Interpretation and Improved Solution Approach for Ill-Conditioned Linear Equations

I. U. Ojalvo* and T. Ting†
University of Bridgeport, Bridgeport, Connecticut 06601

A simple method for obtaining physically realistic solutions to ill-conditioned or singular algebraic systems of linear equations is proposed and interpreted on both physical and mathematical grounds. Solutions obtained by this method are very close, and sometimes identical, to those obtained by the singular-value decomposition method, but are much simpler and more computationally efficient. The relationship of the method presented here, which we call epsilon-decomposition, to singular-value, Bayesian estimation and Levenberg-Marquardt methods for nonlinear parameter computations, is also presented.

Nomenclature

a,b = elements of a matrix
$[I]$ = identity matrix
$\{\boldsymbol{R}\}$ = residual vector
$\{r\}$ = parameters (or solution) of mathematical model of physical system
$\{r_0\}$ = starting estimate of mathematical model parameters
$[S]$ = sensitivity matrix
$[U]$ = modal matrix corresponding to nonzero eigenvalues
$\{\boldsymbol{Y}_a\}$ = analytically determined solution vector
$\{Y_e\}$ = experimentally determined values
$[\lambda]$ = diagonal eigenvalue matrix
$\{\Delta r\} = \{r - r_0\}$
$\{\Delta Y\} = \{Y_e - \boldsymbol{Y}_a\}$
ε = small parameter used to improve numerical ill conditioning
Φ = function to be minimized
$[\,]^T$ = transpose of matrix
$[\,]^{-1}$ = inverse of matrix

Introduction

THIS work concerns itself with an efficient method of obtaining physically realistic solutions to ill-conditioned or singular algebraic systems of linear equations. Such equations may arise from numerical scaling problems, data insufficiency, or poorly defined systems occurring in complex optimization, correlation, and mathematical model studies. For example, suppose we have failed to obtain sufficient independent test data to uniquely define all the unknown parameters of a physical system for which an analytical model exists, or perhaps we have failed to adequately normalize or scale the parameters of a system so that the magnitudes of the parameters sought are many orders of magnitude apart (e.g., the stiffness of a metal part vs its thickness). The resulting equations, under these circumstances, will generally lead to ill conditioning.

In this paper it is assumed that "reasonable" estimates of uncertain parameters are available despite the insufficiency of properly scaled and independent data. The method of singular-value decomposition has been proposed as a procedure for dealing with such systems. However, this approach requires a solution for all the nonzero eigenvalues (and their eigenvectors) of the coefficient matrix. This can become a very lengthy procedure if the rank of the matrix is large. We shall propose a much simpler numerical scheme, which is based on solving a similar but improved-conditioned system of equations. We have called this procedure epsilon-decomposition.

Mathematically stated, let m be the number of linear equations and n the number of unknown parameters. If there is insufficient data to competely define the model, this condition yields a matrix with rank deficiency of at least $n - m$. Alternately, m may be greater than n, but the resulting $n \times n$ least-squared-error system may still be ill conditioned.

It will be shown, subject to a reasonable solution-estimate assumption for the uncertain parameters, that a new approach to solve systems of the types described is possible. Furthermore, a mathematical basis for this approach will be provided as well as a physical and geometrical interpretation of the solution procedure suggested.

Problem Statement

Given a physical system with m pieces of experimental data $\{Y_e\}$, n uncertain parameters $\{r\}$, and the assumption that an analytical model exists. Let the analysis model consists of n estimated parameters $\{r_0\}$, and assume that this model yields the numerical solution values $\{\boldsymbol{Y}_a\}$.

It is desired to determine the "best" fit of parameters $\{r\}$ so that $|\{Y_e - \boldsymbol{Y}_a\}|$ is a minimum. That is, we wish to minimize the residual squared of a one-term Taylor approximation

$$\{Y_e - \boldsymbol{Y}_a\} = \frac{\partial\{\boldsymbol{Y}_a\}}{\partial\{r\}}\{r - r_0\} + \{\boldsymbol{R}\}$$

where the residual $\{R\}$ is the Taylor series remainder plus experimental error. This equation is rewritten as

