N89-25213

AN INTERPRETATION AND SOLUTION OF ILL-CONDITIONED LINEAR EQUATIONS

I.U. Ojalvo and T. Ting
Mechanical Engineering Department
University of Bridgeport
Bridgeport, CT

PRECEDING PAGE BLANK NOT FILMED

INTRODUCTION

Data insufficiency, poorly conditioned matrices and singularities in equations occur regularly in complex optimization, correlation, and interdisciplinary model studies. This work concerns itself with two methods of obtaining certain physically realistic solutions to ill-conditioned or singular algebraic systems of linear equations arising from such studies.

Two efficient computational solution procedures that generally lead to locally unique solutions are presented when there is insufficient data to completely define the model, or a least-squares error formulation of this system results in an ill-conditioned system of equations.

If it is assumed that a "reasonable" estimate of the uncertain data is available in both cases cited above, then we shall show how to obtain realistic solutions efficiently, in spite of the insufficiency of independent data.

The proposed methods of solution are more efficient than singular-value decomposition [1] for dealing with such systems, since they do not require solutions for all the nonzero eigenvalues of the coefficient matrix.

OBJECTIVES

REVIEW MATHEMATICAL FORM OF ILL-CONDITIONING

OUTLINE PHYSICAL SOURCES OF ILL-CONDITIONING

REVIEW SINGULAR-VALUE DECOMPOSITION SOLUTION APPROACH

PRESENT A COMPUTATIONALLY SIMPLER METHOD

OFFER MATHEMATICAL & GEOMETRIC INTERPRETATION OF SOLUTIONS

DISCUSS PHYSICAL INTERPRETATION AND RELATIONSHIP TO:

1. MATH MODEL IMPROVEMENT
2. GRADIENT METHODS
3. BAYESIAN ESTIMATION

PROBLEM STATEMENT

Given the physical system with "m" bits of experimental data, $\{ Y_e \}$ and "n" uncertain parameters $\{ r \}$,

assume that a math model exists which yields the "m" values $\{ Y_a \}$ based upon the "m" estimated parameters $\{ r_o \}$.

It is desired to determine the "best" fit of parameters $\{ r \}$ so that $\| y_e - y_a \|$ is a minimum.

A one-term Taylor approximation yields:

$$\{ Ye - Ya \} = \left[\frac{\partial \{Ya\}}{\partial \{r\}} \right] \{ r - r_o \} + \{ R \}$$

where $\{ R \}$ is the Taylor series truncation error plus experimental error.

The mXn sensitivity matrix [S] is then defined below.

SENSITIVITY MATRIX:
$$[S]_{M \times N} = \left[\frac{\partial \{ Y_A \}}{\partial \{ \Delta R \}} \right]$$

NO. PARAMETERS:
$$\{ \Delta R \}_{N \times 1} = \{ R - R_0 \}$$

SOLUTION AND OUTPUT DATA:
$$\{ \Delta Y \}_{M \times 1} = \{ Y_E - Y_A \}$$

RESIDUAL:
$$\{ R \}_{M \times 1} = \{ \Delta Y \} - [S] \{ \Delta R \}$$

TAYLOR SERIES:
$$\{ Y_E \} = \{ Y_A \} + [S] \{ \Delta R \} + RES$$

LEAST-SQUARES MINIMUM RESIDUAL:
$$[S^T S] \{ \Delta R \} = [S]^T \{ \Delta Y \}$$

1. If $m > n$, there are more data than parameters, and so we shall seek a least-error-squared fit.

2. If $m < n$, then there are an infinite number of solutions and it is not clear which one to use.

Let us pursue case (1) first (i.e. $m > n$). If $[S^T S]$ is numerically invertible, then a least-error-squared solution is possible by efficient triangular factorization of $[S^T S]$ into $[LL^T]$, and subsequent forward and backward substitution. However, if "n" is only moderately large, then solutions can become difficult due to ill-conditioning. Also, if the rank of $[S^T S]$ is less than n, factorization is not possible. Now, if $m < n$ and a solution is somehow obtained, what is its interpretation?

One possible approach to these problems is to use the singular-value decomposition (SVD) solution approach [1]. This technique requires obtaining all the q ($\leq n$) nonzero eigenvalues $[\lambda]$ of $[S^T S]$ and the corresponding nxq modal matrix, [U]. If the rank of $[S^T S]$ is n, then the interpretation of SVD is clear. However, if the rank is less than n (as it may be for $m > n$ and will be for $m < n$), interpretations of the SVD solutions are not obvious.

