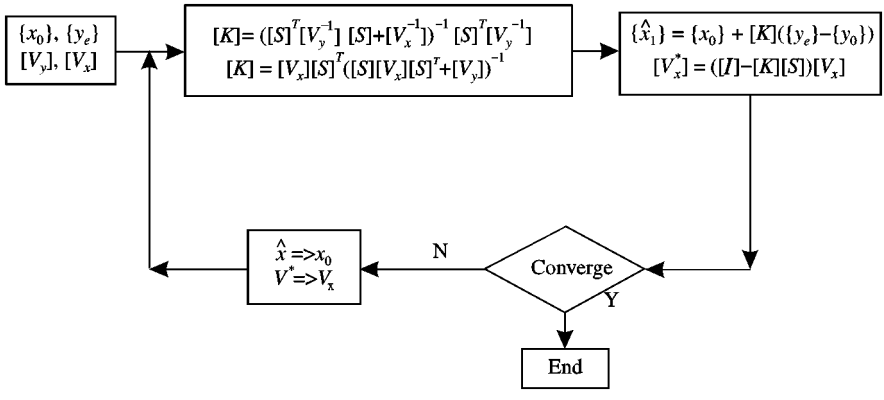

Figure 1. Estimation algorithm.

If the norm $\|[I] - [K][S]\| < 1$, the estimation is thus efficient. With equation (7b)

$$[K][S] = ([S]^{\mathrm{T}}[V_y^{-1}][S] + [V_x^{-1}])^{-1}[S]^{\mathrm{T}}[V_y^{-1}][S]. \tag{12}$$

It can be seen from equation (12) that the norm $\|[K][S]\|$ will be < 1 and thus the estimator will be efficient if the sensitivity matrix is rank full (since the covariance matrix $[V_x]$ is positive definite). Due to the inherent non-linearity of the problem and the initial value dependency of the convergence rate, satisfactory results will only be obtained after several iterations. Figure 1 shows the estimation algorithm.

4.3. LEARNING ABILITY

As shown in Figure 1, at the start, $\{x_0\}$ and $[V_x]$ are more or less subjectively guessed values. $\{y_0\}$ and the sensitivity matrix $[S]$ can be calculated using $\{x_0\}$. As $[V_y^-]$ is known, the gain matrix $[K]$ can be calculated next. Then the estimator learns from $\{\Delta y\}$ the difference between calculated results and measured ones, and modifies the previous judgement. The learning ability is characterized by the gain matrix $[K]$, which places different weightings to each element of $\{\Delta y\}$ to get a synthesized modification for $\{x_0\}$ which can be shown by equation (7a):

$$\hat{x}_j = x_{0,j} + \sum_{i=1}^{n} k_{ji}(y_{ei} - y_{i0}), \quad j = 1, 2, \ldots, m. \tag{13}$$

In general, the more sensitive the parameter, the higher the weighting. But, the covariance matrices $[V_x]$ and $[V_y]$ are also of great importance: from equations (7b) and (7c), it can be found that the smaller the covariance matrix $[V_x]$, the smaller the gain matrix $[K]$ and the smaller the parameter modification; the smaller the covariance matrix $[V_y]$, the larger the gain matrix $[K]$ and the larger

the parameter modification. One of the ways to avoid overmodification of some more sensitive parameters, is to set much smaller variances to these parameters.

4.4. ROBUSTNESS

Setting covariance $[V_y]$ to zero and $[V_x]$ to identify matrix, we have

$$\{\Delta x\} = [S]^{\mathrm{T}}([S][S]^{\mathrm{T}})^{-1}\{\Delta y\}. \tag{14}$$

Zero $[V_y]$ implies no measurement noise. Therefore, in equation (14), $\{y_e\}$ is supposed to be exact, but is in fact noise corrupted. As we know from the learning ability of the estimator, the smaller the covariance matrix $[V_y]$, the larger the parameter modification. Hence, the parameters will be overmodified to reach the unreachable "exact" objective, which probably results in divergence.

Setting $[V_y] = [I]$ and $[V_x] = \infty$,

$$\{\Delta x\} = ([S]^{\mathrm{T}}[S])^{-1}[S]^{\mathrm{T}}\{\Delta y\}. \tag{15}$$

The infinite $[V_x]$ implies that no *a priori* knowledge is known about the parameters, or the differences between initial parameters and true parameters are supposed to be very large. As stated before, the smaller the covariance matrix $[V_x]$, the smaller the parameter modification. That adaptive nature is what we need when the parameters approach their true values. However, equation (15) assumes infinite $[V_x]$, no matter how close the parameters are to their true values, which probably results in divergence too.

Therefore, the least-squares solution, equation (14) or (15) is not so robust as the Bayesian estimator. There are two main reasons. First, non-zero $[V_y]$ is correct in physical reality. That means some calculation error should be allowable. Secondly, the Bayesian estimator will apply more and more strict "constraints", smaller $[V_x]$, to those parameters that approach their true values, so that overmodification can be avoided.

