COMPUTER METHODS IN APPLIED MECHANICS AND ENGINEERING 2 (1973) 339–348

A METHOD FOR SOLVING ILL-POSED INTEGRAL EQUATIONS OF THE FIRST KIND*

Krishna P. Singh
President and CEO
Holtec International
Marlton, NJ

and

Burton Paul

Improperly-posed (or Hadamard-incorrect) problems may arise when numerical solutions are extremely sensitive to a discretization process. The nonconformal contact problem in three-dimensional elastostatics falls into this category. It is shown how such contact stress problems may be formulated and successfully solved using the "Functional Regularization Method" of Tychonov. The Functional Regularization Method requires the use of a parameter, called the Regularization Parameter. Although no general rules for the choice of such a parameter appear to exist, we have determined appropriate bounds on the parameter for a wide class of contact problems (including that of Hertz). The method developed should be capable of extension to more general ill-posed problems. It is also shown that refinements in the discretization process such as reduced mesh lengths or higher order quadrature formulas may postpone, but do not necessarily remove the numerical difficulties associated with the physics of the problem.

Nomenclature

a	Radius of contact region
B	Coefficient matrix (eq. 6)
$f_1(x, y)$ $f_2(x, y)$	Profile functions of surface 1 and 2 respectively
f	Right hand side vector (eq. 6)
F	Normal thrust
H	Kernel of integral equation
k	Elastic parameter
$p(x, y)$	Pressure
p^*	Dimensionless pressure
r^*	Dimensionless distance from origin
δ	Normal approach term
λ	Regularization parameter

* Part of a dissertation submitted by K.P. Singh, as partial fulfilment of requirements for the degree of Ph. D. at the University of Pennsylvania.

$\mu_{\min}$ Numerically smallest eigenvalue of matrix $D = B^t B$
φ A certain banded matrix
Φ Auxiliary function
Ψ Functional
Ω Contact region (region of integration)

1. Introduction

Physically meaningful problems often give rise to numerically unstable equations. Quite frequently, the governing equations cannot be solved analytically, and a numerical solutions must be sought. Numerical solutions necessitate discretization of the continuous operators resulting in discretization errors. If the governing equations are sensitive to start with, then such errors may completely invalidate the solution. Such problems fall into the category known as "Hadamard-incorrect' (Lavrentiev, 1965). Concomitant with the evolution of high speed computing devices, there has been a great upsurge of interest in the solution of these problems. Apart from being mathematically interesting, their occurrence in a variety of important physical situations has made them immensely significant. Lavrentiev (1965) gives solutions for many such problems. Tychonov (1963, 1963, 1965) proposed new criteria for "ill-posedness" of a problem which are far more relaxed than those of Hadamard (see Hadamard, 1952). Tychonov's criteria gave rise to a new approach, the "Functional Regularization Method", for treating improperly posed problems. This concept has been successfully applied in solving problems of nonconformal contact of elastic bodies (Singh, 1972).

In this paper we show that a certain parameter called the "regularization parameter" that occurs in the FR method can be bounded. The bound is perfectly general although it is developed here with particular reference to the contact stress problem. In section 2 an improperly posed problem that arises in contact of elastic bodies is outlined. The limitations of direct numerical solution of this problem and the attributes of an associated coefficient matrix are discussed in sections 3. The functional regularization method is described in section 4, and a bound on the regularization parameter is developed in section 5. Results and conclusions are given in section 6.

2. An improperly posed problem

Consider two convex frictionless surfaces in contact as shown in fig. 1. A common thrust, F, acts along the Z-axis as shown in fig. 1. In the frame of the coordinate axes shown in fig. 1, the two undeformed surfaces are defined by

$$z_1 = f_1(x, y), \qquad z_2 = f_2(x, y). \tag{1}$$

In the absence of a normal thrust, the contact is at a point (point O in fig. 1). When the normal thrust is applied, a contact region Ω is formed. The points far removed from the contact region

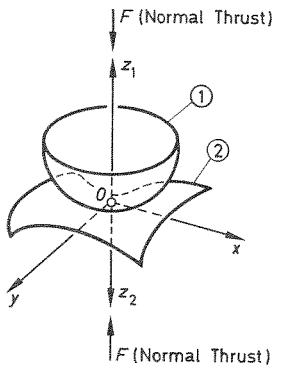

Fig. 1. Nonconformal surfaces in contact.

