P. C. Shah
Research Associate.

F. E. Udwadia
Associate Professor.

Department of Civil Engineering,
University of Southern California,
Los Angeles, Calif.

A Methodology for Optimal Sensor Locations for Identification of Dynamic Systems

The problem of optimally positioning sensors in lumped and distributed parameter dynamic systems for the purpose of system identification from time-domain input-output data is formulated and a methodology for its solution is presented. A linear relation between small perturbations in a finite-dimensional representation of the system parameters and a finite sample of observations of the system time response is used to determine approximately the covariance of the parameter estimates. The locations of a given number of sensors are then determined such that a suitable norm of the covariance matrix is minimized. The methodology is applied to the problem of optimally locating a single sensor in a building structure modeled by a shear beam, such that the estimates of the stiffness distributions, obtained from the records of strong ground shaking and the building response at the sensor location, are least uncertain.

Introduction

The problem of the determination of the dynamic parameters of systems from "input-output" data has become of considerable interest to researchers in many areas of science and engineering. Often, the systems are extended in one or more spatial dimensions, and spatial distributions of the dynamic parameters have to be determined. Some commonly encountered examples of identification of such distributed systems are: determination of stiffness and damping distributions from displacement records in building structures, bridges and dams, determination of rock permeability and porosity from pressure data in petroleum and geothermal reservoirs, and determination of elastic properties of underground rock from seismic data. Usually the input-output data are recorded in the time domain.

There are usually several alternative locations where different sensors can be located. On the other hand, due to economic and space restrictions, the number of sensors is limited. Under these circumstances, the following question arises: for a given number of sensors, where should they be located in the system domain, so that the dynamic parameter estimates resulting from identification using the data obtained thereat have the smallest uncertainty?

This paper is concerned with this problem of optimal sensor location for identification using time domain data. A methodology for its solution is developed. It is assumed that identification has been carried out by obtaining a close match between the observed time response of the system and the simulated model response for a given excitation. It is further assumed that the estimate of the dynamic parameters so determined are in the vicinity of their true values. Then it is possible to obtain an approximate linear relation between small variations in the dynamic parameters and the simulated response.

By representing the system parameters in terms of a finite number of unknowns using a suitable parametrization if necessary, and by suitably sampling the response at a finite number of time instants, this linear relation is expressed as a linear matrix equation. Using this equation, an expression for the covariance matrix representing the uncertainty on the estimates is obtained. The possibility of having *a priori* probabilistic information about the unknowns is pointed out and it is indicated how such information can be incorporated in determination of the covariance matrix. The minimization of a positive definite scalar measure of the covariance matrix is proposed as an optimality criterion for sensor locations.

The proposed methodology is illustrated using the problem of determining the optimal location of a single sensor in a building structure represented as a discrete shear beam, the shear stiffness at different floor levels being the parameters to be identified (Shah and Udwadia [5][1]).

Theory

Let the L-vector $\mathbf{w}(t)$ represent the state of the dynamic system. It may represent the state of either a lumped parameter system or a finite dimensional representation of a distributed parameter system

Contributed by the Applied Mechanics Division for publication in the JOURNAL OF APPLIED MECHANICS.

Discussion on this paper should be addressed to the Editorial Department, ASME, United Engineering Center, 345 East 47th Street, New York, N. Y. 10017, and will be accepted until June 1, 1978. Readers who need more time to prepare a Discussion should request an extension of the deadline from the Editorial Department. Manuscript received by ASME Applied Mechanics Division, May, 1977; final revision, August 1977.

[1] Numbers in brackets designate References at end of paper.

obtained through a suitable discretization. Let the state equation be,

$$\dot{w} = f(w, k, \bar{v}(t)) \qquad w(0) = \text{given} \tag{1}$$

where k is an N-vector of the dynamic parameters of the systems and $\bar{v}(t)$ is an n_1-vector of inputs driving the system.

Identification Problem. Let there be l sensors located in the system domain, which provide noisy measurements of some functions of the state of the system.

$$y_i^{obs}(t) = h_i(w(t); s_i) + \epsilon_i(t) \quad t \in (0, T) \qquad i = 1, 2, \ldots, l \tag{2}$$

where s_i are the sensor locations and $\epsilon_i(t)$ are zero mean observation noise processes. Then, identification consists of determining the estimate $\hat{k}$ of k using these observations.

Usually, this is achieved by determining the estimates $\hat{k}$ which when used in the model equations (1) yield calculated values $w^{cal}(t)$ of the state variables such that the corresponding values

$$y_i^{cal}(t) = h_i(w^{cal}(t); s_i), \quad i = 1, 2, \ldots, l \tag{3}$$

are as close as possible to the observed values $\{y_i^{obs}(t)\}$ for $t \in (0, T)$.

This is usually achieved by solving the following minimization problem:

$$\min_k J = \frac{1}{2} \sum_{i=1}^{l} \int_0^T (y_i^{obs}(\tau) - y_i^{cal}(\tau))^2 W_i(\tau) d\tau \tag{4}$$

where $W_i(\tau)$ are weighting functions which are related to the variances of $\epsilon_i(t)$.

Optimal Sensor Location Problem. The optimal sensor location problem consists of determining the l sensor locations $\{s_i\}$ such that the resulting parameter estimates $\hat{k}(s_1, s_2, \ldots, s_l)$ are as close as possible to the true parameter values k. Since the estimates of $\hat{k}$, depending on the random processes $\{\epsilon_i(t)\}$ are random, the closeness of $\hat{k}$ and k must be considered in a statistical sense. Thus mathematically, the optimal sensor locations are obtained by solving the minimization problem:

$$\min_{\{s_1, s_2 \ldots s_l\}} J_1 = ||E[\Delta k \Delta k^T]|| \tag{5}$$

where

$$\Delta k = \hat{k}(s_1, s_2, \ldots, s_l) - k, \tag{6}$$

$||\cdot||$ denotes a suitable norm of the matrix function, and $E[\cdot]$ denotes expectation with respect to the probability distributions of the observation noise $\{\epsilon_i(\tau)\}$.

Solution to the Optimal Sensor Location Problem. Let us define an l-vector,

$$y^T(t) = [h_1(w(t); s_1), h_2(w(t); s_2), \ldots, h_l(w(t); s_l)] \tag{7}$$

and a similar l-vector $y^{cal}(t)$. Then the l-vector of observations can be written as

$$y^{obs}(t) = y(t) + \epsilon(t) \tag{8}$$

where $\epsilon^T(t) = [\epsilon_1(t), \epsilon_2(t), \ldots, \epsilon_l(t)]$. Let us divide the time interval $[0, T]$ into equal intervals using M points $t_1, t_2, t_3, \ldots, t_M$. We shall determine the estimate error considering observations sampled at these instants.

