NASA Technical Memorandum 83524

Numerical Methods and Computers Used in Elastohydrodynamic Lubrication

Bernard J. Hamrock and John H. Tripp
Lewis Research Center
Cleveland, Ohio

Prepared for the
Tenth Leeds-Lyon Symposium on Tribology
Lyon, France, September 6-9, 1983

NASA

NUMERICAL METHODS AND COMPUTERS USED IN ELASTOHYDRODYNAMIC LUBRICATION

Bernard J. Hamrock and John H. Tripp

National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio 44135

SUMMARY

Some of the methods of obtaining approximate numerical solutions to boundary-value problems that arise in elastohydrodynamic lubrication are reviewed. The highlights of four general approaches (direct, inverse, quasi-inverse, and Newton-Raphson) are sketched. Advantages and disadvantages of these approaches are presented along with a flow chart showing some of the details of each. The basic question of numerical stability of the elastohydrodynamic lubrication solutions, especially in the pressure spike region, is considered. Computers used to solve this important class of lubrication problems are briefly described, with emphasis on recently developed supercomputers.

E-1797

INTRODUCTION

Elastohydrodynamic lubrication is a form of fluid-film lubrication where elastic deformation of the lubricated surfaces becomes significant. It is usually associated with highly stressed machine components such as rolling-element bearings and gears. Historically, elastohydrodynamic lubrication may be viewed as one of the major developments in the field of lubrication in the twentieth century. Its recognition not only revealed the previously unsuspected regime of lubrication in highly stressed nonconformal machine elements, but it also brought order to the complete spectrum of lubrication regimes, ranging from boundary to hydrodynamic.

The present paper attempts to review the methods of obtaining approximate numerical solutions to boundary-value problems that arise in tribology. The central task is to reduce the relevant differential and integral equations to algebraic ones that can be solved by familiar methods. Though the material is primarily illustrated by elastohydrodynamic lubrication problems, it is directly applicable to other areas of lubrication as well as to other engineering disciplines such as heat transfer and fluid mechanics.

The highlights of four main approaches to the elastohydrodynamic lubrication problem, namely the direct method, the inverse method, the quasi-inverse method, and the Newton-Raphson method, are covered, and the advantages and disadvantages of each method are discussed. The important question of numerical stability of solutions, especially in the pressure spike region, is considered. No attempt is made herein to be rigorous or complete. The chief purpose is to introduce the various techniques in a systematic manner and to indicate their general regimes of validity. Computers used to solve this important class of lubrication problem are briefly described with an emphasis on recently developed supercomputers.

SYMBOLS

a,b	Constants used to define density, m^2/N
D	Influence coefficient used in elasticity calculation, m^3/N
E	Modulus of elasticity, N/m^2
E'	Effective elastic modulus, $2\Big/\left(\frac{1-\nu_a^2}{E_a}+\frac{1-\nu_b^2}{E_b}\right)$, N/m^2
f	Iteration damping factor
g	Output/input amplitude ratio used in stability analysis
h	Film shape, m
h_e	Film shape obtained from elasticity calculation, m
h_i	Film shape obtained from inverse Reynolds equation, m
h_0	Film shape constant, m
h_m	Film thickness when $dp/dx = 0$, m
K	Wavelength used in stability analysis
p	Pressure, N/m^2
p_0	Initial pressure, N/m^2
p_1	Inlet pressure, N/m^2
p_2	Pressure in contact and outlet, N/m^2
R	Effective radius, m
R_i	Residual vector
r	Curvature radius, m
T,S	Banded matrices
u	Mean velocity, $(u_a + u_b)/2$, m/s
X_α	Eigenvector
x,y	Coordinate system
α	Pressure-viscosity coefficient of lubricant, m^2/N
δ	Elastic deformation, m
Δ	Pressure difference, $p^{(n+1)} - p^{(n)}$, N/m^2
η	Absolute viscosity at gage pressure, $N\ s/m^2$
η_0	Viscosity at atmospheric pressure, $N\ s/m^2$
λ_α	Eigenvalue
ν	Poisson's ratio
ρ	Lubricant density, $N\ s^2/m^4$
ρ_0	Density at atmospheric pressure, $N\ s^2/m^4$

Subscripts:

a	Solid a
b	Solid b
x,y	Coordinate system

BASIC EQUATIONS

The elastohydrodynamic lubrication problem is as follows: given two elastic, independently rotating solids immersed in oil and pressed together by an external force, find the pressure distribution and film shape in the region of the lubricating contact. This requires calculating the pressure distribution within the conjunction, at the same time allowing for the effects this pressure will have on the properties of the fluid and on the geometry of the elastic solids. The solution also provides the shape of the lubricant film, particularly the minimum film thickness between the solids.

The basic equations used in elastohydrodynamic lubrication are as follows:

Lubrication Equation (Reynolds Equation)

$$\frac{\partial}{\partial x}\left(\frac{\rho h^3}{\eta}\frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{\rho h^3}{\eta}\frac{\partial p}{\partial y}\right) = 12u\frac{\partial}{\partial x}(\rho h) \tag{1}$$

where $u = (u_a + u_b)/2$

Viscosity Variation

$$\eta = \eta_0 e^{\alpha p} \tag{2}$$

where η_0 is the coefficient of absolute dynamic viscosity at atmospheric pressure and α is the pressure-viscosity coefficient of the fluid. Values of αp may be as high as 10.

Density variation

Typically

$$\rho = \rho_0\left(1 + \frac{ap}{1 + bp}\right) \tag{3}$$

For mineral oils the values of a and b yield a maximum density increase of about 35 percent.

Elasticity Equation

$$\delta = \frac{2}{\pi E'}\iint \frac{p(x,y)\,dxdy}{\sqrt{(x - x_1)^2 + (y - y_1)^2}} \tag{4}$$

where

$$E' = 2 \Big/ \left(\frac{1 - \nu_a^2}{E_a} + \frac{1 - \nu_b^2}{E_b} \right)$$

Film Shape Equation

$$h(x,y) = h_0 + \frac{x^2}{2R_x} + \frac{y^2}{2R_y} + \delta(x,y) \qquad (5)$$

where

$$\frac{1}{R_x} = \frac{1}{r_{ax}} + \frac{1}{r_{bx}}$$

$$\frac{1}{R_y} = \frac{1}{r_{ay}} + \frac{1}{r_{by}}$$

Approaches that have been used in analyzing elastohydrodynamically lubricated conjunctions are presented briefly in the following sections.

