INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS, VOL. 4, 377–397 (1984)

ELASTO-HYDRODYNAMIC LUBRICATION: A NON-LINEAR COMPLEMENTARITY PROBLEM

MICHAEL M. KOSTREVA

Senior Member Technical Staff, Statistics & Operations Research Dept., GTE Laboratories, Inc., 40 Sylvan Road, Waltham, MA 02254, U.S.A.

SUMMARY

The classical problem of elasto-hydrodynamic lubrication of cylinders in line contact is formulated as a non-linear complementarity problem. A direct algorithm is applied to the approximation obtained by finite differences. Implementation considerations are emphasized. The new method provides reliable and automatic location of the previously troublesome lubricant free boundary. Numerical results reveal the qualitative behaviour of the pressure distribution and the lubricant film thickness under variation of key physical parameters.

KEY WORDS Lubrication Film Thickness Free Boundary

1. INTRODUCTION

The elasto-hydrodynamic lubrication of cylinders in line contact has received much attention recently in the mechanical engineering literature. Among theoreticians, two main routes of study are distinguished. Following Grubin,[1] many studies made simplifying assumptions about the mathematical model to obtain approximate or partial solutions which may be valid over some restricted domain of interest. Paralleling this work have been attempts to solve the general mathematical model numerically to calculate the pressure distribution in the lubricant and simultaneously the lubricant film thickness. Computer solutions, requiring fewer simplifying assumptions, are more widely applicable. However, there are subtle complicating features present in the mathematical model which have caused computational difficulties to earlier computer approaches. The purpose of this paper is to present a numerical algorithm which is easy to implement and which overcomes these difficulties.

Dowson and Higginson[1,2] approached the computer solution of the EHL model with an inverse iterative technique, after they found straightforward (fixed-point) interation tedious and slowly converging, even for a constant viscosity lubricant. However, they supplied no convergence proof and in some extreme cases the convergence of inverse iteration is doubtful. Stephenson and Osterle[3] obtained solutions for lightly loaded cases using finite differences and an inside–outside iteration coupled with bisection to find the outlet point. Attempts to prove the convergence of their scheme were unsuccessful because of the high degree of non-linearity of the equations and the fact that the outlet point varies from iteration to iteration. Cheng and Sternlicht,[4] treating the heavily loaded case, sought solutions having a sharp pressure peak at a prescribed value and adjusted the speed (for given load and material properties) to match. This approach overlooked the possibility that for some load and material properties there would be no such speed.

Herrebrugh[5] reported that an integral equation approach was successful for the case of

0271–2091/84/040377–21$02.10

Received 1 August 1982
Revised 31 January 1983

constant viscosity. The usefulness of this integral equation approach for the non-constant viscosity case has not been explored.

Some of the most recent computational work has concerned the application of finite element techniques. Taylor and O'Callaghan[6] employed a Galerkin approach with isoparametric elements. Rohde and Oh[7] applied higher order elements (cubic splines, cubic Hermite polynomials) along with Newton–Raphson iteration to solve lightly and moderately loaded cases. However, some oscillation difficulties were noted by Rohde and Oh in the heavily loaded cases using the piecewise cubic Hermite functions.

These two teams of researchers also commented on the location of the free boundary (outlet point) and proposed heuristics to locate the free boundary. Taylor and O'Callaghan[8] applied their techniques to some 'soft' elasto-hydrodynamic lubrication problems (elastic cylinder rolling on a rigid half-plane) treated by Swales, Dowson and Latham.[9] Taylor and O'Callaghan reported some difficulties in locating the free boundary. In particular, extreme case was required in the positioning of the outlet point to avoid solution oscillation and/or divergence.

This brief historical survey indicates a need for a systematic approach to solution—one which can be readily automated, can be proved mathematically to converge and which includes the automatic location of the free boundary.

In the related areas of hydrodynamic lubrication of journal bearings[10–12] and stress analysis of gear systems,[13] computational techniques borrowed from the area of quadratic programming have proved very useful. In these linear problems, solutions and free boundary locations are computed either with 'complementary pivot' algorithms, which use pivot operations similar to the simplex method of linear programming or with a variant of the relaxation method for linear systems of equations. In this paper, the direct algorithm of Habetler and Kostreva[14] will be applied to the non-linear integro-differential problem at hand.

The relaxation method of Christopherson,[10] first derived for the hydrodynamic lubrication problem, is sometimes proposed for use in the elasto-hydrodynamic lubrication problem. However, the criterion for convergence of the method[12] (symmetric, positive definite Jacobian matrix with non-positive off-diagonal elements), does not apply here, even to a linearized version of the problem. Thus a method which converges on a larger class of problems is of interest. To demonstrate its computational feasibility, the method of Reference 14 has been implemented in a FORTRAN program and over a hundred cases have been solved. These cases range over a variety of operating conditions including lubricants of constant and variable viscosity, under lightly, moderately, and heavily loaded conditions. No difficulties have been experienced with respect to the location of the free boundary using the Habetler–Kostreva direct algorithm. Examples demonstrating the capabilities of the method will be included in a later section.

The direct complementarity approach presented in this paper offers many advantages to the computational lubrication engineer. Since it locates the free boundary or outlet point automatically, the user need not be burdened by time consuming trial-and-error guesswork. A modest computer program implements the algorithm, requiring only readily available software. Among the many computer approaches to EHL developed to date, the direct complementarity algorithm is the only one which has been proved mathematically to converge to a solution in a finite number of steps.

The paper is organized as follows. The formulation of the mathematical model is covered in Section 2. Following this is a section describing a new formulation as a complementarity problem and the mathematics needed to solve it. Some important algorithmic implementation considerations are followed by a section on numerical results.

2. MODEL FORMULATION

The mathematical model we study is not much different from those treated earlier by other authors.[1-3,6,7] However, for clarity and completeness we note the simplifications and assumptions inherent in our model.

We consider the case of pure rolling (no sliding) under isothermal conditions. Only steady state solutions will be considered. The lubricant between the two elastic cylinders in line contact will be considered incompressible. It will be assumed that lubricant viscosity reaches its equilibrium value instantaneously according to the pressure dependent law $\mu = \mu_0 e^{\alpha p}$, where μ is viscosity, p is pressure and α and μ_0 are lubricant dependent parameters. The elastic deformation of the cylinders will be assumed to take place instantaneously with small enough magnitude to be described by linear elasticity theory.

Lubricant pressure will be assumed constant along the direction parallel to the axes of the cylinders as well as across the lubricant film gap. Side leakage will be neglected. As usual, the two cylinders of the physical model are analysed as a single equivalent cylinder rolling along a plane.

Any load applied externally will be balanced by the pressure generated in the lubricant film. Lubricant film thickness will be assumed positive over the entire domain of interest.

Between the inlet point denoted by x_a and the outlet point $x_b (x_b > x_a)$, the Reynolds' equation quantifies the relationship between pressure, film thickness and viscosity. The inlet point x_a will be taken as known, whereas the outlet point (free boundary) x_b will be determined as part of the solution.

The Reynolds' equation, the equation of linear elasticity, the pressure viscosity equation and the load constraint are non-dimensionalized (see appendix for notation) by $x = \bar{x}/a$, $h = \bar{h}/\delta$ and $p = \bar{p}/p_{max}$ relative to the Hertzian, dry solution. We also introduce the aggregate variable k which is the dimensionless equivalent of the sum of the terms (h_0 + constant) in the formulation of Dowson and Higginson (Reference 1, p. 66). This leads to the following free boundary problem for the pressure distribution $p(x)$ and the film thickness $h(x)$.

Given the parameters α and λ (see appendix for definitions) and an inlet point x_a, find $p(x)$, $h(x)$, the free boundary x_b and the aggregate variable k satisfying:

$$\frac{d}{dx}\left(\frac{h^3(x)}{e^{\alpha p}}\frac{dp}{dx}\right) = \lambda \frac{dh}{dx} \quad \text{in} \quad [x_a, x_b] \tag{1}$$

$$h(x) = x^2 + k - \frac{2}{\pi}\int_{x_a}^{x_b} p(s) \ln|x - s|\, ds \quad \text{in} \quad [x_a, \infty) \tag{2}$$

$$\frac{2}{\pi}\int_{x_a}^{x_b} p(s)\, ds = 1 \tag{3}$$

with the boundary conditions

$$p(x_a) = 0 \quad \text{and} \quad p(x_b) = 0 = \frac{dp}{dx}(x_b) \tag{4}$$

Let us take note here of a few characteristics of the model which distinguish it from some related models in lubrication theory. Observe that the formulation contains non-local coefficients because $h(x)$ is an integral function of $p(x)$. Non-local coefficients preclude the use of so called 'shooting methods' useful in many iterative boundary-value problem approaches. Another effect of the non-local coefficients is that the Jacobian matrices necessary in any discretization are not banded. Increased computational complexity of the non-linear equation solving is a direct result of the non-local coefficients.