For the given input $\bar{v}(t)$, we have $y(t) = y(t; k)$ and $y^{cal}(t) = y^{cal}(t; \hat{k})$. If the estimate vector $\hat{k}$ is close to the true vector k, then $y = y(k)$ can be expanded in terms of the vector $y^{cal} = y^{cal}(\hat{k})$. At any time $t = t_j$ we have

$$y(t_j) = y^{cal}(t_j) + \left.\frac{\partial y^{cal}(t_j)}{\partial k}\right|_{k=\hat{k}} (k - \hat{k}) + 0||k - \hat{k}||^2 \tag{9}$$

where $\partial y^{cal}(t_j)/\partial k$ is an $l \times N$ sensitivity matrix. Denoting the history mismatch error at time t_j as $m(t_j) = y^{obs}(t_j) - y^{cal}(t_j)$ we obtain from (8)

$$m(t_j) = y(t_j) + \epsilon(t_j) - y^{cal}(t_j). \tag{10}$$

Using relation (9) and denoting the matrix $\partial y(t_j)/\partial \hat{k}$ by B_j we have, to first order,

$$m(t_j) = \epsilon(t_j) + B_j(k - \hat{k}) \tag{11}$$

Equation (11) is valid at each of the discrete times t_j, $j = 1, 2, \ldots, M$.

Denoting $\theta^T = [m^T(t_1), m^T(t_2), \ldots, m^T(t_M)]$, $\eta^T = [\epsilon^T(t_1), \epsilon^T(t_2), \ldots \epsilon^T(t_M)]$, and $A^T = [B_1^T | B_2^T | \ldots | B_M^T]$ we have

$$\theta = \eta + A(k - \hat{k}) \tag{12}$$

where θ and η are column vectors both of dimension lM and A is an $lM \times N$ matrix. Let us define $\xi = \theta - \eta$, so that

$$A(k - \hat{k}) = \xi \tag{13}$$

Since both θ and η are random vectors, so is ξ. Then (13) implies that $k - \hat{k}$ is a random vector with its statistics related to those of ξ.

Also, A is a random matrix since it depends on the random vector $\hat{k}$. However, if k is sufficiently close to $\hat{k}$ the matrix A can be expanded as,

$$A = \bar{A} + \frac{\partial \bar{A}}{\partial k}(k - \hat{k}) + 0||k - \hat{k}||^2 \tag{14}$$

where $\bar{A}$ is the sensitivity matrix evaluated at the "true" value k.

Using (13) and (14) we obtain,

$$\bar{A}(k - \hat{k}) + 0||k - \hat{k}||^2 = \xi \tag{15}$$

The solution of (15) is, to the first order,

$$k - \hat{k} = \bar{A}^{\psi}\xi \tag{16}$$

where $\bar{A}^{\psi}$ is a suitable inverse of $\bar{A}$.

The statistics of the estimate error are

$$E[k - \hat{k}] = E[\bar{A}^{\psi}\xi] = \bar{A}^{\psi}E[\xi] \tag{17}$$

and

$$E[(k - \hat{k})(k - \hat{k})^T] = \bar{A}^{\psi}E[\xi\xi^T]\bar{A}^{\psi T} \tag{18}$$

From the last two expressions it is clear that the statistics of the estimate error are dependent on the inverse $\bar{A}^{\psi}$ used. Since $\bar{A}$ is a nonsquare matrix, it has no unique inverse. However, the Lanczos inverse (Lanczos [3]) has some desirable properties that make it suitable for use here.

The singular value decomposition of matrix $\bar{A}$ leads to the relation,

$$\bar{A} = U\Lambda V^T \tag{19}$$

where U is an $(lM \times N)$ matrix, which has an orthogonal set of lM-vectors $\{u_i, i = 1, 2, \ldots, N\}$ as its columns, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$, with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$, and V is an orthogonal $(N \times N)$ matrix with columns $\{v_i, i = 1, 2, \ldots, N\}$.

The Lanczos inverse is defined in terms of nonzero singular values of $\bar{A}$. Let $\lambda_1, \lambda_2, \ldots, \lambda_r > 0$ and $\lambda_{r+1} = \lambda_{r+2} = \ldots = \lambda_N = 0$, and let $\bar{V} = [v_1, v_2, \ldots, v_r]$, $\bar{U} = [u_1, u_2, \ldots, u_r]$, and $\bar{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$. Then the Lanczos inverse is

$$\bar{A}^{\psi} = \bar{V}\bar{\Lambda}^{-1}\bar{U}^T \tag{20}$$

It can be shown (Jackson [2]), that when $\lambda_i > 0$ for $i = 1, 2, \ldots, N$, Lanczos inverse is identical with the least-square solution of (15). When only $r < N$ of the λ's are nonzero, Lanczos inverse yields a least-square solution having smallest Eucledian norm since it eliminates the corresponding components along the insensitive directions $\{v_{r+1}, v_{r+2}, \ldots, v_N\}$. Let us denote by $\bar{V}_0$ the matrix with these vectors as its columns.

From equations (17), (18), and (20), we have

$$\delta\bar{k} = E[k - \hat{k}] = \bar{V}\bar{\Lambda}^{-1}\bar{U}^T E[\xi] \tag{21}$$

$$\bar{P}_k = E[(k - \hat{k})(k - \hat{k})^T] = \bar{V}\bar{\Lambda}^{-1}\bar{U}^T E[\xi\xi^T]\bar{U}\bar{\Lambda}^{-1}\bar{V}^T \tag{22}$$

If $r = N$, $\overline{U} = U$, $\overline{\Lambda} = \Lambda$ and $\overline{V} = V$, and (21) and (22) yield the statistics of the total estimate error. However, for $r < N$, these expressions give the statistics only for the components of $(\mathbf{k} - \dot{\mathbf{k}})$ which are sensitive to the observations; and there will be an additional error in the estimates given by the projection $\alpha_0 = \overline{V}_0{}^T(\mathbf{k} - \dot{\mathbf{k}})$ along the columns of $\overline{V}_0$:

$$\delta\mathbf{k}_0 = \overline{V}_0\alpha_0 = \overline{V}_0\overline{V}_0{}^T(\mathbf{k} - \dot{\mathbf{k}}) \tag{23}$$