undergo only rigid body translation. The central equation of this problem is

$$k \int_{\Omega} \frac{p(x', y')\mathrm{d}x'\mathrm{d}y'}{[(x-x')^2 + (y-y')^2]^{1/2}} = \delta - f_1(x, y) - f_2(x, y), \tag{2}$$

where k is a constant which depends on the elastic properties of the two bodies, δ is the rigid body approach term, and $p(x, y)$ is the unknown pressure distribution. Eq. (2) holds for every point inside Ω. In particular, at point O, $f_1(x, y) = f_2(x, y) = 0$; hence

$$k \int_{\Omega} \frac{p(x', y')\mathrm{d}x'\mathrm{d}y'}{(x'^2 + y'^2)^{1/2}} = \delta, \tag{3}$$

and δ can be eliminated between eqs. (2) and (3). Thus we have an integral equation of the first kind in two dimensions:

$$\int_{\Omega} H(x, y, x', y')\, p(x', y')\, \mathrm{d}x'\mathrm{d}y' = f(x, y), \tag{4}$$

where

$$f(x, y) = f_1(x, y) + f_2(x, y),$$

and the kernel H is given by

$$H(x, y, x', y') = k\{(x'^2 + y'^2)^{-1/2} - [(x-x')^2 + (y-y')^2]^{-1/2}\}. \tag{5}$$

Equation (4) must be solved for the unknown pressure distribution $p(x, y)$. The function $f(x, y)$ depends on the surface profiles of the two bodies. Analytical solutions of eq. (4) are known for only special forms of the function $f(x, y)$. In general, for an arbitrary function $f(x, y)$ a numerical solution must be attempted. A method of solution utilizing piecewise constant pressure over discrete subregions is briefly described next. Details of the method may be found in Singh (1972).

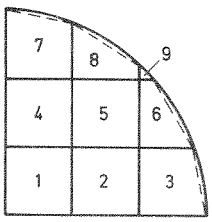

Fig. 2. Circular contact region (9 cells per quadrant).

3. Discretization of the integral equation

The region of integration Ω is subdivided into small cells. The internal cells are squares, but those on the boundary are polygonal to conform with the boundary contour. The curved segments of the boundary are replaced by straight lines. Fig. 2 shows typical cell layouts for a circular contact region formed by the contact of axisymmetric surfaces (e.g. two spheres). In fig. 2 advantage is taken of symmetry about coordinate axes, and only one quadrant of the circle is subdivided. The pressure p_i in cell i is considered to be constant. The centroids of the cells are assumed to be field points. By utilizing a suitable scheme of numerical quadrature, the integral in eq. (4) can be evaluated as a linear combination of cell pressures p_i for every one of the field points. Thus, if there are n cells, n linear algebraic equations in n unknown pressures are generated. In subscript notation, eq. (4) is replaced by

$$B_{ij}p_j = f_i, \qquad i = 1, 2, 3, \ldots, n. \tag{6}$$

It would appear at first sight that eq. (6) could be solved using standard Gaussian elimination procedures. However, such an approach will not be successful. As an illustration, the contact of two spheres of radii R_1 and R_2 is considered. The pressure distribution corresponding to a contact radius $a = 0.006R$ is sought, where

$$R = \frac{2R_1R_2}{R_1 + R_2}\,. \tag{7}$$

The cell layout corresponding to fig. 2 is employed. The solution vector is given in the last column of table 1. The pressure and radial distance of field points in table 1 are given in dimensionless forms, defined by

$$\begin{aligned} &\text{Dimensionless distance from the origin, } r^* = r/R, \\ &\text{Dimensionless pressure, } p^* = kp. \end{aligned} \tag{8}$$

The analytical solution of this problem due to Hertz (1881, 1882) is also given in table 1. It is seen that the numerical solution is grossly in error and physically inadmissible because some of the cell pressures are negative (for cells 3, 6, 7, 8, and 9 in table 1). It was also found that making small perturbations in matrix $\boldsymbol{B}$ causes large variations in the solution vector. Details of numerical ex-

Table 1
Cell pressures for direct solution and regularized solution, spheres in contact (fig. 2), $a = 0.006$.