The kernel in the integral equation which gives $h(x)$ is mildly singular. As long as the Cauchy Principal Value of the integral converges, no difficulties arise with the formulation. However, the singularity may be removed using integration by parts:

$$h(x) = x^2 + (k+1) + \frac{2}{\pi}\int_{x_a}^{x_b} (s-x)\ln|s-x|\left(\frac{dp}{ds}\right)ds$$

This form is recommended for computational purposes.

In order that the computed solutions be physically meaningful, we seek only solutions $p(x)$ which satisfy $p(x) \geqslant 0$ on $[x_a, x_b]$ and $p(x) = 0$ on $[x_b, \infty)$. An approach which enables one to obtain such solutions is complementarity theory. The next section introduces complementarity and reviews the relevant theoretical considerations. A non-linear complementarity problem corresponding to the mathematical model (1)–(4) is then presented and solved with a numerical algorithm.

3. COMPLEMENTARITY

Motivating the complementarity theory applicable to the solution of the elasto-hydrodynamic lubrication problem is the following constrained free boundary problem, corresponding to the model (1)–(4) above.

Given x_a, the inlet point, let x_F denote some point far down stream, so that $x_F > x_b > x_a$. Then if α and λ are given, find $p(x)$ and $h(x)$, x_b and k satisfying:

$$R(p,h,k) = -\frac{d}{dx}\left(\frac{h^3(x)}{e^{\alpha p(x)}}\frac{dp}{dx}\right) + \lambda\frac{dh}{dx} = 0 \quad \text{in} \quad [x_a, x_b]$$

$$R(p,h,k) \geqslant 0 \quad \text{in} \quad [x_b, x_F]$$

$$S(p,h,k) = h(x) - x^2 - k + \frac{2}{\pi}\int_{x_a}^{x_b} p(s)\ln|x-s|\,ds = 0 \quad \text{in} \quad [x_a, x_F]$$

$$T(p,h,k) = 1 - \frac{2}{\pi}\int_{x_a}^{x_b} p(s)\,ds = 0$$

subject to:

$$p(x) \geqslant 0 \quad \text{in} \quad [x_a, x_b]$$
$$p(x) = 0 \quad \text{in} \quad [x_b, x_F]$$
$$h(x) > 0 \quad \text{in} \quad [x_a, x_F]$$
$$k \text{ unrestricted in sign}$$

together with the boundary conditions

$$p(x_a) = 0 \qquad \text{and} \qquad p(x_b) = 0 = p'(x_b)$$

The non-negativity conditions of the above problem are well motivated. That $p(x)$ should be a non-negative function is usually accepted. However, the interpretation of a non-negative Reynolds' operator $R(p, h, k)$ has not been given. Geometrically, this means that in the interval $[x_b, x_F]$, downstream of the cavitation point, the surfaces are *diverging* $\left(\lambda\frac{dh}{dx} \geqslant 0\right)$ and film thickness is an increasing function. This occurs because in $[x_b, x_F]$ the function $p(x)$ is identically zero, which means that $R(p, h, k)$ consists of only one term $\left(\lambda\frac{dh}{dx}\right)$, which has a

positive coefficient λ. Surfaces which are not diverging in the outlet region $[x_b, x_F]$ are not physically meaningful.

The area of mathematical programming known as complementarity theory is a relatively new one and one which has received much study in recent years. A survey paper by Lemke[15] outlines the early theoretical results, most of which were motivated by applications to equilibrium type problems in optimization and game theory. More recent results, relevant to the elasto-hydrodynamic lubrication problem, are contained in the paper of Habetler and Kostreva[14] and the references therein.

Proceeding now to the complementarity theory applicable to solution of the above free boundary problem of elasto-hydrodynamic lubrication we begin with some definitions. Since the free boundary problem requires functions $p(x)$ and $h(x)$ for its solution, the first complementarity problem considered will be in a function space.

Let B be a reflexive, real Banach space with dual space B^*. Let the value of $u \in B^*$ at $v \in B$ be denoted by $\langle v, u \rangle$. Let C be a closed convex cone in B with vertex at 0 and polar cone $C^* = \{u \in B^* \mid \langle v, u \rangle \geq 0 \text{ for each } v \in C\}$. Let $A : D \subset B \to B^*$ be an operator. Then the *complementarity problem* associated with A and C is:

Find $v \in C$ so that $A(v) \in C^*$ and $\langle v, A(v) \rangle = 0$.

Suppose $H_0^1[x_a, x_F]$ represents the set of functions which have a generalized derivative in $L^2[x_a, x_F]$ on the interval $[x_a, x_F]$ and vanish identically elsewhere. The cone of non-negative functions within $H_0^1[x_a, x_F]$ will be denoted $P_0^1[x_a, x_F]$, where $f \geq 0$ means $f \in H_0^1[x_a, x_F]$ and $f(x) \geq 0$ almost everywhere in $[x_a, x_F]$. Making the correspondences yields:

$$B = H_0^1[x_a, x_F] \times H_0^1[x_a, x_F] \times R^1$$

$$C = P_0^1[x_a, x_F] \times H_0^1[x_a, x_F] \times R^1$$

$$C^* = (P_0^1[x_a, x_F])^* \times \{0\} \times \{0\}$$

$$A = \begin{bmatrix} R(p, h, k) \\ S(p, h, k) \\ T(p, h, k) \end{bmatrix} \quad \text{and} \quad v = \begin{bmatrix} p(\cdot) \\ h(\cdot) \\ k \end{bmatrix}$$

Use of the function space $H_0^1[x_a, x_F]$ corresponds to solutions of equations known as weak solutions. That is, if $A : H_0^1[x_a, x_F] \to H^{-1}$ then the equation $A(v) = f$ is to be interpreted as $\langle y, A(v) - f \rangle = 0$ for all $y \in H_0^1[x_a, x_F]$.

The linear complementarity problem for hydrodynamic lubrication of a journal bearing was studied in a function space similar to that above by Cryer and Dempster.[16] Bazaraa, Goode and Nashed[17] posed the non-linear complementarity problem in a Banach space, but cited applications only in finite dimensions. The case treated in this paper is the first physically motivated example of an instance of the non-linear complementarity problem which is infinite dimensional.

The generalized derivatives used in the above also allow for 'spikiness' and mild singularities in the functions $p(x)$, $h(x)$ and their derivatives. Features such as these have been observed in some numerical solutions and measured experimentally, and they are physically justifiable.

The infinite dimensional non-linear complementarity problem rarely yields to exact solutions. Hence, it is mainly of theoretical and modelling interest. For our computations we have discretized the problem to obtain a finite dimensional approximation by means of finite differences as follows.

Let $x_F = x_a + N \Delta x$. Then for $i = 1, 2, \ldots N$, let $p_i = p(x_a + i \Delta x)$ and for $j = i \pm \frac{1}{2}$, let $h_j = h(x_a + j \Delta x)$. For convenience, take $p_0 = k$.

A more compactly written form of the operator A is obtained by making a direct substitution of the h values into the equation for R. This yields, for $i = 1, 2, \ldots, N$,

$$R_i(p_0, p_1, p_2, \ldots, p_N) = -\frac{1}{(\Delta x)^2}\left[\frac{(h_{i+\frac{1}{2}})^3}{e^{\alpha p_{i+\frac{1}{2}}}}(p_{i+1}-p_i) - \frac{(h_{i-\frac{1}{2}})^3}{e^{\alpha p_{i-\frac{1}{2}}}}(p_i - p_{i-1})\right] + \frac{\lambda}{\Delta x}[h_{i+\frac{1}{2}} - h_{i-\frac{1}{2}}]$$

Let

$$R_0 = 1 - \frac{2}{\pi}\sum_{i=1}^{N} w_i p_i \, \Delta x$$

so that R_0 and k form a complementary pair of variables.

Values of p at half points are approximated by averaging:

$$p_{i+\frac{1}{2}} = \tfrac{1}{2}(p_i + p_{i+1}) + O((\Delta x)^2)$$
$$p_{i-\frac{1}{2}} = \tfrac{1}{2}(p_i + p_{i-1}) + O((\Delta x)^2)$$

Numerical integration (by parts) to evaluate $h(x)$ takes the form:

$$h_j = [x_a + j\,\Delta x]^2 + (p_0 + 1) + \frac{2}{\pi}\sum_{i=0}^{N} w_i((i-j)\,\Delta x)\ln|(i-j)\,\Delta x|p_i' \,\Delta x$$

where

$$p_0' = \frac{p_1 - 0}{2\,\Delta x}, \qquad p_1' = \frac{p_2 - 0}{2\,\Delta x}, \qquad p_N' = \frac{0 - p_{N-1}}{2\,\Delta x} \quad \text{and} \quad p_i' = \frac{p_{i+1} - p_{i-1}}{2\,\Delta x} \quad \text{for} \quad 2 \leqslant i \leqslant N-1.$$

The weights w_i of the trapezoid rule were used above in each case. For definiteness, the numerical integrations simply ignore the free boundary x_b, proceeding (as indicated by the above summations) across the entire $[x_a, x_F]$ interval. Once the free boundary has been located correctly, the summations approximate the integrals of the model.