The statistics of this component of the estimate error cannot be determined from the statistics of ξ, and additional information about the true parameter value $\mathbf{k}$ is needed for this purpose. If this information is available in the form of *a priori* probability distribution of $\mathbf{k}$, then the statistics of the component of the estimate error are

$$\overline{\delta\mathbf{k}_0} = E|\overline{V}_0\overline{V}_0{}^T(\mathbf{k} - \dot{\mathbf{k}})| \tag{24}$$

$$\overline{P}_{k_0} = E|\overline{\mathbf{v}}_0\overline{\mathbf{v}}_0{}^T(\mathbf{k} - \dot{\mathbf{k}})(\mathbf{k} - \dot{\mathbf{k}})^T\overline{\mathbf{v}}_0\overline{\mathbf{v}}_0{}^T| \tag{25}$$

where the expectations are taken with respect to the prior probability distribution of $\mathbf{k}$. Furthermore, if we assume that the random vectors ξ and $\mathbf{k}$ are independent, we obtain the statistics of the total estimate error $(\mathbf{k} - \dot{\mathbf{k}})_{\text{total}}$ as

$$E|\delta\mathbf{k}| = \overline{\delta\mathbf{k}} + \overline{\delta\mathbf{k}_0} \tag{26}$$

$$P_k = \overline{P}_k + \overline{P}_{k_0} + \overline{\delta\mathbf{k}}\,\overline{\delta\mathbf{k}_0}{}^T + \overline{\delta\mathbf{k}_0}\,\overline{\delta\mathbf{k}}{}^T \tag{27}$$

The optimal sensor locations can be obtained as those which minimize a suitable norm of P_k. From a practical standpoint, the quantities $\overline{A}$, $\overline{U}$, $\overline{\Lambda}$, and $\overline{V}$ are unknown since the true value $\mathbf{k}$, about which the linearization is done, is not known. Hence, in practice, the sensitivity matrix A computed by linearization about the estimate $\dot{\mathbf{k}}$ must be used. Then, the expressions (17) and (18) are no more rigorously valid since A, depending on $\dot{\mathbf{k}}$, is a random matrix depending on the probability distribution of ξ. However, as indicated in the illustrative example dealing with the identification of building structure stiffnesses, the matrix A, and more specifically, its singular values and the respective subspaces spanned by U, V, and V_0 do not change appreciably for even relatively large changes in the distribution $\dot{\mathbf{k}}$. The value of A and its properties are strongly determined by the structure of the system model and the sensor locations. These observations are also found to be true in the problem of identification of porous rock properties in petroleum reservoirs (Shah [4]). Thus, for a given model equation and a given set of sensor locations, replacing $\overline{A}$ by A in the foregoing analysis does not cause appreciable errors in the evaluation of P_k.

Some Comments on the Statistics of ξ. Since $\theta = \xi - \eta$, statistics of ξ can only be obtained if those of the mismatch θ between the observed and model response are available. However, the latter are not easily available and must either be assumed or inferred from the results of the identification computations.

The mismatch θ depends on three factors: (*i*) the errors made while modeling the system mathematically, (*ii*) the observation errors η, and (*iii*) the degree to which the objective function J is numerically minimized with respect to the parameter $\dot{\mathbf{k}}$. The mismatch due to the modeling errors and the observation error η cannot be in general entirely eliminated by making small changes in the parameter estimates $\dot{\mathbf{k}}$; the component of this mismatch along the directions orthogonal to the subspace spanned by columns of $\overline{U}$ will remain undiminished. If sufficient effort is spent in minimizing J, the first two of the aforementioned components of θ will be mainly contributory to the residual value of J.

Assuming that η is a zero-mean Gaussian random variable and that the modeling errors are small, we may take θ and consequently ξ, to be zero-mean Gaussian random variables. Denoting the covariance matrix of ξ to be P_ξ,

$$\overline{\delta\mathbf{k}} = 0$$

$$\overline{P}_k = \overline{V}\,\overline{\Lambda}^{-1}\overline{U}^T P_\xi \overline{U}\,\overline{\Lambda}^{-1}\overline{V}^T \tag{28}$$

When $P_\xi = \sigma^2 I$, the foregoing expression reduces to

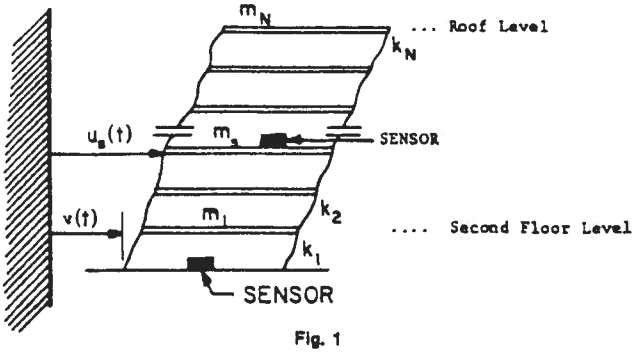

Fig. 1

$$\overline{P}_k = \sigma^2\overline{V}\,\overline{\Lambda}^{-2}\overline{V}^T \tag{29}$$

If, further, the norm of $\overline{P}_k$ is taken to be its trace, then

$$||\overline{P}_k|| = \sigma^2 \sum_{i=1}^{r} 1/\lambda_i{}^2 \tag{30}$$

Comments on the Choice of r. In practice, some of the singular values of $\overline{A}$ are often very close to zero. The use of such small singular values in expression (28) would lead to very large values of the covariance $\overline{P}_k$. On the other hand, if these singular values are taken to be zero, the dimension of $\overline{V}_0$ increases and the value of $\overline{P}_{k_0}$ increases. Thus, generally, there is a value of r for which $||P_k||$ is a minimum. The matrix $\overline{P}_{k_0}$ is inversely related to the commonly used resolution matrix (Jackson [2]). As more and more directions $\mathbf{v}_i$ in the parameter space are excluded from identification by decreasing r, the undetermined component of the parameter vector will become larger and larger and consequently the resolution will progressively worsen.

Calculation of A. The elements of the sensitivity matrix can be computed by two alternative methods:

1 Direct integration of the sensitivity equations.

2 Through the use of an adjoint variable (see the Appendix).

The sensitivity equations are derived by differentiating the system equation with respect to the parameters taken one at a time. The linear equations of sensitivity with respect to each parameter are, in general, coupled and are of the same order as the system equation. Thus, for N parameters, the integration of $(N + 1)$ sets of equations similar to the system equation needs to be done over the observation period. On the other hand, the adjoint variable method requires integration of l adjoint (linear) equations in addition to the system equation, all of which are of the same order. Further, calculation of the sensitivity of an observed state variable at any time t_i with respect to each parameter requires evaluation of a time integral over $(0, t_i)$. Thus, if the number of sensors is small compared to that of the parameters, and if the number of sampling instants M is not very large, the adjoint variable method is computationally more efficient (Shah [4]). The adjoint variable method for systems modeled by a set of linear ordinary differential equations is detailed in the Appendix; the derivation therein can be easily extended to nonlinear models, and to systems described by partial differential equations as well as discrete sets of algebraic equations (Shah [6]).

Illustrative Example

As an illustration of the methodology presented in the foregoing, we present below the results concerning optimal location of a sensor in a building structure for identification using response to earthquake ground motions.