Cell No. i	Dimensionless Radial Distance from origin $r^* \times 10^2$	Dimensionless Pressure, $p_i^* \times 10^2$		
		Analytical Solution, Hertz	Regularized Solution $\lambda = \lambda^* = 0.006219$	Direct Solution $\lambda = 0$
1	0.1414	0.2363	0.2368	0.1248
2	0.3163	0.2066	0.2116	0.0781
3	0.5014	0.1335	0.1231	−0.1516
4	0.3163	0.2066	0.2117	0.0781
5	0.4242	0.1719	0.1659	0.0021
6	0.5382	0.1074	0.0979	−0.2742
7	0.5014	0.1335	0.1231	−0.1516
8	0.5382	0.1074	0.0978	−0.2742
9	0.5883	0.0478	0.0645	−1 325

periments on $\boldsymbol{B}$ may be found in Singh (1972). Study of the character of $\boldsymbol{B}$ revealed that although it is highly sensitive, its determinant is far from zero. The Functional Regularization Method (FR) was used with success (Singh, 1972) in solving eq. (6). A brief description follows of the considerations needed to apply the method to contact stress problems.

4. The Functional Regularization Method

Assume that the cells are numbered sequentially from left to right and from bottom up as shown in fig. 2. Let the cell number of the right-most cell in row i be given by m_i. If there are q rows and a total of n cells, then $m_q = n$. Consider the auxiliary function Φ given by

$$\Phi = \sum_{j=1}^{q} \sum_{i=m_{j-1}+1}^{m_j-1} (p_i - p_{i+1})^2, \tag{9}$$

where $m_0 = 0$. The function Φ is nonnegative and becomes zero (minimum) when all components of p in each row (j = constant) are equal. The FR method avoids a direct solution of eq. (6) by seeking a minimum of the functional

$$\Psi(p_i) = (B_{ik}p_k - f_i)(B_{ij}p_j - f_i) + \lambda\Phi, \tag{10}$$

where λ is a suitably chosen parameter.

The rationale for the method may be explained in the following heuristic terms. We know that it is profitless to solve eqs. (6) $\boldsymbol{Bp} = \boldsymbol{f}$ exactly because the discretization errors in $\boldsymbol{B}$ have an unfortunately pronounced effect on the resulting solution vector. Therefore, we should allow some nonvanishing residual vector $\boldsymbol{\epsilon} = \boldsymbol{Bp} - \boldsymbol{f}$, but we should require that it be small; i.e. $\epsilon^t\epsilon$ should be small. On the other hand we require that p should be a smooth function (no jumps) due to the physics of

the problem; hence Φ should likewise be small. Thus we desire that some linear combination of $\epsilon^t\epsilon$ and Φ as defined in eq. (10) should be as small as possible. The major practical difficulty, at the moment, is that we have no guidance for the choice of the "smoothness weighting factor" λ which appears in eq. (10). The technique which we have generated to choose a numerical value for λ (called the "regularization parameter") will now be explained.

5. Bounding the Regularization Parameter

Finding a p which will minimize Ψ (defined by eq.10) is equivalent to solving the equation set:

$$B_{ij}B_{il}p_j + \frac{\lambda}{2}\frac{\partial\Phi}{\partial p_l} = B_{il}f_i, \qquad l = 1, 2, \ldots, n. \tag{11}$$

In matrix form eq. (11) is given by

$$Dp + \lambda\varphi p = B^t f, \tag{12}$$

where

$$D = B^t B, \tag{13}$$

and φ is a banded matrix with zero elements everywhere except for the diagonal and the immediately adjacent elements. If there are q rows of cells and there are k_i cells in the ith row such that $n = \Sigma_{i=1}^{n} k_i$, then φ may be partitioned in the block-diagonal form

$$\varphi = \lceil \varphi_1 \varphi_2 \dot{\varphi}_3, \ldots, \varphi_8 \rfloor, \tag{14}$$

where the submatrix φ_i is a triadiagonal matrix of order k_i

$$\varphi_i = \begin{bmatrix} 1 & -1 & 0 & & & \\ -1 & 2 & -1 & & \circ & \\ 0 & -1 & 2 & -1 & & \\ & & & \ddots & & \\ & \circ & & -1 & 2 & -1 \\ & & & 0 & -1 & 1 \end{bmatrix} \tag{15}$$

Since D is a symmetric matrix, its Euclidean norm and inverse are given by (Fox, 1965)

$$\begin{aligned} \|D\| &= \mu_{\max}, \\ \|D^{-1}\| &= \frac{1}{\mu_{\min}}, \end{aligned} \tag{16}$$

where $\mu_{\max}$ and $\mu_{\min}$ are the numerically largest and smallest eigenvalues of D, respectively. Thus, if μ_i ($i = 1, \ldots, n$) denote the eigenvalues of D, then $\mu_{\max} = \max(|\mu_1|, |\mu_2|, \ldots, \mu_n)$ and $\mu_{\min}=$