Combining all these relationships yields the following $N+1$ dimensional complementarity problem: Find

$$p = (p_0, p_1, p_2, \ldots, p_N)^T \in R \times R_+^N = C$$

such that

$$R(p) = (R_0, R_1, R_2, \ldots, R_N)^T \in \{0\} \times R_+^N = C^*$$

satisfying

$$\langle p, R(p)\rangle = \sum_{i=0}^{N} p_i R_i(p) = 0$$

The non-negative orthant of N-dimensional Euclidean space $R_+^N = \{x \in R^N \mid x_i \geqslant 0, i = 1, 2, \ldots, N\}$ is the cone most often employed in computational studies in complementarity. The case of an arbitrary orthant in R^N is covered by the theory and algorithm of Habetler and Kostreva,[14] but the present case is slightly more general. Since such side constraints as the load constraint, (3), are not uncommon, we now show how to treat this important case.

First, some terminology will be introduced. A more detailed presentation of the theory is contained in Reference 14 and the references therein. The concepts of P-matrices and P-functions, which are generalizations of positive definite matrices, are useful here. An $n \times n$ matrix is said to be a *P-matrix* if all its principal minors are positive. A non-linear function $f: R^n \to R^n$ is a *P-function* on a set S if for all $a, b \in S$ with $a \neq b$, there exists an index $i = i(a, b)$ such that $(a_i - b_i)(f_i(a) - f_i(b)) > 0$. For each subset $I^{(k)}$, $k = 1, 2, \ldots, 2^n$ of the set

$N=\{1, 2, \ldots, n\}$ the *principal subfunction* $f^{(k)}$ is given by

$$f_i^{(k)}(x)=\begin{cases} f_i(x), & i\in I^{(k)} \\ x_i, & i\in N-I^{(k)} \end{cases}$$

If f is a P-function on R^n such that for each $k=1, 2, \ldots, 2^n$, the principal subfunction $f^{(k)}$ maps R^n onto R^n, we say f is a *non-degenerate P-function* on R^n.

Theorem

Let $I\subseteq\{1, \ldots, n\}$ and $I^c=\{1, \ldots, n\}-I$. Suppose $f: R^n\to R^n$ is a non-degenerate P-function on $S_I=\{x\in R^n \mid f_i(x)=0,\ i\in I\}$. Then the direct algorithm will be successful when applied to the complementarity problem

$$\left.\begin{aligned} f_i(x)&=0 \\ x_i&\ \text{unrestricted}\end{aligned}\right\},\qquad i\in I$$

$$\left.\begin{aligned} f_i(x)&\geq 0 \\ x_i&\geq 0\end{aligned}\right\},\qquad i\in I^c$$

and

$$x_i f_i(x)=0 \qquad i\in I\cup I^c$$

Proof. The algorithm will successfully solve the complementarity problem on an orthant when the function is a non-degenerate P-function. Since f is a non-degenerate P-function on S_I, we may apply this result to the subproblem on I^c. Since P-functions are closed under principal pivot operations and every principal subfunction of a P-function is a P-function, the algorithm is well defined and will be successful on this reduced problem, so long as we remain in S_I. This is easily accomplished by requiring all complementary points encountered by the algorithm to satisfy $f_i(x)=0$, $i\in I$. □

Whether a function f is a non-degenerate P-function on a set or not is somewhat difficult to ascertain. For the case of elasto-hydrodynamic lubrication of rollers, certain values of the parameters α and λ fail to yield non-degenerate P-functions. This fact was discovered by performing eigenvalue analyses of the Jacobian matrix J_R near solution points. When J_R has a negative real eigenvalue, R cannot be a P-function. It is conjectured that for some subset of the range of permissible values for α and λ, R is a non-degenerate P-function. Further analyses of this kind are now being performed and will be reported elsewhere. It should be noted that, even in the cases where R is not a P-function, the algorithm has always performed well, successfully computing solutions to the complementarity problem.

The algorithm itself is simple to describe and intuitively appealing. First, to fix ideas, make the association between the region $\Omega\subseteq[x_a, x_F]$ where Reynolds' equation is satisfied and an index set I which is a subset of $\{1, 2, \ldots, N\}$ using $x=(x_a+i\,\Delta x)\in\Omega$ iff $i\in I$. The solution will be obtained by considering a non-repeating sequence of trial regions until the correct region $\Omega^*=[x_a, x_b]$ is found. Through the above association, a sequence $\{I^{(k)}\}_{k=0}^{M}$ of trial index sets will be considered until the correct one is found.

One begins by choosing an initial trial index set $I^{(0)}$. Next the associated set of equations are solved approximately:

$$\begin{aligned} R_i(p)&=0, \quad \text{for} \quad i\in\{0\}\cup I^{(0)} \\ p_i&=0, \quad \text{otherwise.}\end{aligned} \tag{5}$$

This yields the intermediate solution $p^{(0)}$. Next $p^{(0)}$ is evaluated to see if it satisfies the

non-negativity constraints. If it does, we are done. If not, we must modify the trial region according to some systematic rule. Considered as a set, the non-negativity constraints $p_i \geq 0$, $R_i(p) \geq 0$ are placed in some natural numbering order. Sighting down this list find the index of the first constraint violated by the current intermediate solution. If for this index, say i_k, $p_{i_k} < 0$, form a new index set by

$$I^{(k+1)} = I^{(k)} - \{i_k\}$$

If $R_{i_k}(p) < 0$, then form a new index set according to

$$I^{(k+1)} = I^{(k)} \cup \{i_k\}.$$

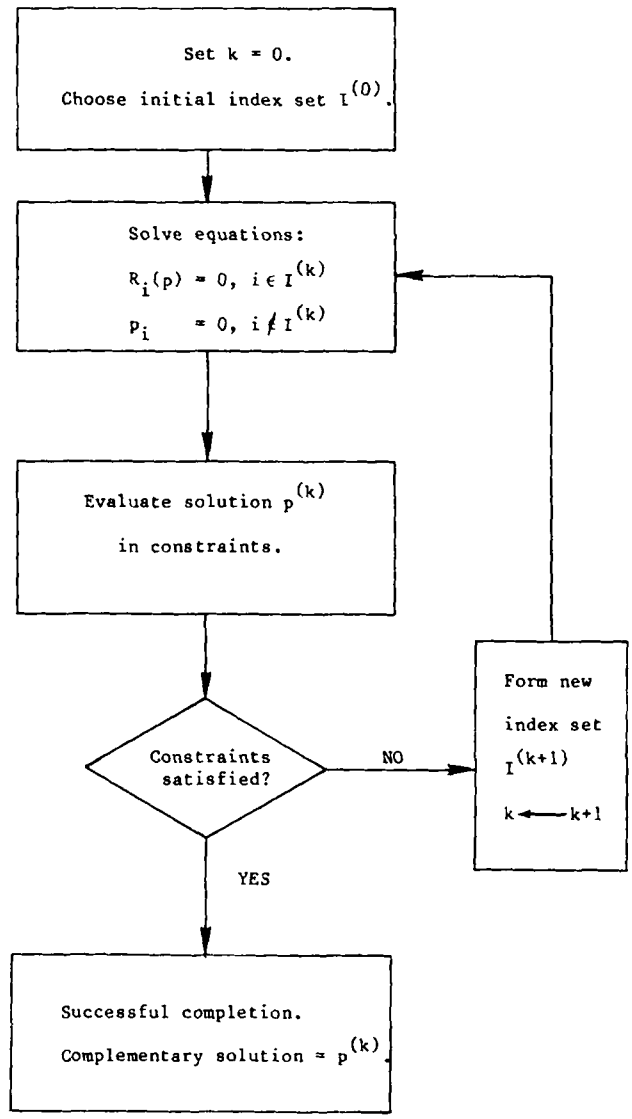

Figure 1. Flow chart—a direct algorithm for the complementarity problem

Using the corresponding trial region, now solve the equations $R_i = 0$, $i \in \{0\} \cup I^{(k)}$ and $p_i = 0$ otherwise. The new intermediate solution is then evaluated against the constraints. If necessary, the index set (trial region) is modified as before. Continue iteratively until no violated constraint can be found. An inductive proof[14] demonstrates that the algorithm terminates in a finite number of steps (index sets or trial regions) without ever retracing a step.

The number of steps required by the algorithm is a function of how well the initial trial region approximates the correct region. Using the procedures outlined in the next section seems to keep the number of trial regions low, usually less than 10. A flow diagram of the algorithm is given in Figure 1.

4. IMPLEMENTATION CONSIDERATIONS

This paper demonstrates the applicability of the direct algorithms of non-linear complementarity theory to large scale, practical engineering problems. Many of the implementation issues addressed here are general in nature and are useful for complementarity algorithms in other contexts, while some of the details relate specifically to elasto-hydrodynamic lubrication.

In any problem containing mildly singular functions a certain amount of care is necessary to prevent function evaluations too close to singularities. The computational molecule used here, which always maintains at least $(\Delta x/2)$ distance from the logarithmic singularity seems to be sufficient for practical values of Δx. The order of approximation, due to the avergaging and centred differences, is seen to be $O((\Delta x)^2)$.