Problem Statement. The building structure is approximated by a lumped system consisting of known masses $\{m_i, i = 1, 2, \ldots, N\}$ at the floor levels, connected by columns with unknown shear stiffnesses $\{k_i, i = 1, 2, \ldots, N\}$. The building is assumed to respond as a shear beam, and the soil-structure interaction is ignored (Fig. 1).

For simplicity of analysis, it is assumed that the records of displacement as functions of time are available: furthermore, the basement motion is assumed to be exactly known, whereas the displace-

ment record at the upper sensor location is assumed to be corrupted by observation noise of known statistics.

The upper sensor is to be located so that the estimates of the stiffnesses $|k_i|$, obtained by matching the model response and the records with a given degree of accuracy after starting with prior estimates sufficiently close to the true values to avoid nonuniqueness of the estimates (Udwadia [8]), have the smallest uncertainty associated with them. Specifically, the trace of the covariance matrix of the stiffness estimates is to be minimized with respect to the upper sensor location.

Let $u^T = (u_1(t), u_2(t), \ldots, u_N(t))$ be the displacement vector. Then the equation governing the motion is

$$M\ddot{u} + Fu = \iota v(t), \quad u(0) = \dot{u}(0) = 0 \tag{33}$$

where $v(t)$ is the motion of the base, $M = \text{diag}(m_1, m_2, \ldots, m_N)$, $\iota = |k_1, 0, \ldots, 0|^T$, $|F_{i,i} = k_i + k_{i+1}, F_{i,i+1} = -k_{i+1}, F_{i+1,i} = -k_{i+1}; i = 1, 2, \ldots, N-1|$, and $F_{N,N} = k_N$. The rest of the elements of F are zero.

Let the upper sensor be located at the sth floor level. Then, the observation record is

$$y^{obs}(t) = u_s(t) + \eta(t), \quad t \in (0, T) \tag{34}$$

where $\eta(t)$ is the observation noise, which is assumed to be zero-mean Gaussian white noise with a uniform variance of σ_1^2. No prior information about $|k_i|$ is assumed.

Solution Technique. The sensitivities of the observations at M time instants uniformly distributed over $(0, T)$ were computed using the adjoint variable method (see the Appendix). Both the system equation and the adjoint equation were integrated using the implicit Crank-Nicholson scheme. Singular value decomposition of $(M \times N)$ sensitivity matrix $A(a_{ij} = \partial y_i/\partial k_j)$ was carried out using the program by Bushinger and Golub [1].

The minimization of trace (P_k) = trace $(\bar{P}_k)$ with respect to s was carried out using exhaustive search. In cases where there are many sensor locations to be determined, exhaustive search is inefficient and a systematic minimization procedure such as a discrete analog of the gradient algorithm should be used.

Results and Discussion. Evidently, the optimal sensor location depends on the true stiffness distribution and on the excitation input $v(t)$. Furthermore, since the sensitivity matrix is calculated by linearization about an estimate of the unknown stiffnesses, the exact solution to the optimal sensor location problem may only be obtained after a good identification has been carried out. On the other hand, in reality, an optimal choice of sensor location must often be determined *prior* to any substantial strong motion data is available. Thus it is very important to determine the influence of changes in the prior stiffness estimates and the inputs on the optimal sensor location.

In view of this, the optimal sensor location was determined for two types of inputs and two stiffness distributions. The first 20 sec of the $N-S$ component of the motion recorded during the 1940 El Centro earthquake and a simulated white noise (see Udwadia and Shah [7]), were alternatively used as the inputs. The stiffness distributions analyzed were (1) uniform ($k_i = 1.5 \times 10^4$) and (2) linearly decreasing with height (from 2.0×10^4 to 1.0×10^4). In all cases, mass was considered to be uniformly distributed with height ($m_i = 3$). Furthermore, in order to study the dependence of the optimal sensor location on the building height, buildings with three different heights were analyzed, $N = 5, 10, 20$.

The integration was carried out using uniform time steps of $\Delta t = 0.05$ sec. To investigate the influence of observation sampling on the optimal sensor location, the calculations were repeated with six different values of M ($M = 13, 20, 24, 32, 45, 99$) for $N = 10$ and $T = 10$ seconds using a linearly varying stiffness distribution. In addition, to study the influence of the sampling instants, the analysis of two cases (with $M = 24$ and $M = 99$) was repeated after shifting all the sampling instants forward in time by half the sampling period. In each case, the conclusion about the optimal sensor location was identical. Hence, all the subsequent computations, the results of which are reported here, were carried out using a single value of $M = 66$ and a fixed sampling policy.

In each case, the total mismatch, ξ between y^{cal} and y^{true}, was assumed to be a Gaussian random vector with zero-mean and covariance $\sigma^2 I$, where the value of σ^2 was uniform for all cases and different observation locations. This assumption is justified on the following basis. The vector η is assumed to be Gaussian with a zero-mean and covariance $\sigma_1^2 I$. Furthermore, it is possible in general to obtain a match between the observations and the model response to such a degree that the residual mismatch is of the same order of magnitude as the observation error (residual value of $J \approx T/\Delta t \sigma_1^2$). For sufficiently large sampling intervals, different elements of the mismatch vector may be assumed zero-mean and statistically independent. We further assume them to be Gaussian. Thus each ξ_i may be assumed to be a difference of two zero-mean, Gaussian random variables with equal variance of σ_1^2. The actual value of σ does not influence the conclusion about optimal sensor location.

The results of the calculations are shown in Figs. 2 and 3. In these figures, the average standard deviation of the error in the individual parameter estimates (= [trace $(\bar{P}_k)/N]^{1/2}$, normalized with respect to the mean of the spatial distribution of the stiffness, is plotted against the sensor location, s. The optimal sensor location is that value of s at which the plot has a minimum. The figures show results for $N = 5$, 10, and 20, using the El Centro and white noise inputs and both stiffness distributions. The results are plotted for several values of r, including $r = N$. (In most cases with the uniform stiffness distribution, all but one point, with $s = 2$, fell beyond the upper range of the vertical scales of the plots when $r = N$; hence, only one point, which is the minimum, is plotted in these cases.)

For $r = N$, $P_k = \bar{P}_k$, and the top curve in each case is applicable when no prior information about the stiffness distribution is available. In this situation, $s = 2$ is the optimal location of the sensor, for all cases with the diverse conditions of input, stiffness distribution, and building heights considered. On the other hand, when $r < N$, $\bar{P}_k$ is only one contribution of the total covariance matrix. However, in the light of the following discussion, whan r is not much smaller than N, this can be seen to be the total covariance when the stiffness distribution is restricted in a certain manner. Then the plots with $r < N$ reinforce the previous conclusion concerning the optimal sensor location. For cases with $r \ll N$, the contribution of $\bar{P}_{k_0}$ to P_k must be considered in the analysis after assuming suitable *a priori* probabilistic information about the stiffness distribution.