$$\{\Delta Y\} = [S]\{\Delta r\} + \{\boldsymbol{R}\} \tag{1}$$

where $\{\Delta Y\} = \{Y_e - \boldsymbol{Y}_a\}$, $\{\Delta r\} = \{r - r_0\}$ and the $m \times n$ sensitivity matrix $[S]$ is then defined as

$$[S] = \frac{\partial\{\boldsymbol{Y}_a\}}{\partial\{r\}}$$

If $m > n$, there are more data than parameters, and we shall seek a least-error-squared solution, i.e.,

$$\text{minimize } \Phi = \{\boldsymbol{R}\}^T\{\boldsymbol{R}\}$$

Presented as Paper 89-1179 at the 30th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Mobile, AL, April 3-5, 1989; received May 30, 1989, revision received Nov. 24, 1989; accepted for publication Dec. 12, 1989.

*Bullard Professor of Mechanical Engineering.
†Assistant Professor of Mechanical Engineering.

or

$$\frac{\partial\{\boldsymbol{R}\}^T\{\boldsymbol{R}\}}{\partial\{r\}} = \{0\}$$

which leads to the n equations

$$[S^TS]\{\Delta r\} = [S^T]\{\Delta Y\} \tag{2}$$

If $m < n$, $[S^TS]$ is singular. Then there are an infinite number of solutions, and it is not clear which one to use.

Let us pursue the first case (i.e., $m > n$). If $[S^TS]$ is numerically invertible, then a least-error-squared solution is possible by efficient Cholesky (triangular) factorization of $[S^TS]$ into $[LL^T]$ and subsequent forward and backward substitution. However, even if n is only moderately large, a solution can become difficult due to ill-conditioning. Also, if the rank of $[S^TS]$ is less than n, factorization is not possible.

One possible approach to these problems is to use the singular-value decomposition solution approach.[1] This technique requires obtaining all the $q\ (\leq n)$ nonzero eigenvalues $[\lambda]$ of $[S^TS]$ and the corresponding $n \times q$ modal matrix $[U]$.

The pseudoleast-error-squared solution is then given by

$$\{\Delta r\} = [U][\lambda]^{-1}[U]^T[S]^T\{\Delta Y\} \tag{3}$$

since $[S^TS]^{-1}$ is replaced by $[U][\lambda]^{-1}\,[U]^T$, where

$$[S^TS][U] = [U][\lambda] \quad \text{and} \quad [U]^T[U] = [I]$$

Epsilon Decomposition

The previously discussed solution approach is laborious since it requires the solution of $[U]$ and $[\lambda]$. A much simpler alternative procedure is as follows. Pick a small value ε, and set

$$[\overline{S^TS}] = [S^TS] + \varepsilon[I] \tag{4}$$

such that $[\overline{S^TS}]$ may be efficiently factored into $[\bar{L}\bar{L}^T]$ where $[\bar{L}]$ is a lower triangular. This is then used to solve

$$[\overline{S^TS}]\{\overline{\Delta r}\} = [S]^T\,\{\Delta Y\}$$

i.e., $\{\overline{\Delta r}\} = [\bar{L}]^{-T}[\bar{L}]^{-1}[S]^T\{\Delta Y\}$.

Next, decrease ε until the elements of $\{\Delta \bar{r}\}$ approach asymptotic values or the factorization of $[\bar{S}^T\bar{S}]$ into $[\bar{L}\bar{L}^T]$ breaks down due to ill-conditioning.

This method has been termed a Levenberg-Marquardt method by Luenberger,[2] who considers it to be a modification of Newton's method and steepest descent (see pp. 225–227 of Ref. 2). However, neither Levenberg nor Marquardt considered this as a method for solving ill-conditioned systems of equations, but rather thought of it as a method for stabilizing iterations for nonlinear systems.

We shall attempt to give some insight into why this method works from both a mathematical and physical view point and its relationship to singular-value decomposition. This will then lead us to a more direct solution approach when $m < n$ and will give us an alternate interpretation of Bayesian estimation[3] techniques as well.

Consider the two degree-of-freedom example with $m = 1$, where

$$[S] = [a \quad b]$$

therefore,

$$[S^TS] = \begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix}$$

and $|S^TS| = 0$.