Because the decision rules of the direct algorithm are based on sign configurations of the independent and dependent variables, a numerical filter was implemented. In this filter, any value less than $\frac{1}{5}(\Delta x)^2$ in absolute value in the complementary point is replaced by zero. Such a filter tends to eliminate index set changes which might be triggered by round-off and/or truncation errors in the numerical solutions. The value $\frac{1}{5}$ was found experimentally.

Direct algorithms for the complementarity problem require the use of a non-linear equation solver, a subroutine which can reliably solve systems of n non-linear equations in n unknowns. Most of the time these solutions are intermediate points which need not be extremely accurate, but should furnish correct sign configurations. A non-linear equation solver which has proven to be quite adaptable for the current use is that given by Powell.[18,19] The algorithm features a larger sphere of convergence than does the Newton–Raphson iterative method. Powell's 'Hybrid', as it has come to be known, is also readily available through many software libraries. Four codes which use variations of Powell's original idea are CO5NAF (The Numerical Algorithm Group (USA) Inc., Downers Grove, IL, U.S.A.), HYBRD (Applied Math Division, Argonne National Lab., Argonne, IL, U.S.A.), NS01AD (Computer Science and Systems Division, A.E.R.E. Harwell, Oxfordshire, U.K.) and ZONE (Bell Laboritories, Murray Hill, NJ, U.S.A.). The wide availability of robust non-linear equation solvers such as these makes the direct algorithm approach practical and relatively simple to implement. The particular version we have used to date is NS01AD. Parameter settings used to obtain results will be given along with the results.

Scaling the equations to be solved is a special concern in Powell's Hybrid. During the calculation, the function values should be approximately the same order of magntiude. However, in our EHL model, the form of the difference equations tends to make this difficult to achieve with fixed scaling factors. Therefore, a dynamic scaling of the equations was used.

Whenever the $R_i(p_0, p_1, \ldots, p_N) = 0$ was to be solved it was replaced by the equivalent scaled equation $\exp(sc \cdot p_i) \cdot R_i(p_0, \ldots, p_N) = 0$. The scale factor sc is defined by

$$sc = \begin{cases} \alpha(0{\cdot}0625\alpha - 0{\cdot}125), & \text{if } \alpha \geq 2 \\ 0, & \text{if } \alpha < 2 \end{cases}$$

and was found through experimentation. Scaling by an exponential factor does not change the solution set of the system of equations but does make it more easily computable with Powell's Hybrid.

Choosing a start point for a non-linear equation solver is always a difficult matter. In the complementarity problem the choice of an initial index set provides additional complication, as does the choice of a start point for intermediate systems. In our approach, we handle intermediate systems by feeding solutions forward: the final solution corresponding to index set $I^{(k)}$ is used as a start point for index set $I^{(k+1)}$. The initial start point and initial index set are handled in one of two ways. If a previous solution is available with α and λ parameters not too distant from the desired values, it may be used as a start point together with the index set (region) it implies. If no such solution is available, a modified Hertzian solution will serve as a start point and initial index set. Such a function can be generated by the formula

$$p_{\text{Hertz}} = \begin{cases} 0{\cdot}06 + 0{\cdot}02x, & x < -1 \\ \sqrt{(1-x^2)}, & -1 \leq x \leq 1 \\ 0, & x > 1 \end{cases}$$

where the linear 'ramp' corresponds to $x_a = -3$. Taking the index set corresponding to the positive p_{Hertz} values seems to work quite satisfactorily whenever the formula is applied.

Sometimes it is desired to solve the model for an (α, λ) pair which is not near any previously solved case and for which the modified Hertzian solution is inadequate. Then a simple continuation procedure can be useful.

Starting from a previous run (α^0, λ^0) one solves a sequence of problems $(\alpha^i, \lambda^i)_{i=0}^M$ such that (α^M, λ^M) is the desired (α, λ) pair. One feeds forward the solution and index set from each succeeding problem. Setting the (α^i, λ^i) along a straight line seems to work reasonably well. A progression from lower to higher values in the parameters seems to permit the continuation to proceed more easily than from higher to lower. This can be understood by inspecting the meaning of (α, λ) and interpreting the continuation physically. The fact that absolutely no derivative information has been required in this continuation process may be attributed to the robustness of Powell's Hybrid algorithm and the ability of the direct algorithm to locate free boundaries through complementarity.

5. COMPUTATIONAL RESULTS

Two categories of numerical results comprise this section. First, some cases which have been solved by other techniques in the literature will be examined to demonstrate the validity of the complementarity approach.* Following this, a more realistic demonstration will be presented where the algorithm is applied to actual bearing data, in which the loads and running speeds are varied.

---

* In these cases, the axes were rescaled to be consistent with the solutions in the literature.

In all the computations performed, the non-linear equations were solved by using the HARWELL subrouting NS01AD with the following input parameter settings:

$$\text{STEP} = 1{\cdot}\text{D}-10$$
$$\text{STPMAX} = 1{\cdot}\text{D5}$$
$$\text{ACC} = 1{\cdot}\text{D}-8$$
$$\text{MAXFUN} = 2000$$

STEP is a parameter which is used to form forward difference approximations to derivatives in NS01AD. STPMAX is a generous estimate of the distance between the initial point and the required solution of the equations. The parameter ACC is the accuracy required. A normal return of NS01AD is made when $\sum_{i=1}^{n}[f_i(x)]^2 \leqslant \text{ACC}$, corresponding to the system of equations $f_i(x) = 0$, $i = 1, \ldots, n$. The maximum number of calls to a function evaluation routine (user supplied) is denoted by MAXFUN.

Although the number of difference equations varies from case to case, the grid size was fixed at $\Delta x = 0{\cdot}05$. All the calculations were performed in double precision on an IBM 370 model 3033 computer.

In 1962, Stephenson and Osterle[3] obtained the solution for $\alpha = 1{\cdot}838$ and $\lambda = 1{\cdot}642$. Later Taylor and O'Callaghan[6] and Rohde and Oh[7] obtained matching solutions using finite elements rather than finite differences for discretization. Figure 2 shows the solution obtained with the method of this paper. It is not distinguishable from the earlier solutions.

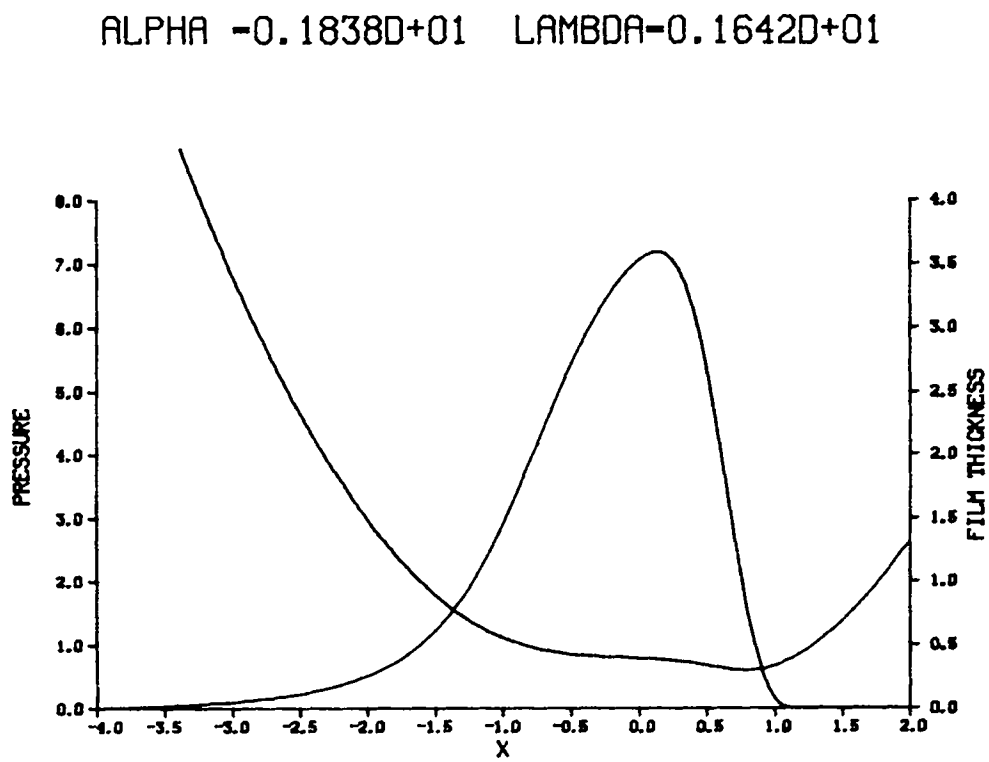

Figure 2

Next we examine two solutions which were presented by Rohde and Oh. These cases have somewhat higher α values corresponding to higher loads. In Figure 3, what Rohde and Oh call a 'medium load' case ($\alpha = 3{\cdot}468$ and $\lambda = 1{\cdot}065$) is shown. Some discrepancies appear. In particular, note the location of the free boundary and the location and shape of the 'pressure spike'. To evaluate their solution, its graph (see Reference 7, p. 325) was digitized and the result was input as a starting point for the direct complementarity algorithm. Convergence was to the same solution (Figure 3) which was originally computed from a Hertzian starting point. Thus we conclude that, in this case, $x_b > 1{\cdot}0$, which disagrees with their solution graph. The solution they present mislocates the free boundary, and hence violates $\frac{dh}{dx} > 0$ in $[x_b, x_F]$, which is a meaningful geometric constraint corresponding to diverging cylinders.