Figs. 4 and 5 show the v-vectors plotted against the height for a structure with $N = 10$, for both inputs and both stiffness distributions. The results for two sensor locations, $s = 2$ and $s = 10$, are illustrated. In each plot, the corresponding singular values λ are indicated. It is evident from these plots that in all cases, as λ decreases, the v-vectors have more and more spatial oscillations. (This implies that the data are not sensitive to the highly oscillatory components of the stiffness distribution; and consequently such components are not well determined by the data.) Then, restricting $(k - \bar{k})$ to have no components along the columns of $\bar{V}_0$ is tantamount to the restriction that the stiffness distribution not be very oscillatory. Thus $\bar{P}_k$ is the total covariance of the estimate error under such a restriction on the stiffness distribution.

Another important observation from Figs. 4 and 5 is that the v-vectors associated with several of the larger singular values are almost identical for the diverse conditions and for sensors located as far apart as the second floor and the roof level. (Since these are eigenvectors, a change of sign is immaterial.) Furthermore, if we consider the subspace of E^N spanned by the first r of the v-vectors, then changes in the order of occurance of similar vectors in this group may be ignored. Then, the respective subspaces spanned by the columns of $\bar{V}$ and $\bar{V}_0$ for a given r are almost identical for all the cases illustrated. Thus, fixing a value of $r < N$ implies almost identical restriction on the stiffness distribution for all the cases. These observations about the v-vectors were found to be true for all the cases analyzed with different sensor locations and with $N = 5$, 10, and 20.

As noted earlier, the calculated covariance of the estimate error

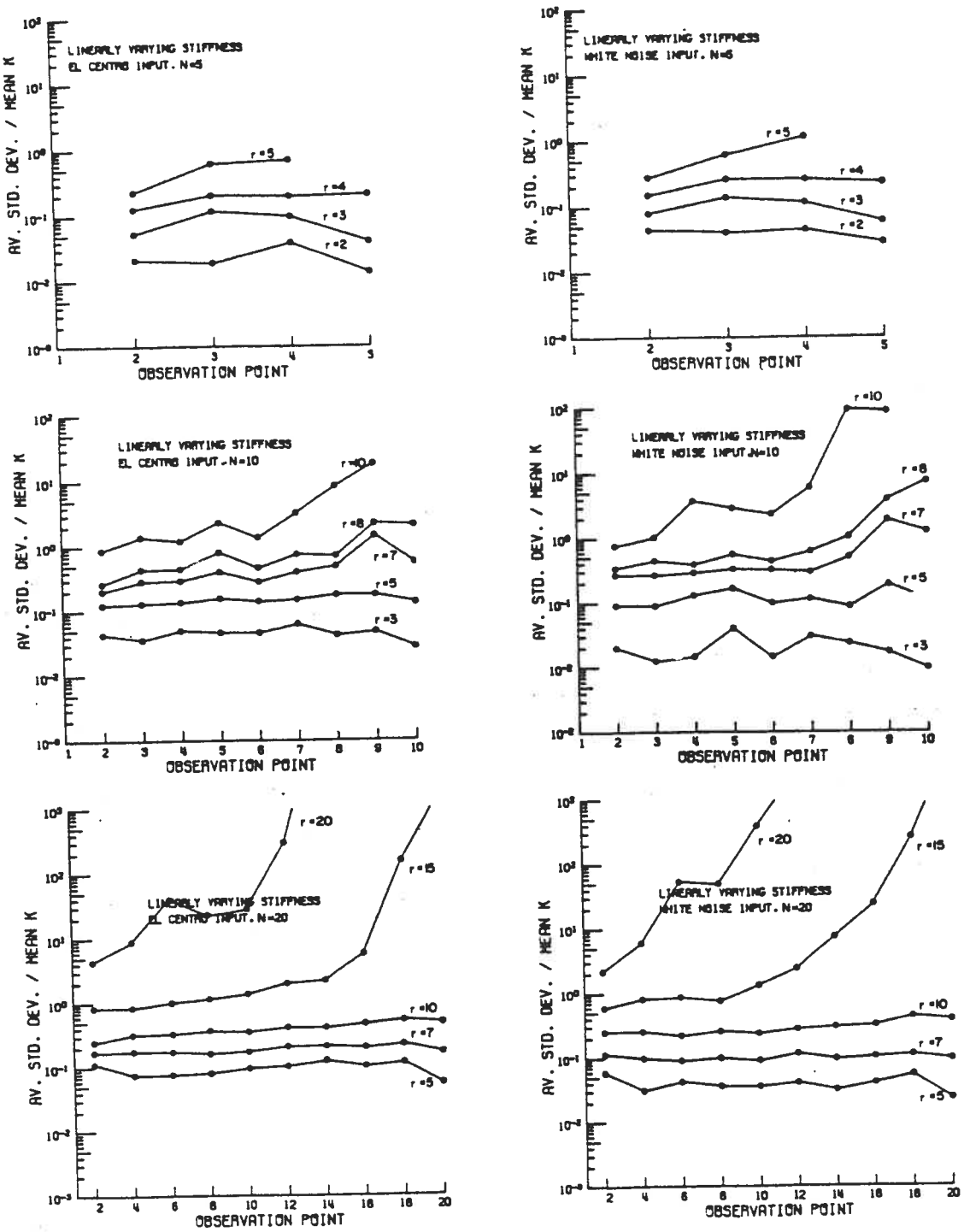

Fig. 2

depends on the inverse $A^{\dagger}$ of A used in analysis. Hence, this inverse should reproduce to the best possible extent, the properties of the iterative correction procedure employed in the identification scheme. It can be shown that the Lanczos inverse with $r < N$ used in the foregoing, is consistent with the iterative corrections made in the first-order gradient algorithm, which is a very efficient scheme for identification (Udwadia and Shah [7]). The gradient algorithm makes corrections in the parameter estimates,

$$\Delta \mathbf{k} = \alpha A^T \theta = V \Lambda U^T \theta = \alpha \sum_{i=1}^{N} \lambda_i (u_i{}^T \theta) \mathbf{v}_i \tag{35}$$

where α is a scalar parameter determined by one-dimensional minimization of J in the direction of $\Delta \mathbf{k}$. Thus the correction $\Delta \mathbf{k}$ has the largest component along that $\mathbf{v}_i$ which has the largest singular value, and the corrections in the subspace spanned by the columns of $\overline{V}_0$ will be very small. Thus the foregoing covariance analysis with $r < N$ is consistent with the gradient algorithm in that both ignore corrections along the less sensitive directions in the parameter space.