It is suggested that in the Bayesian estimator, the value of $[V_y]$ should be close to measurement noise, the value of $[V_x]$ should be close to the actual situation, or close to the required parameter variance, and this covariance can be used in the entire iteration procedure. Of course, we can still monitor the difference of covariance between two iterations, and use this difference as convergence criterion.

It should be pointed out that references [6, 7] proposed a method to improve the ill-conditioning of the sensitivity matrix:

$$\{\Delta x\} = [S]^{\mathrm{T}}([S][S]^{\mathrm{T}} + [\varepsilon])^{-1}\{\Delta y\}. \tag{16}$$

Equation (16) is obviously a special case of equation (7c) by setting $[V_y]$ to $[\varepsilon]$ and $[V_x]$ to identify matrix.

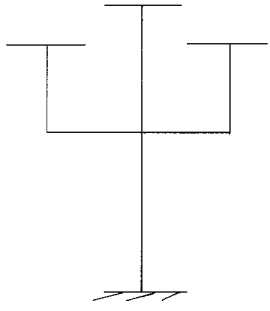

Figure 2. Finite element model of a satellite antenna.

TABLE 1

Resonance frequencies of initial finite element model

	Finite element (Hz)	Experiment (Hz)	Diff %	Mac %
1	18·38	18·05	1·82	96·3
2	20·50	21·85	− 6·18	96·0
3	33·95	32·66	3·91	60·2
4	36·97	37·71	4·96	46·9
5	39·028	40·01	− 2·47	85·8
6	44·41	43·74	4·54	49·9
7	44·55	49·72	− 10·40	84·0
8	47·88	45·09	6·18	25·8
9	51·00	55·06	− 7·38	96·4
10	68·41	72·51	− 5·65	96·8
11	77·44	80·31	− 3·58	97·7
12	90·20	68·99	31·31	99·9

5. A CASE STUDY

The finite element model of a satellite antenna is shown in Figure 2. There are totally 38 beam elements. It was required to correlate first 12 computed resonance frequencies and mode shapes with the experimental results. The problem is a multi-objective and multi-parameter optimization problem. 12 resonance frequencies are taken as objectives and 38 second moment of inertia I_z for each beam element as parameters.

Table 1 lists the calculated resonance frequencies of the initial finite element model, the experimental data and MAC values (MODAL ASSURANCE CRITERION, MAC) $\{\phi_e\}^{\mathrm{T}}\{\phi_a\}/(\{\phi_e\}^{\mathrm{T}}\{\phi_e\})\ \{\phi_a\}^{\mathrm{T}}\{\phi_a\}, \{\phi_a\}$ is computed modes and $\{\phi_e\}$ is measured modes, [8]). Table 2 lists the updated results after five iterations.

6. CONCLUSION

This paper derives a Bayes type of multi-objective and multi-design parameter optimization method by maximizing a joint probability density function, and

TABLE 2

Resonance frequencies of updated finite element model

	Finite element (Hz)	Experiment (Hz)	Diff %	Mac %
1	18·03	18·05	− 0·12	94·9
2	21·81	21·85	− 0·20	95·8
3	32·75	32·66	0·26	89·8
4	37·55	37·71	− 0·42	64·2
5	40·01	40·01	0·01	98·9
6	43·47	43·74	− 0·62	94·9
7	45·37	45·09	0·62	92·4
8	49·69	49·72	− 0·05	99·3
9	55·05	55·06	− 0·01	97·2
10	68·69	68·69	− 0·01	99·8
11	72·73	72·51	0·30	74·7
12	80·32	80·31	0·02	97·8

discusses its estimation efficiency, unbiasedness, learning ability and robustness. A case study shows that it is an efficient optimization and estimation method in engineering.

REFERENCES

1. J. COLLINS, G. HART, T. HASSELMAN and B. KENNEDY 1974 *AIAA Journal* **12**. Statistical identification of Structures.
2. T. HASSELMAN 1983 *Winter Annual Meeting of ASME, Boston, MA*, 119–133. ASME Bound Vol Model testing and model refinement (ASME, New York) AMO-Vol. 59. A perspective on dynamic verification.
3. H. HUA 1993 *Doctoral Dissertation, VUB, Brussels*, 123–176. Identification of plate rigidities of anisotropic rectangular plates, sandwich panel and circular orthotropic discs using vibration data.
4. N. E. NAHI 1976 *Estimation Theory and Applications*, 175–190, 272–280. Huntington, New York. Robert E. K. Publishing Co.
5. D. E. CATIN 1989 *Estimation, Control, and Discrete Kalman Filter*, 133–160. Springer-Verlag, Berlin.
6. I. U. OJAVO 1990 31*th AIAA SDM Conference*, 1–6. Improved solution for system identification equation by epsilon-decomposition.
7. D. W. MARQUADT 1963 *Journal of the Society for Industrial and Applied Mathematics* **11**, 431–411.
8. R. J. ALLEMANG and D. L. BROWN 1982 *First IMAC, Florida, U.S.A.* 110–116. A correlation coefficient for modal vector analysis.