Finally, a 'heavily loaded' case was solved. For comparison a selection of model parameters of $\alpha = 4{\cdot}077$ and $\lambda = 0{\cdot}03624$ led to a solution which was indistinguishable from that of Rohde and Oh. See Figure 4.

Next consider a roller bearing with a 5·1 inch diameter inner race and rollers of diameter 0·875 in. If these are steel, we assume a Poisson's ratio of 0·3 and a Young's modulus of $2{\cdot}97 \times 10^7$ lb/in^2. Suppose the roller length is 0·85 in. A lubricant with a viscosity of $0{\cdot}174 \times 10^{-5}$ (lb s)/in^2 at atmospheric pressure and a pressure viscosity coefficient of $0{\cdot}70 \times 10^{-4}$ in^2/lb is to be used in the bearing. Three different loading conditions (100 lb, 175 lb and 250 lb) and three different speeds (500 rpm, 2500 rpm and 5000 rpm) will be considered. These cases are meant to represent a typical engineering 'check-out' of a hypothetical roller

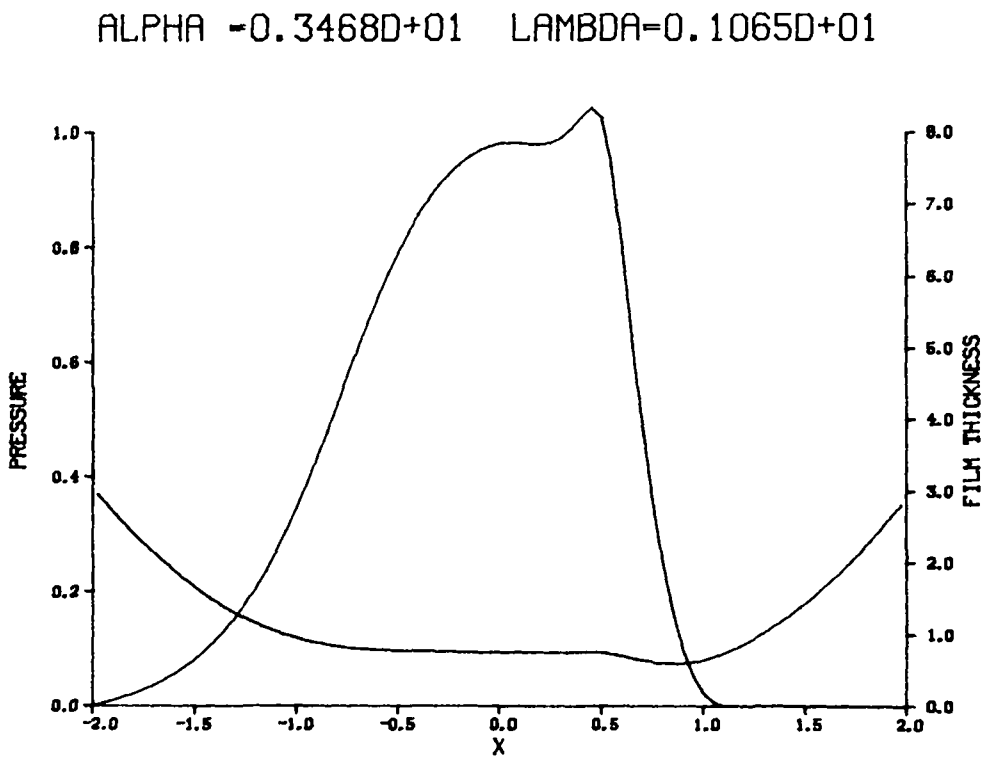

Figure 3

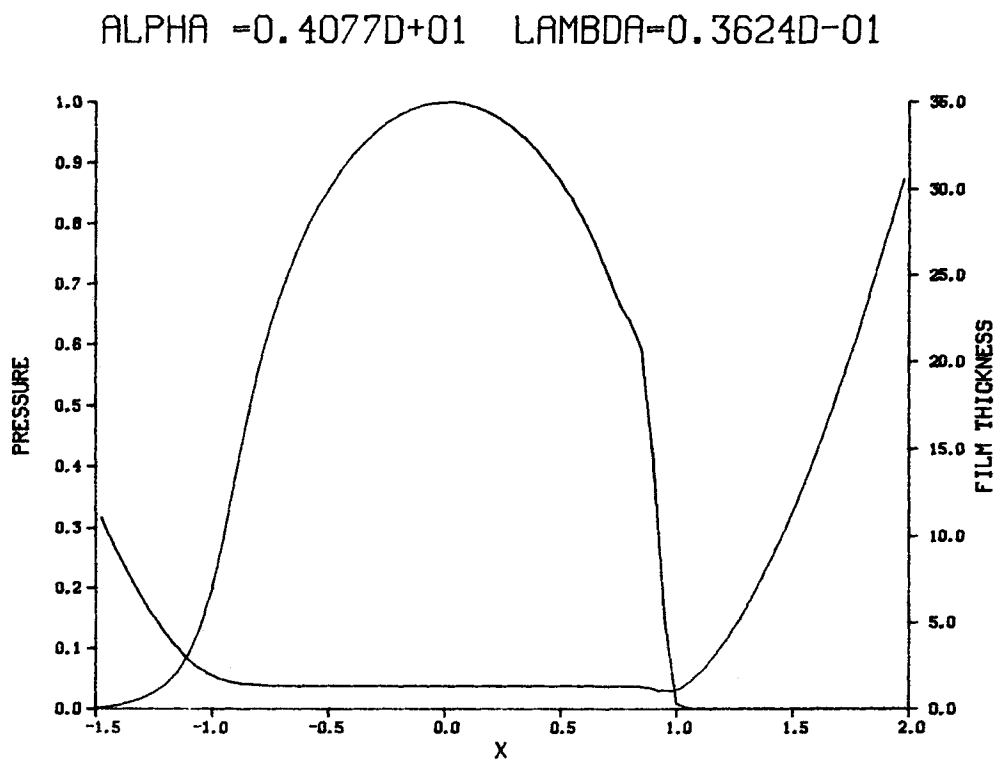

Figure 4

bearing with the model. Table I summarizes the model parameters corresponding to these running conditions.

Studying these solutions (Figures 5–13) one can gain some interesting insights into the qualitative behaviour of the solutions as a function of speed and load. Note that for most cases there is no appearance of a 'plateau' region in $h(x)$, the lubricant film thickness. Thus Grubin's[1] simplifying assumption is not valid in such cases. Increasing speed while holding load constant produces an increase in film thickness, a drift upstream of the pressure

Table I

Run No.	Load (lbs)	Speed (rpm)	α(Alpha)	λ(Lambda)
1	100	500	2·832	6·057
2	100	2500	2·832	30·29
3	100	5000	2·832	60·57
4	175	500	3·746	1·978
5	175	2500	3·746	9·889
6	175	5000	3·746	19·78
7	250	500	4·477	0·9692
8	250	2500	4·477	4·846
9	250	5000	4·477	9·692