By Eq. (4)

$$[\overline{S^TS}] = \begin{bmatrix} a^2+\varepsilon & ab \\ ab & b^2+\varepsilon \end{bmatrix}$$

and so

$$[\overline{S^TS}]^{-1} = \frac{1}{\varepsilon(\varepsilon + a^2 + b^2)}\begin{bmatrix} b^2+\varepsilon & -ba \\ -ab & a^2+\varepsilon \end{bmatrix}$$

Therefore, the least-error-squared solution becomes

$$\{\overline{\Delta r}\} = [\overline{S^TS}]^{-1}[S]^T\{\Delta Y\}$$

or

$$\{\Delta r\} = \lim_{\varepsilon\to 0}\{\overline{\Delta r}\} = \lim_{\varepsilon\to 0}\begin{Bmatrix} \dfrac{a\,\Delta y}{a^2+b^2+\varepsilon} \\ \dfrac{b\,\Delta y}{a^2+b^2+\varepsilon} \end{Bmatrix} = \begin{Bmatrix} \dfrac{a\,\Delta y}{a^2+b^2} \\ \dfrac{b\,\Delta y}{a^2+b^2} \end{Bmatrix}$$

A geometrical interpretation of this solution shows that the ideal solution occurs for the point on the line $[S]\{\Delta r\} = \{\Delta Y\}$, which is closest to the origin (see Fig. 1).

Note, if $\varepsilon \neq 0$, then the solutions fall short of this ideal point.

A comparison of our proposed epsilon-decomposition approach for this same problem, with the method of singular-value decomposition yields the identical result but through different computational steps as follows.

The nonzero eigenvalue λ of $[S^TS]$ is $\lambda = a^2 + b^2$ and its corresponding normalized eigenvector $\{U\}$ is

$$\{U\} = \frac{1}{(a^2+b^2)^{1/2}}\begin{Bmatrix} a \\ b \end{Bmatrix}$$

Thus, the pseudoinverse of $[S^TS]$, called $[S^TS]^{-P}$ here, is given by

$$[S^TS]^{-P} = \{U\}[\lambda]^{-1}\{U\}^T = \frac{1}{(a^2+b^2)^2}\begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix}$$

This does not yield $[I]$ when multiplied by $[S^TS]$. However, it does yield the same solution for $\{\Delta r\}$ as epsilon-decom-

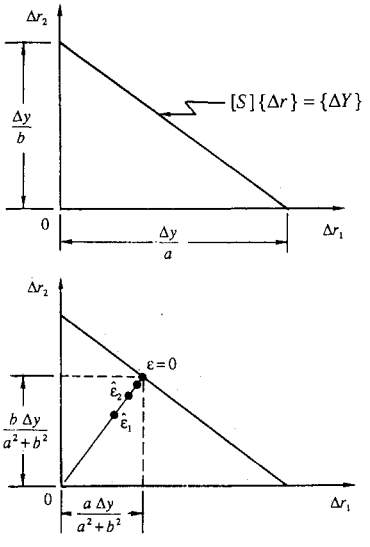

Fig. 1 Geometric interpretation.

position (as ε approaches zero), i.e.,

$$\{\Delta r\} = [S^T S]^{-P}[S]^T\{\Delta Y\} = \frac{1}{(a^2+b^2)}\begin{Bmatrix} a\,\Delta y \\ b\,\Delta y \end{Bmatrix}$$

The similarity of epsilon-decomposition with singular-value decomposition for very small ε may be established through a comparison of the modal expansions of $[S^T S]$ and $[\overline{S^T S}]$, i.e., compare

$$[\overline{S^T S}] = [U][\lambda][U]^T, \qquad [U]^T[U] = [I]$$

$$[S^T S]^{-P} = [U_1][\lambda_1]^{-1}[U_1]^T, \qquad \lambda_i \neq 0$$

with $[\overline{S^T S}]$ and its inverse

$$[\bar{S}^T\bar{S}] = [\bar{U}]([\lambda] + \varepsilon[I])[\bar{U}]^T$$

and

$$[\overline{S^T S}]^{-1} = [\bar{U}]([\lambda] + \varepsilon[I])^{-1}[\bar{U}]^T$$

Thus, epsilon-decomposition would seem to give accurate results as long as ε is small compared to the smallest nonzero eigenvalue of $[S^T S]$.