DIRECT METHOD

The direct method solves the Reynolds equation for the pressure distribution arising from a given film shape. A flow diagram of the direct method used by Hamrock and Dowson (1976) is presented in Fig. 1. The modified pressure shown on the left of Fig. 1 contains a damping factor to control numerical convergence. This pressure is then used to determine changes in the film shape. However, for maximum Hertzian contact stresses exceeding 0.5 GPa, even with this damping factor, the direct method is sometimes found to diverge. Nevertheless, Hamrock and Dowson (1981) were able to obtain useful formulas that cover a complete spectrum of contact geometries (ranging from point to line contacts), materials (hard and soft), and lubricant availability (fully flooded or starved conditions). These theoretical film thickness formulas were found to have a pleasing agreement with the experimental findings of Dalmaz and Godet (1978), Kunz and Winer (1977), and Koye and Winer (1980) even at relatively large maximum Hertzian contact stresses (1.5 GPa). The reason for this seems to originate partially from the linearity of the minimum film thickness - load relationship when plotted on a log-log scale.

INVERSE METHOD

The Reynolds equation (Eq. (1)) is normally regarded as determining the pressure distribution corresponding to a given film shape. In the inverse method adopted by Dowson and Higginson (1959), however, the equation is used to find the film shape responsible for the generation of a given pressure distribution. That is, for a given load, an initial pressure profile (slightly different from Hertzian) is chosen. The film thickness is calculated twice, once by using the elasticity equation and again from the Reynolds equation. An elastohydrodynamic lubrication solution is obtained when the discrepancy between these two film profiles is sufficiently small. If the film shapes are not in agreement, the pressure profile is modified to improve the agreement between them. A flow diagram of the inverse method is shown in Fig. 2.

Some limitations of the inverse method are listed below:

(1) It is not suited for lightly loaded cases where the film shape in the contact region is not parallel.

(2) Although the computational method of Dowson and Higginson (1959) produced an acceptable solution in a small number of cycles, the procedure was not fully automatic. Judgement was needed in effecting the necessary modification of the pressure curve on the basis of the discrepancies between the elastic and inverse Reynolds film shape calculations.

(3) It is only suitable for one-dimensional problems as pointed out by Rohde and Oh (1975). Where more than one spatial variable is involved, the Reynolds equation cannot be so readily integrated to express the film profile as a function of pressure.

QUASI-INVERSE METHOD

With regard to limitation (3) given above, it should be pointed out that Evans and Snidle (1982) have recently used a quasi-inverse method in solving a two-dimensional elastohydrodynamic lubrication problem and have obtained good results for heavily loaded conditions. In contrast to solving an algebraic cubic equation as Dowson and Higginson (1959) did for the one-dimensional problem, Evans and Snidle (1982) have solved the Reynolds equation for the two-dimensional problem as a first-order differential equation in the film shape. They use a direct method in the inlet region, where the pressure in the conjunction is low, and the inverse method elsewhere. Figure 3 shows the boundaries between the computing regions used in the quasi-inverse method. The circle in Fig. 3 represents the corresponding Hertzian area of dry contact. The curve AA is the starting point for the inverse solution of the Reynolds equation. The curve BB, which is downstream of AA, is the downstream boundary of the region over which a direct solution to the Reynolds equation is obtained. The procedure is summarized as follows:

(1) Assume an initial pressure $p_0(x,y)$.

(2) Use elasticity to calculate film shape $h_e(x,y)$.

(3) Use the direct method to calculate pressure $p_1(x,y)$ upstream of BB.

(4) Adjust the film shape constant h_0 until integrated load agrees with input load.

Assume that the pressure downstream of BB initially is $p_0(x,y)$.

(5) Use the inverse hydrodynamic approach to calculate the film shape $h_i(x,y)$ downstream of AA.

(6) Adjust pressure downstream of BB with

$$p_2(x,y) = p_0(x,y)\left[2 - \frac{h_e(x,y)}{h_i(x,y)}\right]$$

(7) Between AA and BB use a weighted mean of $p_1(x,y)$ and $p_2(x,y)$ for the new pressure.

(8) Continue calculations, returning to (2) until little change in pressure occurs throughout the conjunction.

The position of BB in Fig. 3 is located where the viscosity-pressure exponent $\alpha p \approx 1$ or 2. The boundary AA is positioned two or more grid points upstream of BB.

It appears that the Evans and Snidle (1982) approach is able to get converged solutions for Hertzian contact stresses up to 1.5 GPa. This is a threefold increase in stress over the direct method and within the range that nonconformal contacts such as rolling element bearings and gears experience. A limitation of the approach is that a good initial pressure profile is needed since the pressure in the zone between curves AA and BB is constrained to r very slowly by virtue of their very small separation.

NEWTON-RAPHSON METHOD

Both in its direct and inverse forms, the Reynolds equation is nonlinear (except the isoviscous, incompressible direct case), so that the Newton method along with its variations suggests itself as a powerful approach to the problem. Applied to an algebraic equation $f(p) = 0$, the Newton method provides a systematic and usually convergent improvement on an approximation $p^{(n)}$ to one of the roots. The simple algorithm that results can be expressed as $p^{(n+1)} = p^{(n)} - f(p^{(n)})/f'(p^{(n)})$, which shows that the correction is always proportional to the residual error $f(p^{(n)})$. The Newton method for a functional equation likewise produces an algorithm for which the correction to the solution depends on the residual, with the chief difference being that the correction function itself satisfies a differential equation. With suitable boundary conditions this equation supplies convergent corrections to each previous approximation.