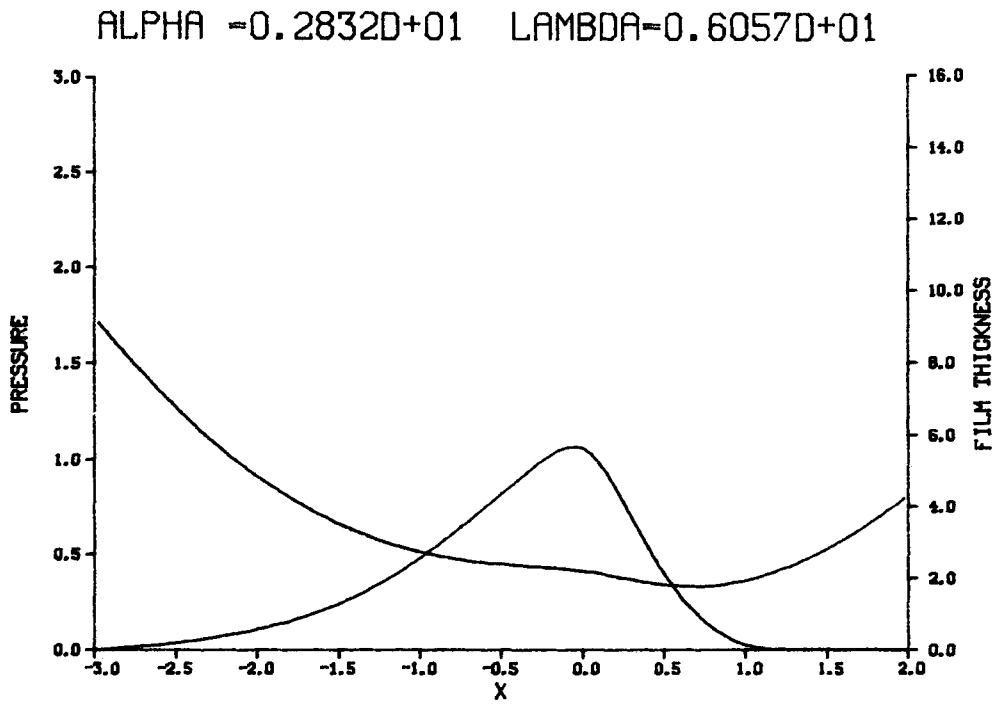

Figure 5. Run No. 1

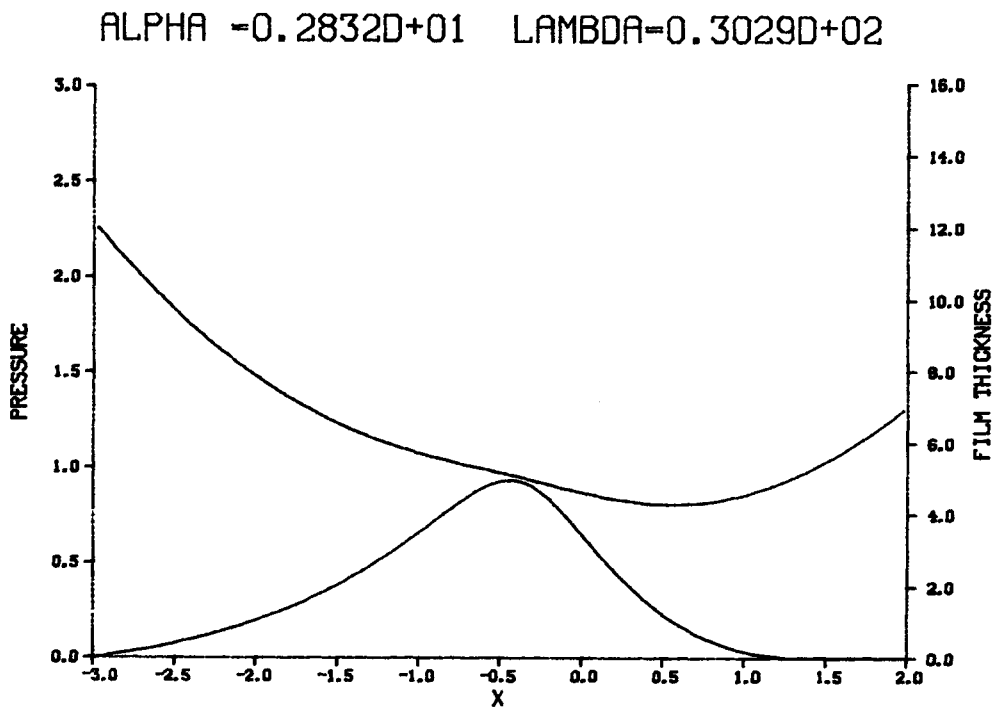

Figure 6. Run No. 2

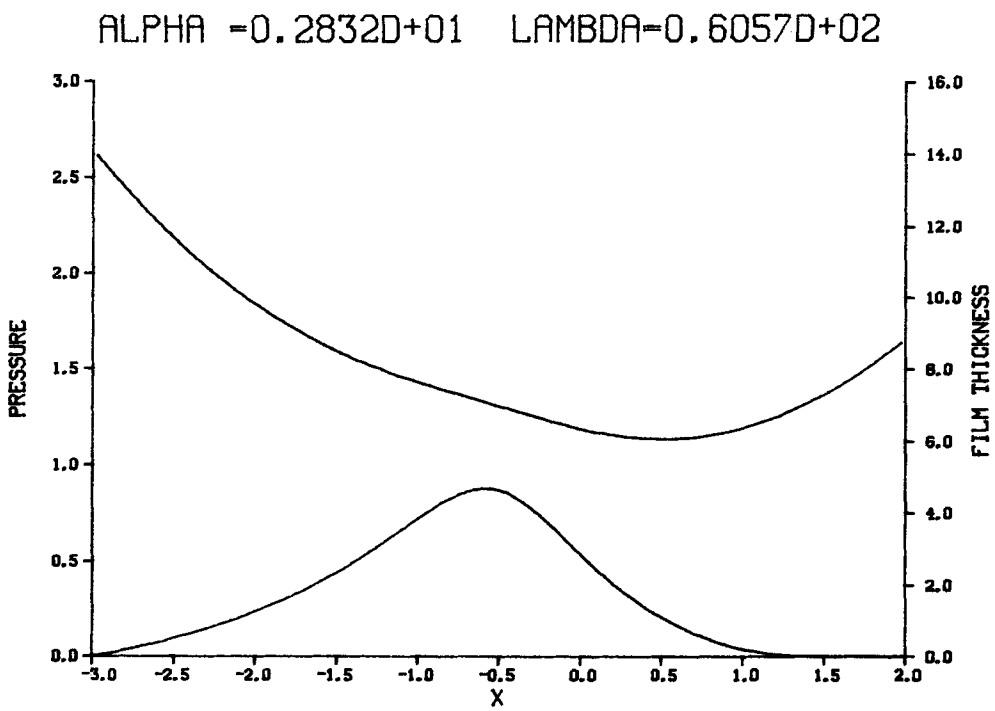

Figure 7. Run No. 3

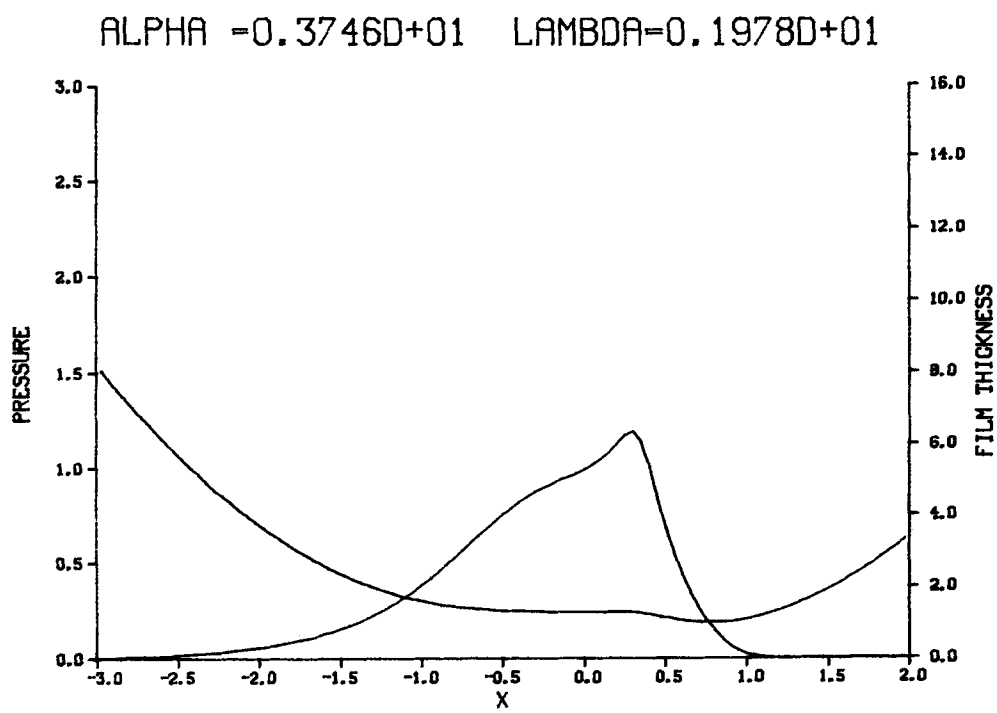

Figure 8. Run No. 4

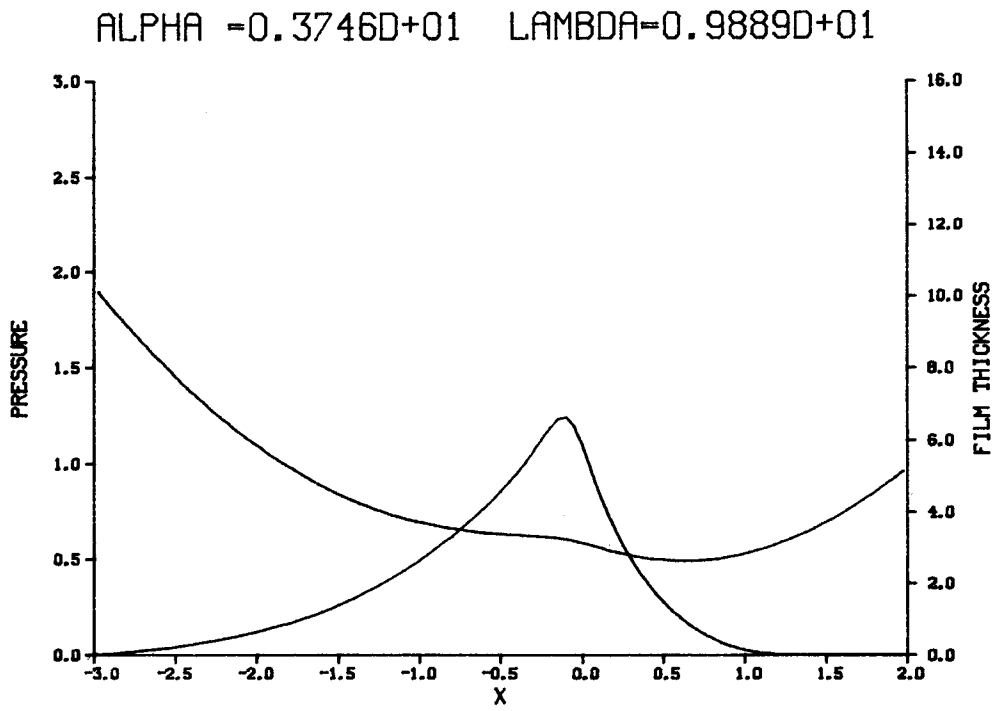

Figure 9. Run No. 5

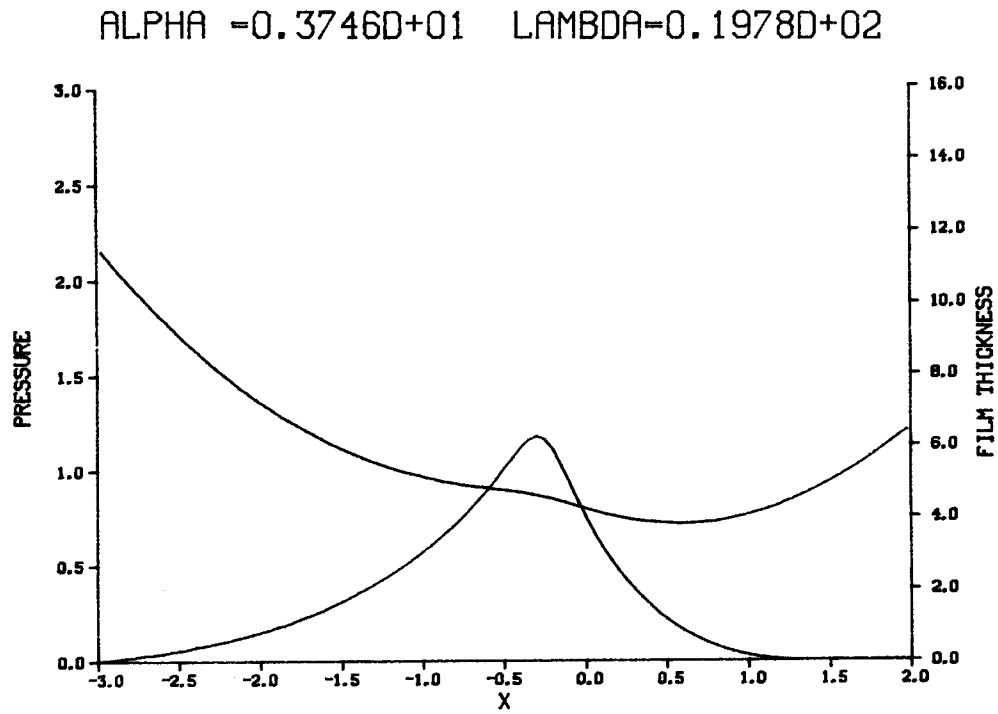

Figure 10. Run No. 6

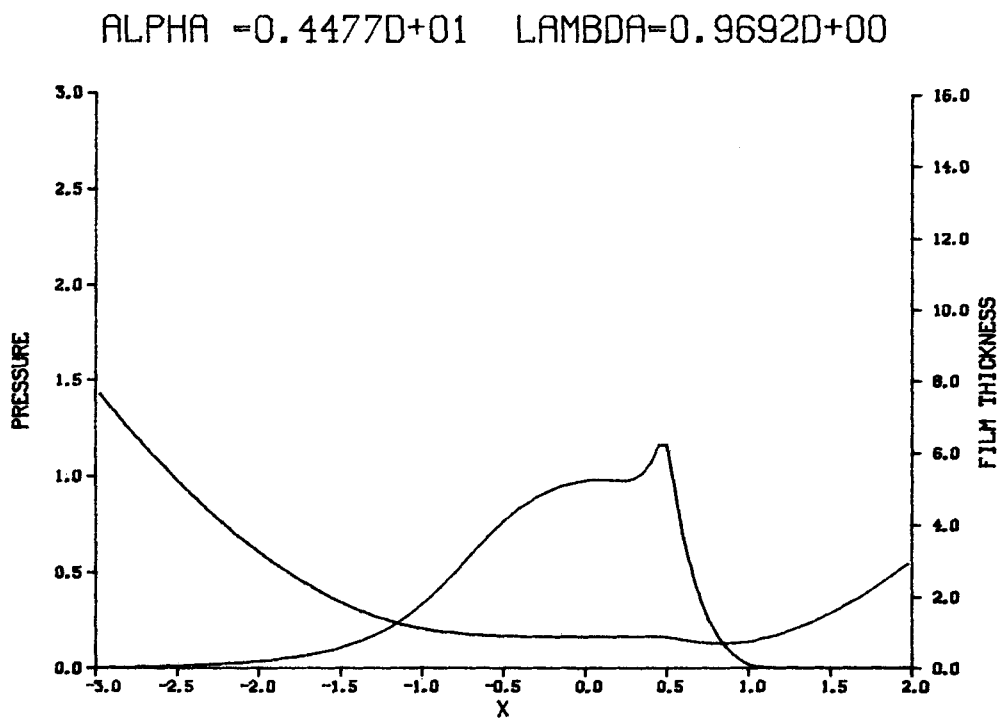

Figure 11. Run No. 7

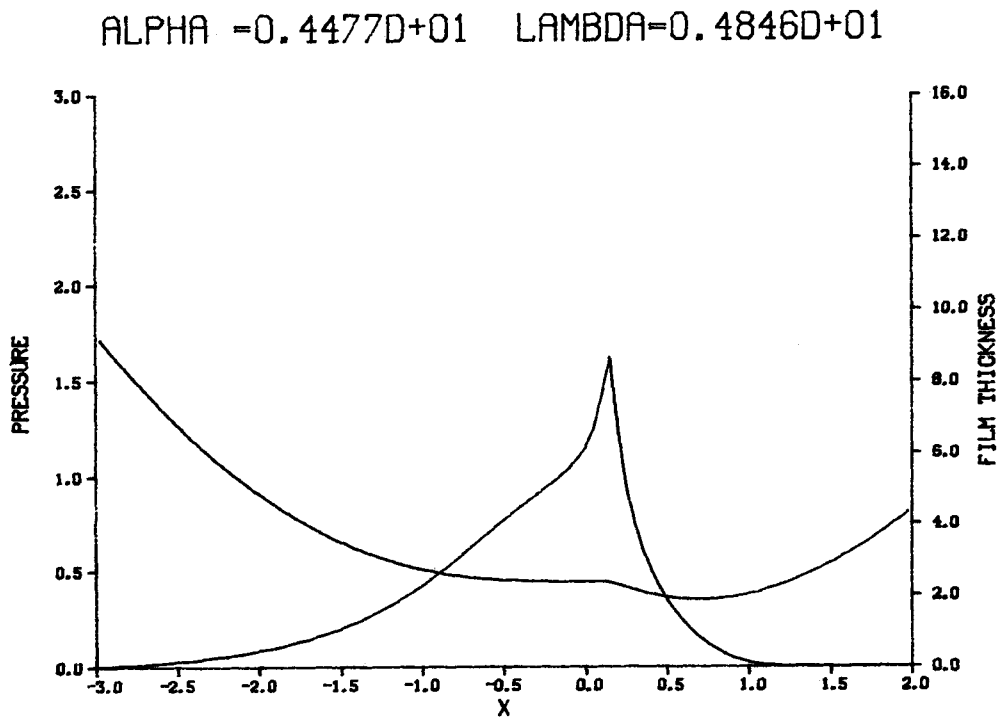

Figure 12. Run No. 8

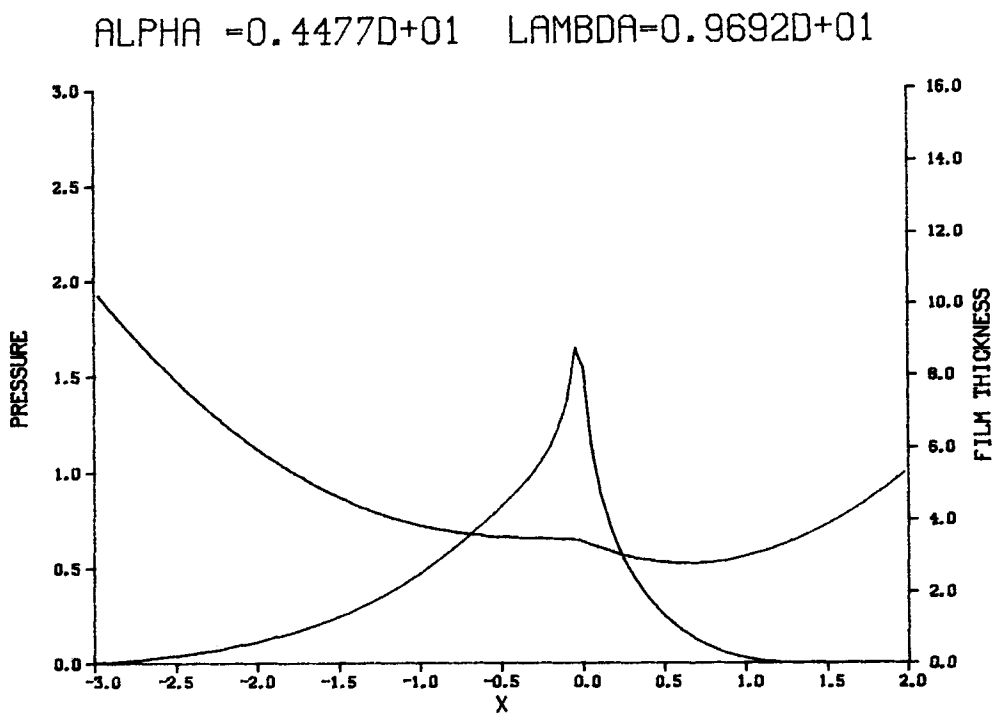

Figure 13. Run No. 9

distribution and a lessening of spikiness in $p(x)$. The free boundary x_b moves downstream with increasing speed. Increasing load leads to decreased film thickness, pronounced spikiness of the pressure distribution and an increased resemblance at low speed of $p(x)$ to the dry Hertzian solution. Load increases move the free boundary x_b upstream.

A quantity of great interest to the engineer performing an elasto-hydrodynamic analysis of a bearing is the minimum film thickness. Dowson and Higginson[1] claim that the minimum film thickness can be fairly accurately represented over the whole range of theoretical solutions by a simple algebraic formula. An earlier film thickness formula derived by Grubin is widely known and used. Since these handbook formulae are the main analytic tools available short of computer models, a comparison of our model with the formulae is made here for the hypothetical bearing described above.

As seen from Table II, the model agrees with the formulae under a set of predictable circumstances, namely low speed and moderate to high loads. These circumstances agree with the hypothesis of Grubin which was that the lubricated film thickness h would merely be a translation of the geometry of the Hertzian (dry) case. We see from the graphs (Figures 5–13) Grubin's hypothesis is not valid in most cases. No 'plateau' region is apparent for most cases, so Grubin's analysis, not suprisingly, is not valid. This, however, does not explain the disagreement with the formula of Dowson and Higginson. Their formula was derived by curve fitting using their numerical solutions. Judging from their results, two explanations are possible. Perhaps not enough computer solutions were used for the curve fitting. The graph in which they display their comparison of computer model vs. formula minimum film thicknesses contains only 12 points. (see Reference 1, p. 97). Since six physical parameters