M > N	:	1.	$S^T S$	WELL CONDITIONED (NUMERICALLY & PHYSICALLY)
		2.	$S^T S$	ILL-CONDITIONED (NUMERICALLY, BUT NOT PHYSICALLY)
M < N	:	3.	$S^T S$	ILL-CONDITIONED (NUMERICALLY & PHYSICALLY)

CASE OF WELL-CONDITIONED NUMERICALLY, BUT NOT PHYSICALLY, IS NOT CONSIDERED

IF CASE 1: $[S^T S] = [L][L]^T$, L = CHOLESKY DECOMPOSITION = $\begin{bmatrix} \ddots & 0 \\ & \end{bmatrix}$

IF CASE 2: $[S^T S]^{-1} \approx U \; [\lambda]^{-1} \; U^T$ SINGULAR-VALUE DECOMP.

N X N NXQ QXQ QXN

$\lambda = (Q \leq N)$ NON-ZERO EIGENVALUES OF $[S^T S]$

[U]=CORRESPONDING MODAL MATRIX

EPSILON DECOMPOSITION

The previously discussed procedure is laborious at best. A much simpler procedure is now presented. First, pick a small value ε, and set

$$[\underline{S^T S}] = [S^T S] + \varepsilon [I]$$

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

$$[\underline{S^T S}] \{\Delta r\} = [S]^T \{\Delta y\}$$

Next, decrease ε until $\{\Delta r\}$ approaches an asymptote, or the factorization of $[\underline{S^T S}]$ into $[\underline{L}\underline{L}^T]$ 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 Reference 2).

(SIMPLER PROCEDURE)

$$\overline{[S^T S]} \equiv [S^T S] + \varepsilon [I] = [\bar{L}] [\bar{L}]^T$$

$$\varepsilon_1 > \varepsilon_2 > \varepsilon_3 > \ldots > 0$$

BUT, WHAT DO SOLUTIONS MEAN?

EXPLANATION BASED ON SIMPLE 2-DOF INTERPRETATION

TWO DOF EXAMPLE

Consider the two degree-of-freedom (DOF) example with m=1, where we are given:

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

Therefore,

$$[\underline{S^T S}] = \begin{bmatrix} a^2 + \varepsilon & a \cdot b \\ a \cdot b & b^2 + \varepsilon \end{bmatrix}, \quad |S^T S| = 0$$

$$[S^T S]^{-1} = \frac{1}{\varepsilon\,(\varepsilon + a^2 + b^2)} \begin{bmatrix} b^2 + \varepsilon & -a\,b \\ -a\,b & a^2 + \varepsilon \end{bmatrix}$$

Thus, the least-error-squared solution becomes $\{\Delta \tilde{r}\} = [\underline{S^T S}]^{-1}[S]^T \{\Delta y\}$

or

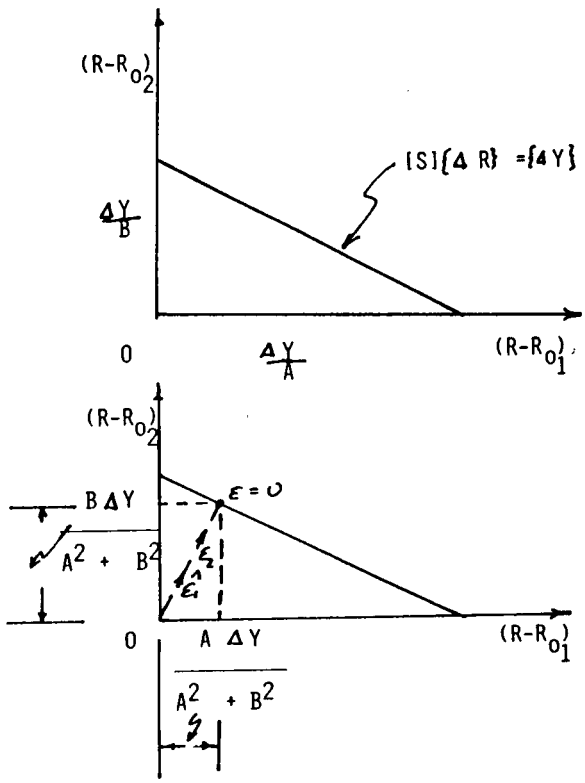
$$\lim_{\varepsilon \to 0} \{\Delta \tilde{r}\} = \lim_{\varepsilon \to 0} \begin{Bmatrix} \frac{a\,\Delta y}{a^2 + b^2 + \varepsilon} \\ \frac{b\,\Delta y}{a^2 + b^2 + \varepsilon} \end{Bmatrix} = \begin{Bmatrix} \frac{a\,\Delta y}{a^2 + b^2} \\ \frac{b\,\Delta y}{a^2 + b^2} \end{Bmatrix}$$

GEOMETRIC INTERPRETATION

A geometrical interpretation of the solution shows that the ideal solution occurs for the point on the line $[S]\{\Delta r\} = \{\Delta y\}$ which is closest to the origin.