Geometrical Interpretation

The previous equations suggest that an n degree-of-freedom solution approach should not only minimize the sum of the equation residuals squared, but should also minimize the distance on this line to the origin of $\{\overline{\Delta r}\}$ space, i.e.,

$$\text{minimize } \bar{\Phi} = \{\boldsymbol{R}\}^T\{\boldsymbol{R}\} + \varepsilon\{\overline{\Delta r}\}^T\{\overline{\Delta r}\}$$

which leads to

$$\frac{\partial \Phi}{\partial\{r\}} = \{0\}$$

or

$$([S^T S] + \varepsilon[I])\{\overline{\Delta r}\} = [S]^T\{\Delta Y\} \tag{5}$$

If $[S^T S]$ is not ill-conditioned, simply set ε to zero in Eq. (5) and solve by $[LL^T]$ decomposition. However, if it is ill conditioned, chose an ε and solve by $[\bar{L}\bar{L}^T]$ decomposition. Keep reducing ε unitl the solution asymptotes, or this type of factorization breaks down numerically.

The choice of selecting a starting ε is arbitrary as a practical matter, but the two degree-of-freedom example suggests that ε should be small compared to the trace of $[S^T S]$. However, it is dangerous to generalize in such matters. One may, of course, obtain the smallest nonzero eigenvalue of $[S^T S]$ and chose ε so that it is very small compared to it but not so small as to cause numerical problems during numerical decomposition.

Numerical Application

A finite element model for a cantilevered beam containing 28 beam elements of distinct area moments of inertia about the Z axis (I_{ZZ}) was selected for numerical application of the proposed method. This original set of 28 sectional properties was perturbed slightly from their baseline values to generate simulated test data. The results of a modal analysis for the perturbed model were then used to represent a set of experimental modal data. Ten eigenvalues corresponding to natural bending modes about the Z axis were selected as the elements of the experimental data vector $\{\boldsymbol{Y}_a\}$. Thus, the design sensitivity matrix $[S]$, computed by any one of the methods in Refs. 4 or 5, was a 10×28 matrix. The correlation problem, therefore, yielded an undetermined problem, which is normally solved by the minimum "norm" solution for $\{\Delta r\}$, i.e.,

$$\{\Delta r^*\} = [S]^T[SS^T]^{-1}\{\Delta Y\} \tag{6}$$

In this case $[SS^T]$ becomes a 10×10 nonsingular matrix. As long as $[SS^T]$ is not singular, Eq. (6) will yield the minimum change solution for $\{\Delta r\}$. If we try to solve the problem by the least-squares method, i.e.,

$$\{\Delta r\} = [S^T S]^{-1}[S^T]\{\Delta Y\} \tag{7}$$

the lower-upper LU decomposition of the 28×28 matrix $[S^T S]$ will break down due to its being singular.

However, Eq. (7) may be easily solved using the epsilon-decomposition approach, i.e.,

$$\{\overline{\Delta r}\} = ([S^T S] + \varepsilon[I])^{-1}[S^T]\{\Delta Y\} \tag{8}$$

The LU decomposition of $([S^T S] + \varepsilon[I])$ is valid for a large range of ε values, and it will be shown that the approximate solution $\{\overline{\Delta r}\}$ of Eq. (8) is "close" to the "minimum change" solution, $\{\Delta r^*\}$ of Eq. (6).

Table 1 presents a comparison of the minimum change solution $\{\Delta r^*\}$ with the approximate solutions $\{\overline{\Delta r}\}$ for a wide range of ε values. For the sake of brevity, only six of the 28 parameter changes are shown in Table 1.

Table 1 shows clearly that when the ε chosen is relatively small, say, smaller than 1×10^{-4} in this case, that epsilon-decomposition provides a good practical minimum change solution. There is also an indication that if ε is made too small, the quality of the approximation begins to deteriorate prior to decomposition breakdown. For this example, the LU decomposition broke down at approximately $\varepsilon = 1 \times 10^{-18}$, and some errors approached 10 to 15% when $\varepsilon = 1 \times 10^{-17}$. It is believed that this breakdown was caused by the finite precision of the computer.