Table II. Minimum film thickness

Run No.	$h^{\min}_{\text{model}}$	δ(scaling factor)	Model (μ in)	Grubin (μ in)	D & H (μ in)
1	1·7584192	0·45896127(−5)	8·1	9·2	8·7
2	4·2755541		19·6	29·7	26·9
3	6·0481208		27·8	49·2	43·7
4	0·9884915	0·8031824(−5)	7·9	8·8	8·1
5	2·6114315		20·9	28·2	25·0
6	3·8071227		30·6	46·8	40·6
7	0·6754051	0·11474034(−4)	7·75	8·5	7·7
8	1·8615249		21·4	27·3	23·9
9	2·7817967		31·9	45·3	38·8

enter the formula, they may have run computer models in only a small section of the regime of interest. Another possibility for the lack of agreement is that their computer solutions may not have converged. This is more speculative since no details are given about the actual solutions used in their curve fitting. However, this cause is not completely out of the question owing to the *ad hoc* nature of their numerical solution procedure.

Our results show that the formulae of Grubin and Dowson–Higginson provide satisfactory answers over only a limited range of parameters. Significant deviations from accurate numerical calculations occur even at moderate speeds. For high speeds, formulae calculated minimum film thickness is 42 per cent to 76 per cent too large. Fortunately, the complementarity method provides a computationally feasible alternative.

CONCLUDING REMARKS

Reformulation of the mathematical model for elasto-hydrodynamic lubrication of cylinders in line contact as a non-linear complementarity problem has produced a number of interesting insights. First, the complementarity approach has a meaningful interpretation in terms of physical and geometric quantities. Next, solutions are reliably and automatically computable now with the direct algorithm presented here. Finally, when the method is appropriately computerized, lubrication engineers can conveniently obtain accurate solutions over a wide range of parameters.

ACKNOWLEDGEMENT

The author thanks Dr. Dah-Chen Sun of the Mechanical Research Department of General Motors Research Laboratories for introducing him to the elasto-hydrodynamic lubrication problem and for many interesting conversations on this work.

APPENDIX: NOMENCLATURE

Symbol	Definition	Formula	Units
a	Hertzian half-width	$2\sqrt{\left(\frac{RP}{E}\right)}$	length

Symbol	Definition	Formula	Units
E_1, E_2	Young's moduli of two cylinders		$\frac{\text{force}}{(\text{length})^2}$
E	effective Young's Modulus	$\frac{\pi}{\left(\frac{1-v_1^2}{E_1}+\frac{1-v_2^2}{E_2}\right)}$	$\frac{\text{force}}{(\text{length})^2}$