Note that if $\mathcal{E} > 0$, then the solutions fall short of this ideal point.

A comparison of our proposed epsilon-decomposition approach with the method of singular-value decomposition yields the identical result for $\{\Delta r\}$, but through different computational steps.



MATHEMATICAL INTERPRETITION

The previous two-pages suggest that an n-DOF solution approach is not only to minimize the sum of the residuals squared but also to minimize the distance on this line to the origin of $\{\Delta r\}$ space, i.e.

$$\underline{\Phi} = R^T R + \varepsilon \Delta r^T \Delta r$$

If $[S^T S]$ is not ill-conditioned, simply set ε to zero and solve by $[LL^T]$ decomposition. However, if it is ill-conditioned, choose an ε and solve by $[\underline{LL}^T]$ decomposition. Keep reducing ε until the solution asymptotes to certain values, or this type of factorization breaks down numerically.

Choosing a starting ε is arbitrary, but the two-DOF example suggests that ε should be small compared to the smallest eigenvalue of $[S^T S]$. However, it is dangerous to generalize.

We have shown that this procedure works, not only when $m > n$, but also when $m < n$ (which was initially called case 2 here). However, a more direct solution is possible if $m < n$ and this is shown on the following page.

$$\{R\} = \{\Delta Y\} - [S]\{\Delta R\}$$

$$\bar{\Phi} = \{R\}^T\{R\} + \varepsilon \{\Delta R\}^T\{\Delta R\}$$

$$\left\{\frac{\partial \bar{\Phi}}{\partial \{\Delta R\}}\right\} = \{0\}, \quad ([S^T S] + \varepsilon [I])\{\Delta R\} = [S]^T\{\Delta Y\}$$

UNDER-DETERMINED SYSTEMS

For the case where $m < n$, $[S^T S]^{-1}$ will not exist. However, if the rank of $[S^T S]$ is m, then its inverse will exist, and it is possible to show that $[S^T S]^{-1}[S]^T$ is equivalent to $S^T [SS^T]^{-1}$ (as ε approaches zero). Therefore, it is not necessary to use epsilon in determining $\{\Delta r\}$, for $m < n$, when the rank of $[SS^T]$ is m. However, if its rank is less than m, we may still wish to use the present method on either $[S^T S]$ or $[SS^T]$. Our method will yield almost identical results if epsilon is sufficiently small, as compared to the smallest non-zero eigenvalue of either $[S^T S]$ or $[SS^T]$.

$M < N$ CASE, $\underset{M \times M}{|S\,S^T|} \neq 0$

GIVEN: $[S^T S]\{\Delta R\} = [S]^T\{\Delta Y\}$, $\underset{NXN}{|S^T S|} = 0$, $\{\Delta \overline{R}\} = \overline{[S^T S]^{-1}}\,[S]^T\{\Delta Y\}$

CONSIDER: $\lim_{\varepsilon \to 0} \overline{[S^T S]}^{-1} [S^T] = \lim_{\varepsilon \to 0} (S^T S + \varepsilon I)^{-1} [S^T]$

$\therefore [S^T S]\,[S]^T = \lim_{\varepsilon \to 0} (S^T S + \varepsilon I)\, S^T$ . . . . . (1)

ASSUME: $|S\,S^T| \neq 0$, POST-MULTIPLY EQ (1) BY $[S\,S^T]^{-1}$

$\therefore S^T = \lim_{\varepsilon \to 0} (S^T S + \varepsilon [I])\, S^T\, [S\,S^T]^{-1}$ . . . . (2)

ASSUME: $|\overline{S^T S}| \neq 0$, $\varepsilon > 0$

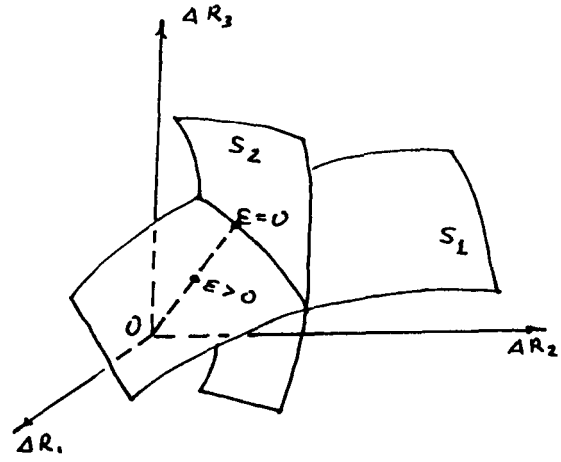
PREMULT. EQ(2) BY $\lim_{\varepsilon \to 0} \overline{[S^T S]}^{-1}$

THEN, $\lim_{\varepsilon \to 0} \overline{[S^T S]}^{-1}\, S^T = S^T\, [S\,S^T]^{-1}$

AND SO $\{\Delta R\} = S^T\, [S\,S^T]^{-1}\, \{\Delta Y\}$

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 the resulting nxn system is ill-conditioned, then the proper solution should be closest to the initial math 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, $|r-r_o|$ = min, relative to the size the m-equation residual, $|R|$, minimization. Thus, if we wish to minimize $|R|$ ϵ should be made as small as practically possible while still delivering a unique solution for $\{\Delta r\}$.