$\min(|\mu_1|, |\mu_2|, \ldots |\mu_n|)$. We define ϵ_f as the Euclidean norm of the relative error vector in the solution of the equation

$$Dp = B^t f, \tag{17}$$

That is,

$$\epsilon_f = \frac{\|Dp - B^t f\|}{\|B^t f\|}. \tag{18}$$

Experience has shown that for typical contact stress problems an error of $\epsilon_f = 10^{-2}$ is permissible. In especially well-behaved problems (e.g. contacting spheres) ϵ_f may be driven down as low as 10^{-4}. Where ϵ_f is excessively large (greater than 0.1), it may be concluded that the problem is too improperly posed to permit a meaningful regularized solution. We have not yet encouneterd such a case among the many contact problems studied.

Thus ϵ_f is a central quantity in the regularization method, and serves as an indicator of the accuracy and relevance of the solution to the physical problem considered. Hence it is logical to obtain a bound on λ in terms of ϵ_f.

Premultiplication of eq. (12) by D^{-1} yields

$$p + \lambda D^{-1}\varphi p = D^{-1} B^t f, \tag{19}$$

or

$$p = D^{-1} B^t f - \lambda D^{-1}\varphi p.$$

Taking norms and appealing to the triangle inequality of matrix norms,

$$\|p\| \leq \|D^{-1} B^t f\| + \lambda \|D^{-1}\varphi p\|, \tag{20}$$

or

$$\|p\| \leq \|D^{-1}\| \, \|B^t f\| + \lambda \|D^{-1}\| \, \|\varphi\| \, \|p\|. \tag{21}$$

Let

$$\|p\| = \beta_1, \qquad \|B^t f\| = \beta_2, \qquad \|\varphi\| = \beta_3. \tag{22}$$

From eq. (16), (21) and (22)

$$\beta_1 \leq \frac{\beta_2}{\mu_{\min}} + \frac{\beta_3 \beta_1 \lambda}{\mu_{\min}}, \tag{23}$$

or

$$\frac{\beta_2}{\beta_1} \geq (\mu_{\min} - \beta_3 \lambda). \tag{24}$$

Furthermore, eq. (12) yields

$$\lambda \varphi p = B^t f - Dp.$$

Taking norms and using the matrix inequality $\|Ax\| \leq \|A\| \|x\|$

$$\lambda \|\varphi\| \|p\| \geq \|B^t f - Dp\| . \tag{25}$$

From relations (18) and (25)

$$\lambda \beta_3 \beta_1 \geq \epsilon_f \beta_2 , \tag{26}$$

or

$$\frac{\beta_2}{\beta_1} \leq \frac{\lambda \beta_3}{\epsilon_f} .$$

From inequalities (24) and (26) a lower bound on λ is found by eliminating β_2/β_1 :

$$\frac{\lambda \beta_3}{\epsilon_f} \geq \mu_{\min} - \beta_3 \lambda , \tag{27}$$

or

$$\lambda \geq \frac{\epsilon_f \mu_{\min}}{\beta_3 (1 + \epsilon_f)} . \tag{28}$$

To find a bound on $\beta_3 = \|\varphi\|$, appeal is made to Gerschgorin's Theorem (called the Levy-Hadamard Theorem by Bodewig, 1959), which states that all eigenvalues of a general matrix G of order $n \times n$ lie in a closed domain C_D in the complex plane consisting of the union of all disks Q_i ($i = 1, 2, \ldots, n$) with the center of Q_i located at G_{ii} (the ith diagonal element of G), and the radius of disk Q_i is given by

$$r_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |G_{ij}| . \tag{29}$$

Since φ is symmetric, all its eigenvalues are real. From the structure of φ and the Gerschgorin-Levy-Hadamard theorem, it is readily concluded that all the eigenvalues of φ lie between 0 and 4. In fact, φ is a singular matrix, and its maximum eigenvalues asymptotically approach 4 for increasing n. Thus

$$\|\varphi\| = \beta_3 \leq 4. \tag{30}$$

Substituting in ineq. (28) yields

$$\lambda \geq \frac{\epsilon_f \mu_{\min}}{4(1 + \epsilon_f)} . \tag{31}$$

Since we desire to keep $\epsilon_f \ll 1$, we may write

$$\lambda > \epsilon_f \mu_{\min}/4. \tag{32}$$

In other terms

$$\epsilon_f < 4\lambda/\mu_{\min}. \tag{33}$$

Since $\mu_{\min} = \|D^{-1}\|$ may be calculated for any given matrix D, we keep the "error" ϵ_f sufficiently low by keeping λ sufficiently small. On the other hand, if λ is too small, the "smoothing constraint" is too weak to provide a physically meaningful solution in a poorly posed problem. As a practical guide we use the greatest value of λ which will ensure, via ineq. (33), that $\epsilon_f < 0.01$. Thus a lower bound on λ is obtained in terms of ϵ_f and the numerically smallest eigenvalue of D.