The conclusion about sensor location was verified using a simulated stiffness estimation problem with linearly decreasing "true" stiffness and $N = 10$. The estimation was done alternatively using a sensor located at m_1 and at m_N. In each case, the observations were simulated by integrating the model equations with the El Centro input using the true values of the stiffnesses, and adding a zero-mean purely

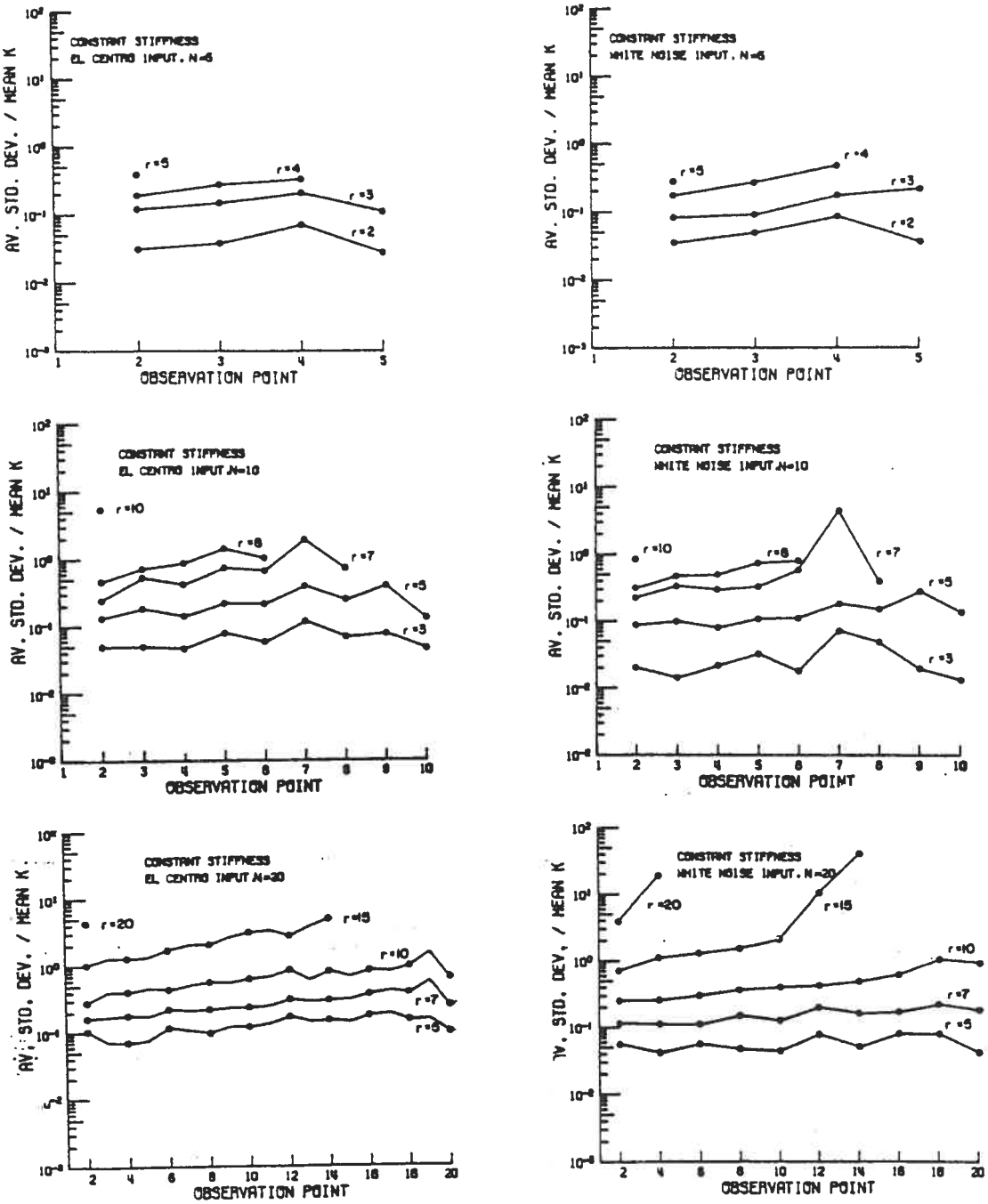

Fig. 3

random Gaussian sequence to the response at the sensor location at each time step. The standard deviation of this observation noise was approximately equal to 12 percent of the RMS roof level displacement. In each case, the value of J was reduced to the level of its expected residual value using a discrete analogue of the first-order gradient algorithm; in each instance, the initial guess was a uniform distribution as shown in Fig. 6. This matching was done without any explicit penalty constraints on the stiffness distribution, implying an absence of any additional *a priori* information [7].

It is evident from Fig. 6 that the stiffness estimates obtained with a sensor located at the second floor level are closer to the true values, than those obtained with the sensor located at the roof level. This result is in agreement with our expectations based on the foregoing analysis.

Conclusions

1 An approximate analysis is developed for optimally positioning sensors in lumped and distributed parameter systems for their identification from time-domain input-output data.

2 It is found that when the observations are insensitive to some components of the unknown parameters, additional information, in terms of an *a priori* probability distribution, about the unknowns is necessary to determine the accuracy of the parameter estimates.

3 The methodology developed in this work is applied to the problem of optimal location of a sensor in a building structure. The sensor is to be so located that the shear stiffnesses of the columns at the different floor levels can be most accurately determined from the displacement record obtained by the sensor and from the basement