Physical Interpretation of Proposed Procedure

Strictly speaking, the unique solution of an ill-conditioned system of equations is nonsense. What we have shown is that a unique solution is possible if an additional condition is introduced, in the form of a constraint, upon the minimization problem; namely, that the solution must not only satisfy the original m equations in a least-squared-error sense, but if

Table 1 Typical solutions $\{\overline{\Delta r}\}$ vs ε for 6 of the 28 variable changes for a cantilever beam problem

Param. number	$\{\Delta r^*\}$	1.E0	1.E-2	1.E-4	1.E-8	1.E-16	1.E-17
2	0.00462	0.00010	0.00337	0.00460	0.00462	0.00455	0.00433
7	−0.00225	−0.00004	−0.00131	−0.00224	−0.00225	−0.00220	−0.00190
12	0.00597	0.00005	0.00340	0.00592	0.00594	0.00594	0.00581
17	−0.00695	−0.00001	−0.00400	−0.00690	−0.00695	−0.00694	−0.00705
22	0.00719	0.00009	0.00439	0.00714	0.00719	0.00726	0.00797
27	−0.00774	−0.00009	−0.00561	−0.00771	−0.00774	−0.00781	−0.00766

the resulting $n \times n$ system is ill conditioned, then the proper solution should be the one that is closest to the initial mathematical model. This latter condition is sufficient to uniquely determine the solution in most cases. In our minimization formulation, epsilon represents a weighting of the closeness-to-the-origin constraint, $|\{r - r_0\}| = \min$, relative to the size of the m-equation residual $|\{\boldsymbol{R}\}|$ minimization. Thus, if we wish to minimize $|\{\boldsymbol{R}\}|$, ε should be made as small as practically possible while still delivering a unique solution.

Physically speaking, our approach implies that the initial solution guess $\{r_0\}$ is reasonable and that if the formulation leads to some insufficiency in uniquely defining the system (i.e., an ill-conditioned $[S^T S]$ matrix), then the criterion of closeness of $\{r\}$ to $\{r_0\}$ should be imposed to uniquely define the system. By making ε as small as is practical, we are simply weighting this closeness of $\{r\}$ to $\{r_0\}$ condition as secondary to the least-squared-error criterion. Viewed in this way, we may consider a different weighting of the various parameters $\{r\}$ through the diagonal weighting matrix $[W_r]$. Continuing with this concept for the various residuals $\{\boldsymbol{R}\}$ as well, through the weighting matrix $[W_Y]$, we arrive at a Bayesian[3] formulation, i.e., to find the minimum of Φ where Φ is given by

$$\Phi = \{\boldsymbol{R}\}^T[W_Y]\{\boldsymbol{R}\} + \{r - r_0\}^T[W_r]\{r - r_0\}$$

Thus, $\varepsilon[I]$ may be thought of as an $\{r - r_0\}$ weighting matrix in Bayesian estimation.

Conclusion

We have presented a new numerical technique for solving ill-conditioned systems of equations that does not rely upon computation of the system's nonzero eigenvalues and eigenvectors (as is necessary with singular-value decomposition). We have named this new technique epsilon-decomposition.

While others have proposed this same or similar procedures, it is felt that we have given it a more formal and less heuristic derivation. In addition, we have supplied a physical interpretation that gives insight into the results that this method produces as well as its relation to Bayesian estimation methods.

The work described is useful for mathematical modeling and optimization studies in which complex and insufficient information regularly result in underdetermined and ill-conditioned systems of linear algebraic equations. In addition, the requirement of a reasonable starting guess is also often met. Therefore, the present work should have broad applicability.

Acknowledgments

The authors wish to acknowledge the partial support of this effort from the State of Connecticut through Department of Higher Education Grant 014 and a matching grant by Sikorsky Aircraft.

References

[1]Dahquist, G., and Bjurck, A., *Numerical Methods*, Prentice-Hall, NJ, 1974.

[2]Luenberger, D. G., *Linear and Nonlinear Programming*, 2nd ed., Addison-Wesley, Reading, MA, 1984.

[3]Isenberg, J., "Progressing from Lease Squares to Bayesian Estimation," American Society of Mechanical Engineers Paper 79-WA/DSC-b, New York, Dec. 1979.

[4]Nelson, R. B., "Simplified Calculation of Eigenvector Derivatives," *AIAA Journal*, Vol. 14, No. 9, 1976, pp. 1201–1205.

[5]Ojalvo, I. U., "Efficient Computation of Mode-Shape Derivatives for Large Dynamic Systems," *AIAA Journal*, Vol. 25, No. 10, 1987, pp. 1386–1390.