Though the Newton-Raphson method is capable of solving the most general form of the Reynolds equation, we limit our demonstration of the method here to the one-dimensional, incompressible case. For this situation equation (1) can be written as

$$f(p) = \frac{\partial}{\partial x}\left(\frac{h^3}{\eta}\frac{\partial p}{\partial x}\right) - 12u\frac{\partial h}{\partial x} = 0 \tag{6}$$

The Newton-Raphson method implies that

$$f(p^{(n)}) + \Delta\frac{df(p^{(n)})}{dp} = 0 \tag{7}$$

where

$$\Delta = p^{(n+1)} - p^{(n)} \tag{8}$$

The index n refers to the n^{th} iterative approximation. But

$$\Delta \frac{df(p^{(n)})}{dp} = \frac{d}{d\varepsilon}\left[f'(p^{(n)} + \varepsilon\Delta)\right]\Big|_{\varepsilon = 0} \qquad (9)$$

By making use of Eqs. (6) and (9), Eq. (7) can be rewritten as

$$0 = \frac{d}{dx}\left(\frac{h^3}{\eta}\frac{dp^{(n)}}{dx}\right) - 12u\frac{dh}{dx} + \frac{d}{dx}\left[-\frac{\alpha\,\Delta h^3}{\eta}\frac{dp^{(n)}}{dx} + \frac{h^3}{\eta}\frac{d\Delta}{dx} + \frac{3h^2\delta(\Delta)}{\eta}\frac{dp^{(n)}}{dx}\right] - 12u\frac{d\delta(\Delta)}{dx} \qquad (10)$$

In Eq. (10) $\delta(\Delta)$ is the elastic deformation due to the pressure difference between the present and previous cycles at a given location.

Figure 4 is a flow diagram showing how the Newton-Raphson method applies to the elastohydrodynamic lubrication problem. There are three main loops. The innermost loop solves for the pressure difference $\Delta^{(n+1)}$ from Eq. (10) while the viscosity and film thickness at any x location are held constant. Within this loop the elastic deformation $\delta(\Delta)$ due to the pressure difference Δ is continuously updated. With this converged solution on the pressure difference in the inner loop, the new pressure throughout the conjunction can be evaluated, and the viscosity and film shape are recalculated in the second loop. This loop is continued until there is little change in the pressure throughout the conjunction. The final loop requires that the load obtained from integrating the pressure is in agreement with the inputted load.

Rohde and Oh (1975) applied higher order finite elements (cubic splines, cubic Hermite polynomials) with the Newton-Raphson iteration to solve lightly and moderately loaded cases (maximum Hertzian contact stress less than 0.8 GPa). Numerical instabilities were found for larger maximum Hertzian contact stresses.

CONVERGENCE AND STABILITY

Hitherto several approaches to the elastohydrodynamic lubrication problem have been considered, each of which finds application in some regime of the operating parameters contained in the model equations. However, since none of the methods produces a solution (i.e., a compatible pressure distribution and film shape) in closed mathematical form, questions of convergence and stability of the various results assume considerable importance. Moreover, in arriving at an acceptable solution, it should not be overlooked that while uniqueness of solutions is generally assumed, it has never been rigorously demonstrated in the case of the particular nonlinearities presented by the elastohydrodynamic lubrication problem. The rather weak and transparent basis for this assumption is simply that, as a matter of experience, only one numerical solution is ever found for any specified set of operating parameters.

Linear Analysis

In view of the complications introduced by nonlinearity, combined with the great variety of numerical approximation schemes within the general cate-

gories of finite element and finite difference, we do not attempt the most general analysis of convergence or stability here. Instead, we examine the behavior of a linearized method, such as the Newton-Raphson method, which can be suitably reduced to algebraic form to be solved by an iterative scheme. Some comments are also included on the stability of the pressure distribution under conditions such that several extrema, including a spike, may occur in the conjunction.

Consider for example the innermost iterative loop of the Newton method shown in the flow diagram of Fig. 4. The differential equation (Eq. (10)) for the pressure correction is linear in both Δ and δ, and as implemented in Fig. 4 its solution resembles that of time march methods in the transient problem with iteration count replacing time advance. An eigenvalue analysis of the convergence of such a loop is therefore appropriate. An equivalent linearized approach using an explicitly time-dependent perturbation was adopted by Kostreva (1983), who was able to construct a stability map in the space of two nondimensional parameters. Curiously, however, most of the existing elastohydrodynamic solutions lie well within the unstable region of this map.

Each term of Eq. (10) belongs to one of three types: (1) linear in Δ, (2) linear in δ, or (3) independent of both Δ and δ. In the algebraic reduction of the derivatives involved for types (1) and (2), their contributions to Eq. (10) can be written in matrix form, respectively, as $T_{ij}\Delta_j$ and $S_{ij}\delta_j$, where the summation convention is used. The matrices T and S will usually be banded with elements given by the initial (or current) approximation to pressure p and film shape h. Single-subscripted quantities are vectors whose elements are the nodal values on the chosen discrete mesh. Type (3) terms are nonzero only when the pressure distribution fails to satisfy the Reynolds equation and hence are known as the vector of the residuals R_i. In this representation Eq. (10) becomes

$$T_{ij}\Delta_j + S_{ij}\delta_j - R_i = 0 \tag{11}$$

If T is nonsingular, the direct solution of Eq. (11) is

$$\Delta = A + B\delta \tag{12}$$

where $A = T^{-1}R$ and $B = -T^{-1}S$. Since the elastic displacement vector δ is not initially known, Δ is computed iteratively according to the prescription

$$\Delta^{(n+1)} = A + B\delta^{(n)} \tag{13}$$

Similarly for the inverse method the solution of Eq. (11) appears as follows:

$$\delta^{(n+1)} = A' + B'\Delta^{(n)} \tag{13'}$$

where A' and B' require the nonsingularity of S.