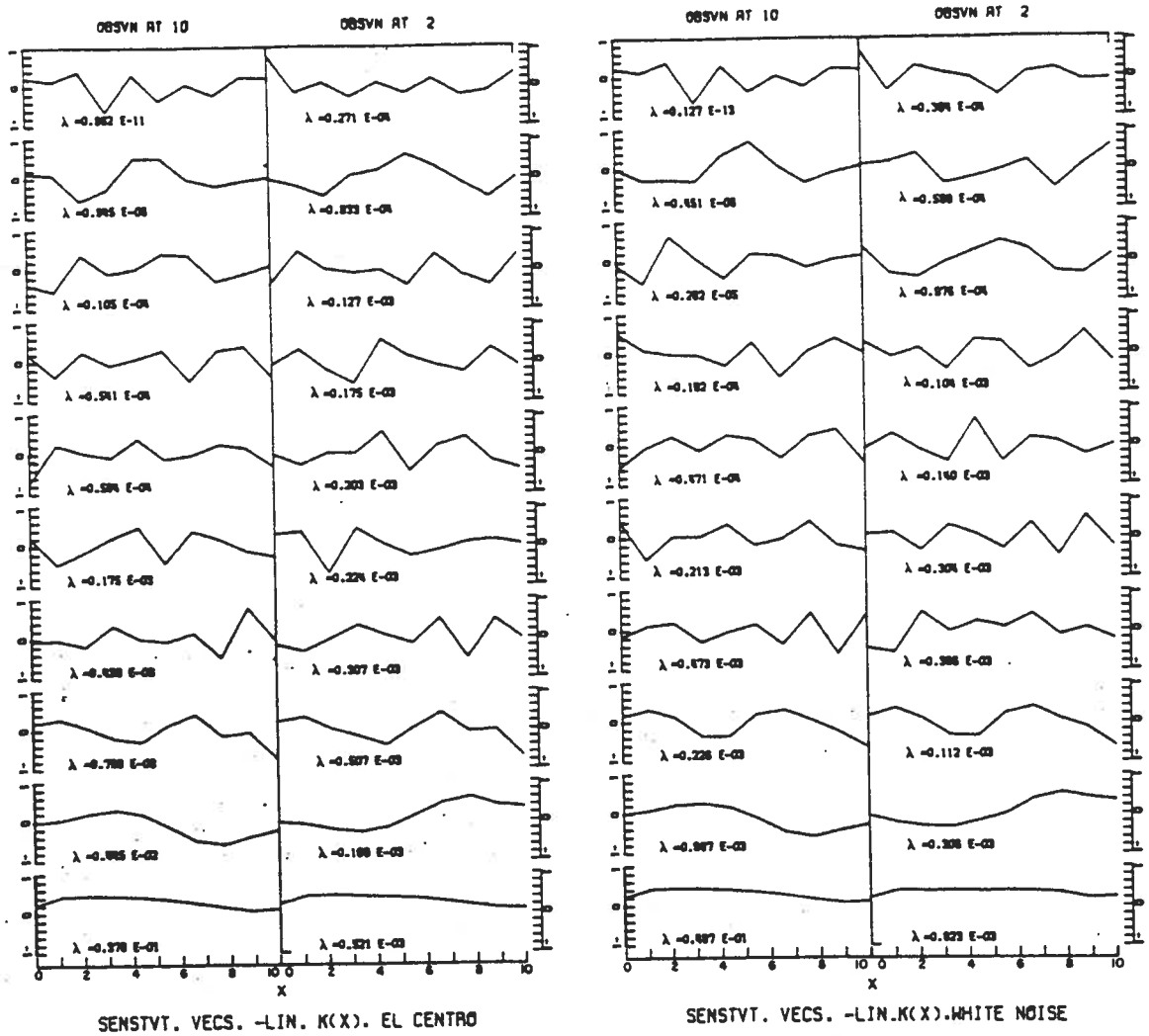

Fig. 4

input record. It has been found that for a variety of structures with different heights, different stiffness distributions and for different inputs, the optimal sensor location is at the floor level immediately above the basement.

Acknowledgments

The research reported in this paper was funded in part by grants received from the National Science Foundation.

References

1 Bushinger, P. A., and Golub, G. H., "Singular Value Decomposition of a Matrix," *Communications of the ACM*, Vol. 12, No. 10, Oct. 1969, pp. 564–565.

2 Jackson, D. D., "Interpretation of Inaccurate, Insufficient and Inconsistent Data," *Geophys. J. R. Astr. Soc.*, No. 28, 1972, pp. 97–100.

3 Lanczos, C., *Linear Differential Operators*, D. Van Nostrand Co., 1961, London.

4 Shah, P. C., "Estimation of Properties in Petroleum Reservoirs," PhD Thesis, California Institute of Technology, Pasadena, 1976.

5 Shah, P. C., and Udwadia, F. E., "Optimal Sensor Location for Identification of Building Structures Using Response to Earthquake Ground Motions," Paper No. 10-06, Sixth World Conference on Earthquake Engineering, New Delhi, India, Jan. 1977.

6 Shah, P. C., "Models and Identification Algorithms for Distributed Parameter Systems," Symposium on Applications of Computer Methods in Engineering, Session No. 11, University of Southern California, Los Angeles, Aug. 1977, pp. 613–724.

7 Udwadia, F. E., and Shah, P. C., "Identification of Structures Through Records Obtained During Strong Earthquake Ground Motion," ASME *Journal of Engineering for Industry*, Vol. 98, No. 4, 1976, pp. 1347–1362.

8 Udwadia, F. E., and Sharma, D. K., "Some Uniqueness Results Related to Building Structural Identification," *SIAM Journal of Applied Mathematics*, Vol. 24, No. 1, Jan. 1978.

APPENDIX

Computation of the Sensitivity Matrix

1 **Using Sensitivity Equations.** System equations are

$$M\ddot{u} + F(k)u = f(k)v(t) \quad t \in (0, T) \tag{36}$$

$$u(0) = 0, \quad \dot{u}(0) = 0$$

Differentiation with respect to the parameter k_i yields,

$$M\frac{d^2}{dt^2}\left(\frac{\partial u}{\partial k_i}\right) + F(k)\left(\frac{\partial u}{\partial k_i}\right) = -\frac{\partial F}{\partial k_i}u + \frac{\partial f}{\partial k_i}v(t), t \in (0, T)$$

$$\frac{\partial u}{\partial k_i}(0) = 0, \quad \frac{\partial \dot{u}}{\partial k_i}(0) = 0 \tag{37}$$

The integration of (36) is carried out once; subsequently, the integration of the system (37) is carried out once for each i. Thus, for N parameters, the integration of $(N + 1)$ systems of differential equations is required, all of which are of the same order.

2 **Using Adjoint Variable.** From (36), the equation for a variation $\delta u(t)$ in $u(t)$ due to a small variation δk in k is (denoting the corresponding variations in F and f by δF and δf),