(NON-UNIQUE SOLUTIONS TO S_1 & S_2 EXIST ALONG INTERSECTION)

$\bar{\Phi} = \{R\}^T\{R\} + \epsilon \{\Delta R\}^T [\Delta R\}$ = MIN (MAKES PROBLEM WELL-POSED & SOLUTION UNIQUE)

$\{\Delta R\} = \{0\}$ IS THE INITIAL GUESS

RELATIONSHIP TO EXISTING METHODS

Physically speaking, our approach implies that the initial solution-guess, $\{r_o\}$, 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 $\{r_o\}$ shall be imposed to uniquely define the system. By making ε as small as is practical, we are simply weighting this closeness of $\{r\}$ to $\{r_o\}$ condition as secondary to the least-square-error criterion. Viewed in this way, we may consider a different weighting of the various parameters of $\{r\}$ through the diagonal weighting matrix $[\sim W_r \sim]$. Continuing this concept for the various residuals, $\{R\}$, as well as through the weighting matrix $[W_y]$ we arrive at a Bayesian formulation i.e. to find the minimum of Φ where Φ is given by:

$$\Phi = \{R\}^T [W_y] \{R\} + \{r - r_o\}^T [\sim W_r \sim] \{r - r_o\}.$$

NEWTON METHODS: $\varepsilon_1 > \varepsilon_2 \cdots$ TRAVEL ALONG GRADIENT TO SOLUTION FROM INITIAL GUESS

BAYESIAN ESTIMATION:

MINIMIZE: $\Phi = \{R\}^T [W_Y] \{R\} + \{\Delta R\}^T [W_R] \{\Delta R\}$

$[W_Y]$: WEIGHTED RESIDUALS

$[W_R]$: ε [I] PROPOSED SOLUTION APPROACH

CONCLUSIONS

We have derived a numerical technique for solving ill-conditioned systems of equations which does not rely upon computation of the system's non-zero eigenvalues and eigenvectors (as is necessary with singular-value decomposition) . While others have proposed this same procedure, it is felt that we have given it a more formal and less heuristic derivation. In addition, we have supplied a physical interpretation which gives insight into the results which the method yields, as well as its relation to Bayesian estimation methods.

AVOIDED USE OF SINGULAR-VALUE DECOMPOSITION
(COMPUTATIONALLY HARDER)

PROVIDED MATHEMATICAL BASIS FOR PROPOSED SOLUTION APPROACH

PRESENTED A GEOMETRIC INTERPRETATION

PRESENTED A PHYSICAL INTERPRETATION

IMPROVED INSIGHT INTO SOLUTION OF ILL-CONDITIONED EQS.

REFERENCES

1. Dahlquist, G. and Bjurck, A., "Numerical Methods", Prentice-Hall, N.J., 1974

2. Luenberger, D.G., "Linear and Nonlinear Programming", 2nd Ed., Addison-Wesley, Reading, Mass, 1984.

BIBLIOGRAPHY

Isenberg, J., "Progressing from Least Squares to Bayesian Estimation", ASME Paper 79-WA/DSC-1, Dec. 1979.

SYMBOLS

a, b	Elements of a matrix
$\{r\}$	Parameters (or solution) of math model being sought
$\{r_o\}$	Starting estimate of math model parameters
$\{\Delta r\}$	$r - r_o$
$\{Y_a\}$	Analytically determined solution vector
$\{Y_e\}$	Experimental determined values
$\{\Delta Y\}$	$Y_e - Y_a$
R	Residual
S_1, S_2	2-dimensional surfaces in 3-dimensional space
ε	Epsilon parameter
Φ, $\underline{\Phi}$	Functions to be minimized
[I]	Identity matrix
[L]	Lower triangular matrix
[S]	Senitivity matrix
[U]	Modal matrix corresponding to non-zero eigenvalues
$[W_r]$	Initial parameter confidence weighting matrix
$[W_y]$	Experimental data confidence weighting matrix
$[\lambda]$	Diagonal eigenvalue matrix
T	Transpose of a matrix (used as a superscript)
-1	Inverse of matrix (used as a superscript)
-p	Pseudo-inverse (used as a superscript)