To complete the iteration loop, the elastic equation relating Δ and δ is needed. The elasticity equation such as Eq. (4) can be expressed as

$$\delta^{(n)} = D\Delta^{(n)} \tag{14}$$

which is not iterative. The solution of Eq. (13), making use of Eq. (14), is thus

$$\Delta^{(n)} = \sum_{m=0}^{n-1} (BD)^m A \tag{15}$$

which shows explicitly the proportionality of Δ to the modified residual $A \equiv T^{-1}R$. The limiting value of $\Delta^{(n)}$ becomes

$$\Delta = \lim_{n\to\infty} \Delta^{(n)} = \lim_{n\to\infty} \sum_{m=0}^{n-1} (BD)^m A$$

or

$$\Delta = (1 - BD)^{-1} A \tag{16}$$

The intermediate estimates $\Delta^{(n)}$ are just partial sums in the expansion of the inverse operator on the right side of Eq. (16).

Similarly for the inverse method, making use of Eqs. (13') and (14) gives

$$\delta = (1 - B'D^{-1})^{-1} A' \tag{16'}$$

Because of the ill condition of D, this segment of the inverse elastohydrodynamic lubrication problem is never actually solved this way. Consequently the inverse method discussed earlier ends up with two film shapes for each $\Delta^{(n)}$ instead of generating the usual pressure $\rightleftarrows$ film shape loop.

From Eq. (16) it is easy to see how the inner loop of Fig. 4 behaves. Let the eigenvector decomposition of the residual vector A be written as

$$A = \sum_{\alpha} A_\alpha X_\alpha \tag{17}$$

where A_α is the (scalar) coefficient of the α^{th} eigenvector X_α of the product matrix BD corresponding to eigenvalue λ_α. Inserting Eq. (17) into Eq. (15) and summing the series yield

$$\Delta^{(n)} = \sum_{\alpha} A_\alpha \left(\frac{1 - \lambda_\alpha^n}{1 - \lambda_\alpha}\right) X_\alpha \tag{18}$$

which converges if $|\lambda_\alpha| < 1$ for all α. When this holds, Eqs. (16) and (18) both lead to

$$\Delta = \sum_{\alpha} A_\alpha (1 - \lambda_\alpha)^{-1} X_\alpha \tag{19}$$

This converged limit is quickly attained if all $|\lambda_\alpha| << 1$. The most important contributions to the sum are unfortunately the slowest to converge, and it may thus be useful to test for $\max |\lambda_\alpha|$. When this is close to 1, and particularly if it exceeds 1, the number of inner-loop cycles should be curtailed. Control is then passed to the second iteration loop, where the matrix BD and the residual vector A are recomputed from the new pressure and its film shape.

This behavior of the inner loop of the Newton-Raphson method remains essentially unaltered by the introduction of a damping factor f. If the loop is reentered with the weighted pressure distribution $(1 - f)\Delta^{(n)} + f\Delta^{(n-1)}$, there is no change in the limiting iterate of Eq. (16). Only the intermediate steps by which this is reached are affected. It is worth noting, however, that any smallness criterion for the pressure changes used in determining an exit from the inner loop should, for consistency, be multiplied by (1 - f).

Nonlinear System Response

Useful as this analysis can be in carrying out the first iteration, it leaves unanswered the original question of overall convergence of the method. Specifically, it does not address the modification in the eigenvalues of BD when the middle iteration loop updates the pressure. Some limited appreciation of the stability at this stage can be obtained by regarding this loop as an amplifier whose input $p^{(n)}$ produces output $p^{(n+1)}$. The parameters of the amplifier are dependent on the input, reflecting the nonlinearity or feedback aspects of elastohydrodynamic lubrication, and the objective is to achieve unit gain at any frequency by adjusting the parameters (matrix elements). Since the components of p are nodal pressures, frequency here refers to spatial oscillation in the solution vector.

For illustrative purposes the treatment can be reduced to its simplest form by considering the same one-dimensional incompressible flow problem as in Eq. (6). The Reynolds equation in this situation can be integrated once to give

$$\frac{dp}{dx} = 12u\eta \frac{h - h_m}{h^3} \tag{20}$$

A typical solution of this equation is displayed in Fig. 5, which shows that over a length AB comparable to the Hertzian diameter the elastohydrodynamic lubrication film thickness may be substantially constant. The pressure and hence also the viscosity are large in this parallel film region, so that characteristics of the pressure profile are highly sensitive to the small difference $(h - h_m)$, where h_m is the film thickness at the pressure extrema.

By superimposing a low-amplitude pressure ripple on the pressure input to the middle loop, the output ripple can be calculated with the aid of Eq. (20) and the one-dimensional counterpart of Eq. (4). The output/input amplitude ratio or gain g in a region of almost constant film thickness is then found to be given by the approximate proportion

$$g \propto \frac{u\eta_0 e^{\alpha p}}{h^3 K^2} \qquad (21)$$

where K is the wave number (reciprocal of wavelength) of the ripple. The constant of this approximate proportionality depends on the contact geometry.

Equation (21) shows qualitatively how the middle iteration loop is expected to behave in the region AB of Fig. 5, where the pressure gradient is still small and where the pressure extrema may occur. For given u, η_0, and α values, h is essentially determined independently of load so that interest focuses on the p and K dependence of g. The larger g grows, the greater becomes the amplification of any ripple, with consequent appearance of oscillations in the pressure distribution. The equation confirms the general experience that such are associated with high pressures, particularly on the upstream side of the spike in the vicinity of B. It further suggests that oscillation becomes of less concern at short wavelengths. Since the smallest wavelengths involved are of the order of the mesh size, numerical noise introduced by the algebraic representation of the basic equations or by finite computer accuracy is well attenuated. Greater difficulty is encountered at wavelengths comparable to the length of the parallel film region, and this ultimately may lead to distortion of much of the pressure curve there.

Although the validity of Eq. (21) breaks down for still larger wavelengths, it is clear that the system response to such long-wavelength disturbance is given by its response to changes in h_0, the additive constant in the film shape Eq. (5). This behavior is controlled separately in the third iteration loop of Fig. 4 and as such is not treated by present considerations.