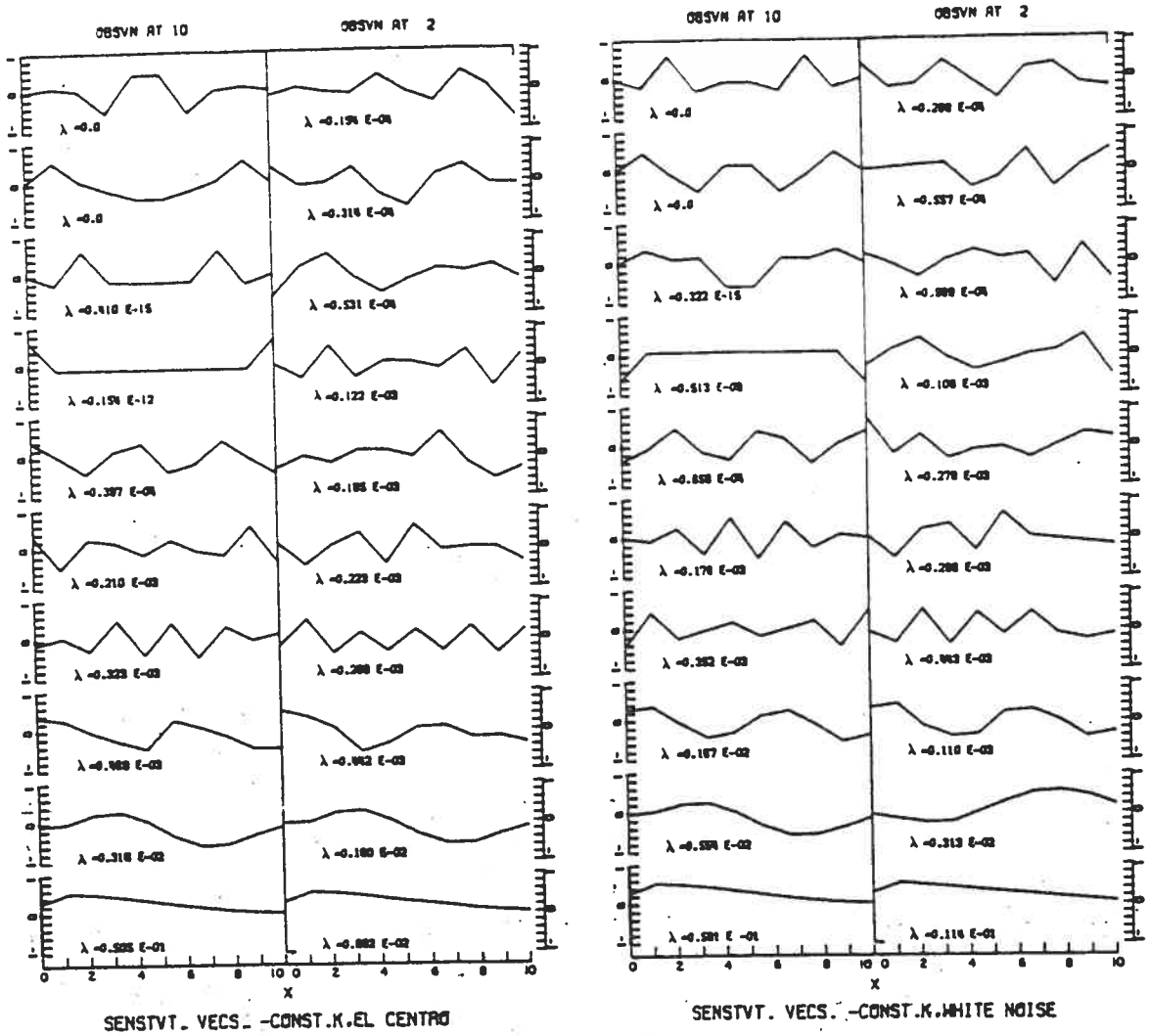

Fig. 5

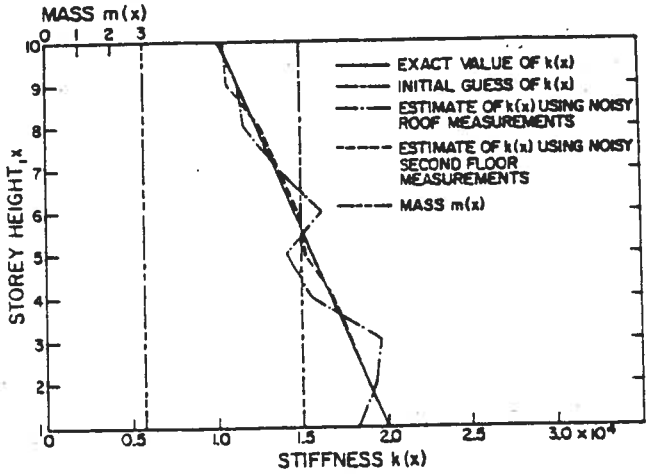

Fig. 6

$$M\delta\ddot{u} + F(k)\delta u = -\delta F u + \delta f v$$

$$\delta u(0) = \delta\dot{u}(0) = 0 \tag{38}$$

Premultiplying (38) by an adjoint vector $\lambda^T(t)$ and integrating over $(0, t_1)$, we obtain

$$\int_0^{t_1} \lambda^T M\delta\ddot{u}\,dt + \int_0^{t_1} \lambda^T F(k)\delta u\,dt = \int_0^{t_1} -\lambda^T \delta F u\,dt + \int_0^{t_1} \lambda^T \delta f v\,dt$$

Integration by parts yields,

$$\lambda^T M\delta\dot{u}|_{t=t_1} - \dot{\lambda}^T M\delta u|_{t=t_1} + \int_0^{t_1} \ddot{\lambda}^T M\delta u\,dt + \int_0^{t_1} \lambda^T F\delta u\,dt = \int_0^{t_1} -\lambda^T \delta F u\,dt + \int_0^{t_1} \lambda^T \delta f v\,dt$$

If we define $\lambda(t)$ so as to satisfy the differential equation,

$$M^T\ddot{\lambda} + F^T\lambda = 0 \quad t \in (0, t_1)$$

$$\lambda(t_1) = 0 \tag{39}$$

$$M^T\dot{\lambda}(t_1) = -[0, 0, \ldots, 1, 0, 0, \ldots, 0]^T$$

where the unity in the last vector is in the jth position, then we obtain

$$\delta u_j(t_1) = \int_0^{t_1} -\lambda^T \delta F u\,dt + \int_0^{t_1} \lambda^T \delta f v\,dt = \sum_i \delta k_i \left[\int_0^T -\lambda^T \frac{\delta F}{\delta k_i} u\,dt + \int_0^T \lambda^T \frac{\delta f}{\delta k_i} v\,dt \right] \tag{40}$$

from the expression (40) it follows that

$$\frac{\partial u_j(t_1)}{\partial k_i} = \int_0^{t_1} \lambda^T \left[\frac{-\partial F}{\partial k_i} u + \frac{\partial f}{\partial k_i} v \right] dt \tag{41}$$

Thus once $\lambda(t)$ is determined by solution of (39), the sensitivity coefficient of u_j at each time is obtained by a quadrature over $(0, t_1)$. If the system is time invariant, then the solution of the adjoint system (39) need only be done once with $t_1 = T$; solution for any other value of t_1 can be found from it be merely shifting the argument.

Thus, in addition to the system equation (36), it is necessary to integrate the adjoint equation (39) once for each sensor location j. Furthermore, for each observation time t_1, the quadrature over $(0, t_1)$ in (41) has to be performed, once for each parameter k_i. However, the computational effort in evaluation of this integral can be substantially reduced by noting that only n_e scalar quadratures of the form $\int_0^{t_1} u_p \lambda_s \, dt$ are required, where n_e is the number of nonzero elements in $F(\mathbf{k})$; the expressions for different values of i on the right-hand side of (41) are linear combinations of such integrals. When the number of sensors l is small, and the number M of samples is not very large, the adjoint variable method is more efficient than direct integration of the sensitivity equations. A detailed comparison of the relative computational requirements of the methods can be found in Shah [4].

The derivation of the adjoint variable method presented here can be extended in a straightforward manner to problems with systems modeled by either partial differential equations or by discrete sets of algebraic equations.