Ineq. (31) may be used for an initial estimate of λ. Then eq. (12) is solved. If any of the pressures turn out to be negative, then λ is slightly incremented. The process is continued until all the pressures become positive, the corresponding $\lambda = \lambda^*$ being the "optimum" value of the regularization parameter for the problem.

6. Discussion and results

The essence of the FR method lies in properly utilizing certain available information from the physics of the problem. In the problem of elastic contact of fairly smooth non-adhesive surfaces, it is known *a priori* that all field point pressures should be positive, and the "pressure surface" should be reasonably smooth. This smoothness constraint is introduced through the functional Φ and an adjustable weighting parameter λ. The positiveness of pressures was used to determine when λ was large enough. The minimum value of λ which produces all positive pressures is designated by λ^*. It was observed that taking λ moderately larger than λ^* does not hurt the solution. It was found that taking λ equal to 3 times the optimum value causes less than 2% change in the peak pressure. Table 1 summarizes the results for the field point pressures for $\lambda = \lambda^*$ and $\lambda = 0$.

As would be expected, the accuracy of the solution improves with increased cell density. Table 1 shows that the results are reasonably good using even a rather coarse cell layout (fig. 2). Total computation time on an IBM 370/165-machine (UNICOLL CORP, Philadelphia) was 0.9 seconds using the mesh of fig. 2. To obtain more refined solutions, a finer cell grid must be employed. For example, using 33 cells per quadrant gives excellent results for the "spheres-in-contact" problem (Singh, 1972). Any additional mesh refinement may cause substantial increase in the computation time and may not be worthwhile for most practical solutions.

We have conjectured that the ill-posedness in this problem is introduced by the errors made in the discretization process. To test this hypothesis, we utilized a more accurate discretization technique than that sketched above. For example, instead of considering pressure to be piecewise constant in each cell, we assumed pressure to be a quadratic function of nodal points. This "Quadratic Interpolation Method" is described in detail in Singh (1972). It was established that more accurate quadrature methods made only a marginal difference. The Quadratic Interpolation Method gave a

reasonably good direct solution (i.e., by solution of eqs. 6) for the cell layout of fig. 2 (9 cells per quadrant), but broke down at the next higher cell density of 15 cells per quadrant. Thus by using refined discretization schemes, we only postponed the difficulty. In order to obtain a reliable solution procedure, recourse to some other method, e.g., the Functional Regularization Method, was unavoidable.

References

[1] E. Bodewig, Matrix calculus (North-Holland, Amsterdam, Interscience, New York, 1959).
[2] L. Fox, An introduction to numerical linear algebra (Oxford Univ. Press, New York, 1965).
[3] J.H. Hadamard, Lectures on Cauchy's problem (Dover Publications, N.Y., 1952).
[4] H. Hertz, Miscellaneous papers (Trans. by D.E. Jones and G.A. Schott) (Macmillan, London, 1896). Translated from
(i) On the contact of elastic solids, Journal fur die reine und angewandte Mathematik 92 (1881) pp. 156–171.
(ii) On the contact of rigid elastic solids and an hardness, Verhandlungen des Vereins zur Beforderung des Gewerbefleisses (Nov. 1882).
[5] I.B.M. Scientific Subroutine Package, System/360 (360-CM-03x) Version II, Programmers Manual, H20-0205-2 (IBM Corp., White Plains, New York, 1968).
[6] M. Lavrentiev, Some improperly posed problems of mathematical physics (Springer, 1965).
[7] K.P. Singh, Contact stresses in elastic bodies with arbitrary profiles, Ph. D. Thesis, University of Pennsylvania (1972).
[8] A.N. Tychonov, "Regularization of incorrectly posed problems", Soviet Math. 4 (1963) pp. 1624–1627.
[9] A.N. Tychonov, "Solution of incorrectly posed problems and the regularization method", Soviet Math. 4 (1963) 1035–1038.
[10] A.N. Tychonov, "Incorrect problems of linear algebra and a stable method for their solution", Soviet Math. 6 (1965) pp. 988–991.