SUPERCOMPUTERS USED IN EHL

The computers used to solve elastohydrodynamic lubrication problems in the 1950's were either mechanical or the first of the electronic computers and were considerably different from the supercomputers available today. The present-day powerful computing machines known as supercomputers have peak computing speeds exceeding 100 million operations per second. This is to be contrasted with the first commercial electronic computer delivered in 1951, which had a peak computing speed of around 600 operations per second. This means that the speed of large-scale scientific computers has doubled, on the average, every 2 years. Although the current performance levels of such machines owe much to the rapid advance of microelectronics, new concepts in computer architecture have been equally important. The term 'architecture' refers to the logical organization of the computer as it is seen by the programmer. The architectural innovations of greatest significance are those that enable the machine to carry out many similar operations in parallel, or concurrently. This is referred to as vector computation. The latest supercomputers allow the programmer to prescribe many different elementary steps to be executed simultaneously, whereas earlier computers could only handle a single sequence of elementary steps executed one at a time. This is referred to as scalar computation. Levine (1982) gives a thorough and useful description of supercomputers.

The most recent supercomputers provide a quantum jump in the speed of computation, and their effect in solving elastohydrodynamic lubrication prob-

lems in the next decade should be significant. Hamrock (1982) demonstrated the effect of using such computers in performing elastohydrodynamic film shape calculations. Both scalar and vector versions of the film shape code were presented. The film shape calculations were chosen since in elastohydrodynamic lubrication solutions over 80 percent of the computation time is spent in the straightforward calculation of elastic deformation.

Before proceeding it might be well to point out more specifically what is meant by scalar and vector coding. Scalar coding is conventional sequential coding that has been used universally for digital computers since their earliest development. On the other hand, the process of writing programs in vector code consists of organizing programs so that the maximum possible number of arithmetic and logical operations can be processed in parallel. Such 'vectorization' is most easily introduced and illustrated by an example.

Consider arrays A and B, each consisting of 100 numbers, which are to be added to form array C, where $C_i = A_i + B_i$, $i = 1, \ldots, 100$. The traditional 'scalar' computer executes five assembly language instructions each 100 times. There are two reads from memory (A_i and B_i), one addition, one store to memory (for C_i), and an instruction that increments a counter, tests, and branches back to load the next pair of input operands. Thus 500 scalar instructions are executed in consecutive sequence to add arrays A and B. A vector computer's compiler can generate 'vector' object code, which executes very differently. The vector code for adding the 100 pairs of operands still uses the first four of these instructions for each pair. But instead of continually branching back for the next pair, execution proceeds by continuously streaming operands from central memory into the central processor, where the addition takes place, and continuously streaming answers back into memory in 'pipeline' fashion. During execution of the vectorized addition some elements of A and B are being read from central memory, some elements of A and B are undergoing addition in the vector pipeline, and some are being written in memory.

This brief example illustrates the major conceptual difference between the two modes of operation. It should be clear from this that many kinds of repeated instructions can be processed in vector form, with consequent enormous savings in overall time. It is also apparent that there are other situations in which the pipeline mode will not operate. Perhaps the most common of these in elastohydrodynamic lubrication problems arise when a recursive or iterative process requires the i^{th} result before it can generate the $(i + 1)^{th}$, as in direct integration of a differential equation. Often a way may be found to avoid such conditions, but even when this is not possible and the computer is forced into the scalar mode, its operation speed is still considerably superior to previous generation computers.

Table I shows the scalar and vector computation times expressed in nanoseconds for varying vector 'length', which here corresponds to the number of nodes in the finite difference representation of the elastohydrodynamic lubrication solutions. A considerable decrease in computation time is seen for the vector computation over that of the scalar computation. The last column of the table gives the ratio of scalar to vector times and indicates that, as the vector length increases, the ratio also increases significantly, although eventually an asymptotic condition is approached. The run times for the two types of code indicate that over a 50-to-1 speedup of scalar to vector compu-

tation time is obtained for vector lengths typically used in elastohydrodynamic lubrication analysis.

CONCLUSIONS AND OUTLOOK

The aim herein has been to examine the principal approaches devised to tackle the technologically important elastohydrodynamic lubrication problem. Without entering into details of numerical approximation schemes applicable to the nonlinear equations involved, it is apparent that under the general headings of finite difference and finite element, sufficiently powerful methods exist to produce some useful solutions to the problem. Executing such schemes by a multigrid approach could lead to further substantial improvements over the conventional fixed-mesh methods most widely used at present. The availability of supercomputers combining high speed and large memory capacity removes most obstacles to implementing the chosen scheme, so that the efficiency or convergence rate of the method is no longer the overriding concern. The question of central importance has now become the stability of the solution method.

Experience with finite element techniques in elastohydrodynamic lubrication applications is accumulating, as evidenced by the two sessions of this conference devoted to them, and it is likely that such techniques could have some advantages over the currently more familiar finite difference methods. Thus, it often appears that significantly fewer nodal values are required in the finite element representation than for finite differences. Nevertheless the strong impetus behind finite element analysis, namely, its suitability for problems of little or no symmetry, as, for example, in classical elastostatics, is lacking in elastohydrodynamic lubrication, where nominal bearing geometry is often required to be highly symmetrical.

Practical experience indicates an advantage to using an inverse method at higher maximum pressures and conversely a direct method at lower pressures. The linearized Newton-Raphson method of improving solutions to functional equations runs in either the direct or inverse mode according to which approach to the hydrodynamic equation is taken. So far, no serious attempt has been made to invert the elasticity equation, but Saint-Venant's principle strongly suggests that such an approach would be hopeless.

Turning to questions of stability, again the issue is not between finite difference and finite element. Both approaches can be analyzed in the same manner. A beginning at such analysis has been made by Kostreva (1983), whose stability map uses a criterion based on the eigenvalue behavior of a certain linearized form of the Reynolds equation.

A feature shared by many elastohydrodynamic lubrication calculations has been that their implementation by computer is not fully automatic. The solution evolves stepwise from the initial guess, and the criterion for passing to the next step (iteration) requires some judgement or a further guess not easily reduced to a sequence of simple branching decisions. As such the pathway to a final solution has been neither completely described nor properly optimized and the nonquantifiable skills of the programmer have played a prominent role. To exploit more fully the advantages of automatic computation, it is thus necessary to develop improved understanding of the convergence properties of the various steps and the ability to decide if the current step is actually

moving toward or away from a solution. Since several constraints are imposed which it is difficult to tighten simultaneously, the process at intermediate stages may seem to be diverging from the viewpoint of some single constraint. It is common, for example, to find a pressure profile making improvements in the film shape while diverging from the required load, to which adjustments are made only at a later stage. An optimized strategy might handle the load at the same time. In brief, criteria are needed to enable control to pass automatically between the iterative loops.

It was to this end that the analysis of the behavior of one type of iteration loop was undertaken here. A criterion for convergence in terms of eigenvalues of the operator representing the net effect of the loop was derived. Although it still remains to develop this into concise executable form, progress toward this goal is being made and further effort in this direction is certainly worthwhile.

Similarly, the treatment given here of the much more difficult stability problem associated with the high-pressure region of the contact is a beginning attempt to understand the conditions for which reliable solutions can be generated there. Further work, however, is needed to improve the analysis as the spike region itself is entered, where the pressure gradient grows large.

If the potential of such approaches to program automation is fulfilled, then prospects are excellent for considering additional aspects of the elastohydrodynamic lubrication problem, such as

(1) Non-Newtonian effects of the fluid
(2) Surface roughness effects
(3) Thermal effects

With the use of supercomputers for the incorporation of these coupled effects it can be predicted that run times will be dramatically reduced during this decade. Such investigations will then become much more feasible than in the past, with a corresponding gain in the reliability and range of application of the results.

REFERENCES

Dalmaz, G.; and Godet, M. (1973): Traction, Load, and Film Thickness in Lightly Loaded Lubricated Point Contacts, J. Mech. Eng. Sci., vol. 15, no. 6, pp. 400-409.

Dowson, D.; and Higginson, G. R. (1959): A Numerical Solution to the Elastohydrodynamic Problem, J. Mech. Eng. Sci., vol. 1, no. 1, pp. 6-15.

Evans, H. P.; and Snidle, R. W. (1982): The Elastohydrodynamic Lubrication of Point Contacts of Heavy Loads, Proc. R. Soc. Lond., vol. A382, pp. 183-199.

Hamrock, B. J. (1982): Film Shape Calculations on Supercomputers, Advances in Computer-Aided Bearing Design, C. M. Chang and F. E. Kennedy, eds., The American Society of Mechanical Engineers, pp. 67-74.

Hamrock, B. J.; and Dowson, D. (1976): Isothermal Elastohydrodynamic Lubrication of Point Contacts, Part I - Theoretical Formulation, J. Lubr. Technol., vol. 98, no. 2, pp. 223-229.

Hamrock, B. J.; and Dowson, D. (1981): Ball Bearing Lubrication - The Elastohydrodynamics of Elliptical Contacts, John Wiley & Sons, New York.

Kostreva, M. M. (1983): Pressure Spikes and Stability Considerations in Elastohydrodynamic Lubrication Models. To be presented at the ASME/ASLE Lubrication Meeting in Hartford, Connecticut, Oct. 18-20, 1983, and later published in the J. Lubr. Technol., ASME.

Koye, K. A.; and Winer, W. O. (1980): An Experimental Evaluation of the Hamrock and Dowson Minimum Film Thickness Equation for Fully Flooded EHD Point Contacts, J. Lubr. Technol., vol. 103, no. 2, pp. 284-294.

Kunz, R. K.; and Winer, W. O. (1977): Discussion on pp. 275-276 to Hamrock, B. J. and Dowson, D. Isothermal Elastohydrodynamic Lubrication of Point Contacts, Part III - Fully Flooded Results, J. Lubr. Technol., vol. 99, no. 2, pp. 264-275.

Levine, R. D. (1982): Supercomputers, Sci. Am., vol. 46, no. 1, pp. 118-135.

Rohde, S. M.; and Oh, K. P. (1975): A Unified Treatment of Thick and Thin Film Elastohydrodynamic Problems by Using Higher Order Element Methods, Proc. R. Soc. Lond., vol. A343, pp. 315-331.

TABLE I. - COMPARISON OF SCALAR AND VECTOR COMPUTER RUN TIMES FOR DIFFERENT VECTOR LENGTHS

Vector length, NX	Scalar time, nsec	Vector time, nsec	Scalar time / Vector time
10	241	70	3.44
30	1 179	155	7.61
50	2 726	243	11.22
100	9 252	483	19.28
300	73 357	1 933	37.95
500	198 282	4 188	47.35
700	383 997	7 242	53.02
1 000	776 588	13 363	58.11
3 000	6 890 354	99 988	68.91
5 000	19 084 530	266 602	71.58
7 000	37 358 446	513 218	72.79
10 000	76 171 860	1 033 169	73.73

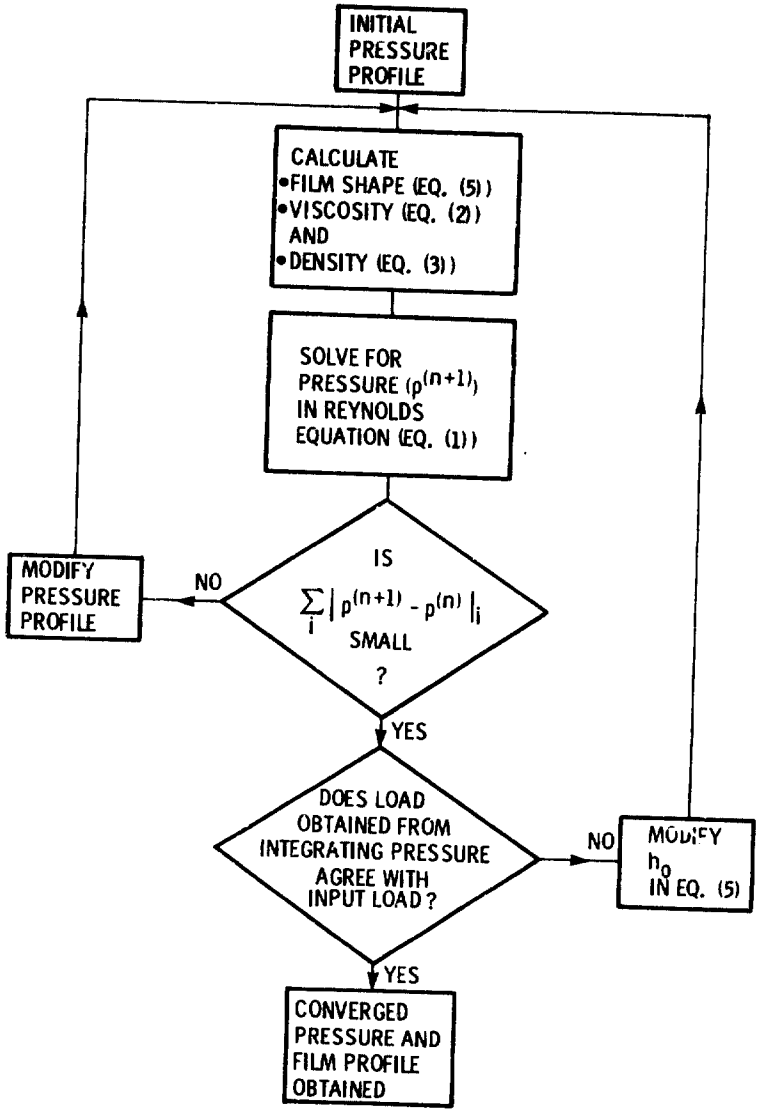

Figure 1. - Flow diagram of direct method.

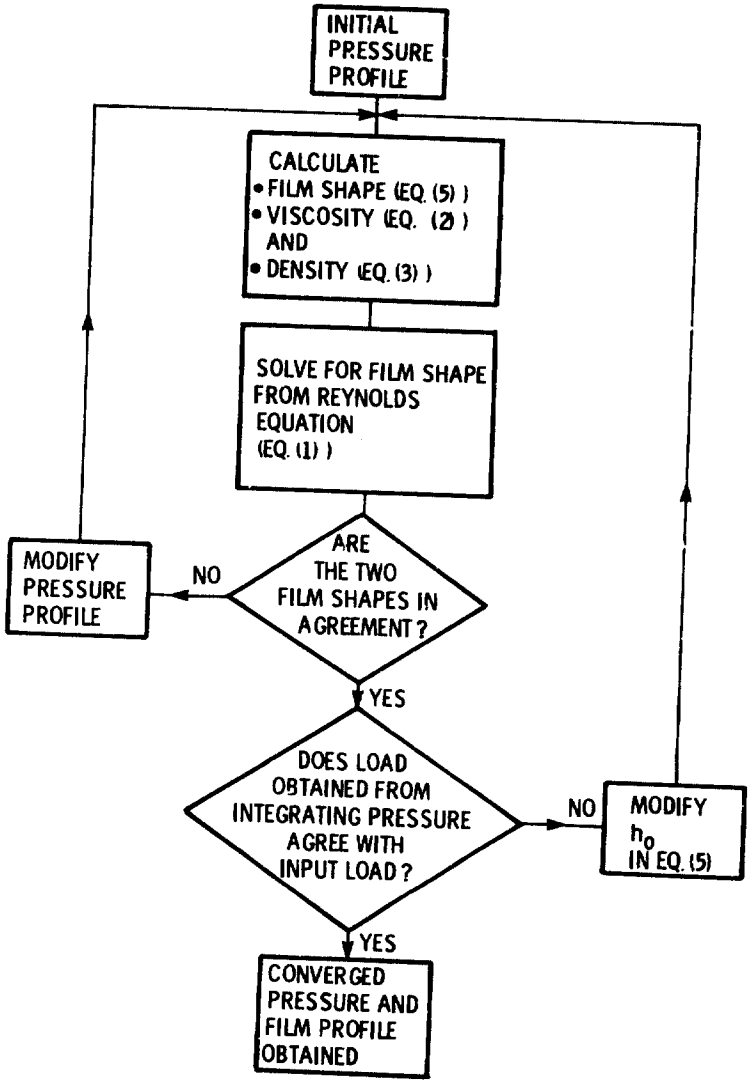

Figure 2 - Flow diagram of inverse method.

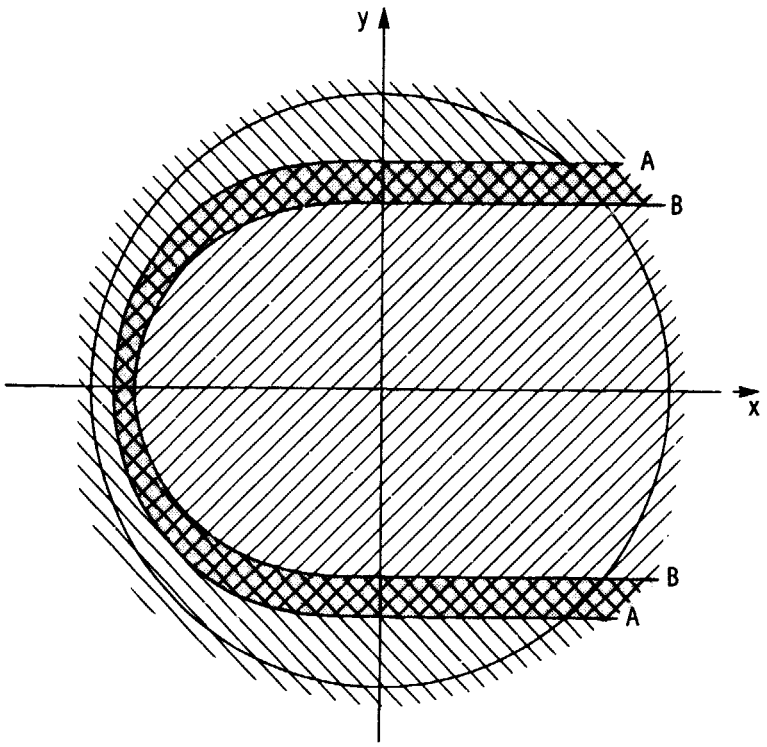

Figure 3. - Contact area showing boundaries between computing regions used in quasi-inverse method. (From Evans and Snidle (1982))

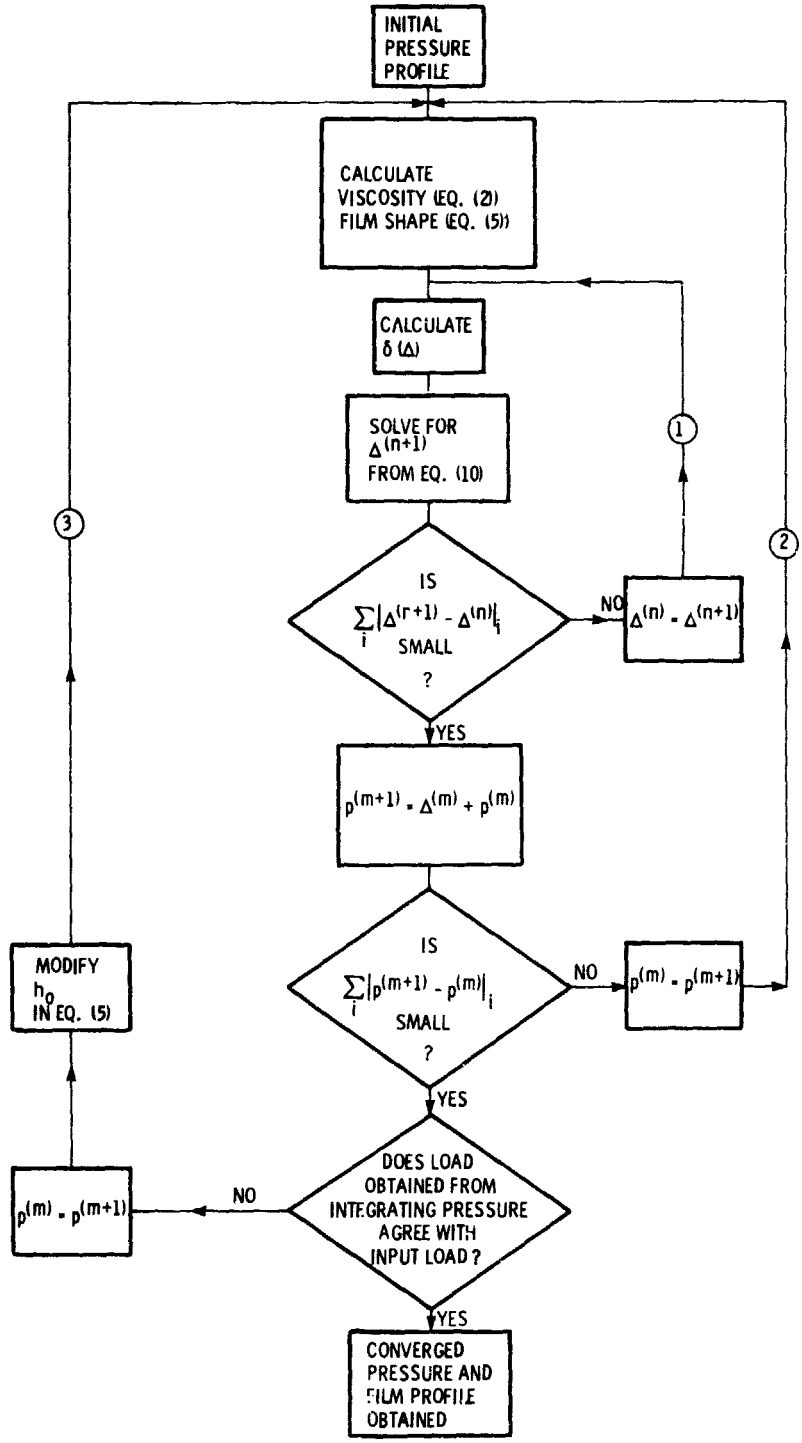

Figure 4. - Flow diagram of Newton-Raphson method.

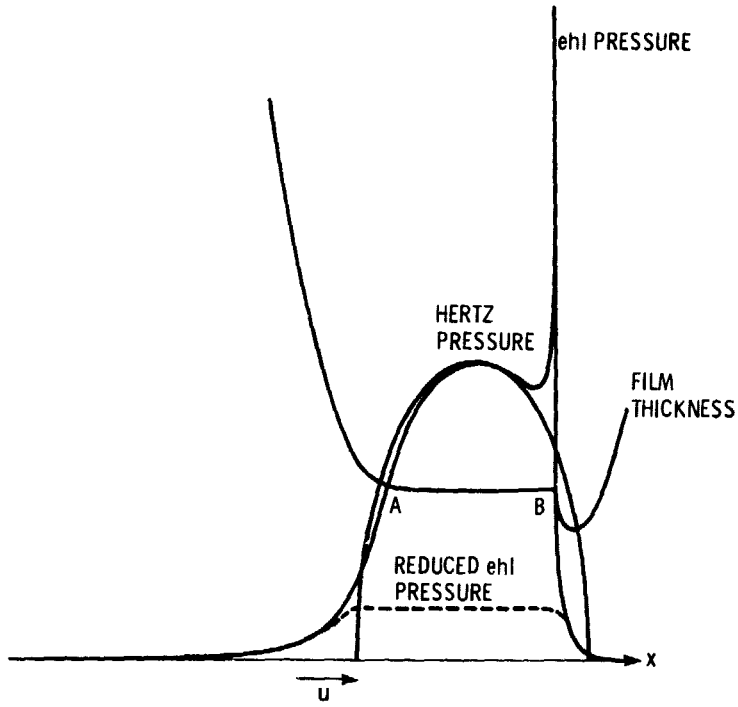

Figure 5. - Example of an ehl pressure distribution and film profile. The operating parameters have been chosen to produce a minimum in the pressure curve between the Hertzian maximum and the spike near the exit. Between points A and B, the film thickness is nearly constant - as a consequence of the similarity of the pressure to the Hertzian distribution, shown for comparison.