$\bar{h}$	film thickness		length
h	dimensionless film thickness	$\frac{\bar{h}}{\delta}$	—
P	load on cylinders		$\frac{\text{force}}{\text{length}}$
$\bar{p}$	pressure in lubricant	$\frac{\text{force}}{(\text{length})^2}$	
p	dimensionless pressure	$\frac{\bar{p}}{p_{\max}}$	—
$p_{\max}$	Hertzian maximum pressure	$\left(\sqrt{\frac{EP}{R}}\right)\Big/\pi$	$\frac{\text{force}}{(\text{length})^2}$
R_1, R_2	radii of two cylinders		length
R	effective radius	$1\Big/\left(\frac{1}{R_1}+\frac{1}{R_2}\right)$	length
S	generic subset of Euclidean space		—
s	integration variable		—
$\bar{x}$	distance along surface of cylinder		length
x	dimensionless independent variable	$\frac{\bar{x}}{a}$	—
x_a	dimensionless lubricant inlet point		—
x_b	dimensionless lubricant outlet point (cavitation point, free boundary)		—
x_F	dimensionless point downstream of outlet point		—
u_1, u_2	surface speed of two cylinders		$\frac{\text{length}}{\text{time}}$
u	entrainment velocity	$\frac{u_1+u_2}{2}$	$\frac{\text{length}}{\text{time}}$
$\bar{\alpha}$	pressure viscosity coefficient		$\frac{(\text{length})^2}{\text{force}}$
α	dimensionless pressure viscosity coefficient	$\bar{\alpha}p_{\max}$	—
δ	Hertzian reference deformation	$\frac{2P}{E}$	length

Symbol	Definition	Formula	Units
λ	dimensionless model parameter	$\frac{6\pi\mu_0 uRE}{P^2}$	—
μ_0	reference viscosity of lubricant		$\frac{\text{(force)(time)}}{\text{(length)}^2}$
μ	pressure dependent viscosity	$\mu_0 e^{\bar{\alpha}\bar{p}}$	$\frac{\text{(force)(time)}}{\text{(length)}^2}$
ν_1, ν_2	Poisson's ratios of two cylinders		—

REFERENCES

1. D. Dowson and G. R. Higginson, *Elasto-Hydrodynamic Lubrication*, SI Edition, Pergamon Press, Oxford, 1977.
2. D. Dowson and G. R. Higginson, 'A numerical solution to the elasto-hydrodynamic problem', *J. Mech. Eng. Sci.*, **1**, 6–15 (1959).
3. R. R. Stephenson and J. F. Osterle, 'A direct solution of the elasto-hydrodynamic lubrication problem', *ASLE Trans.* **5**, 365–374 (1962).
4. H. S. Cheng and B. Sternlicht, 'A numerical solution for the pressure, temperature, and film thickness between two infinitely long, lubricated rolling and sliding cylinders, under heavy loads', *ASME J. Basic Eng.*, **87**, 695–707 (1965).
5. K. Herrebrugh, 'Solving the incompressible and isothermal problem in elastohydrodynamic lubrication through an integral equation', *ASME J. Lubrication Tech.* **90**, 262–270 (1968).
6. C. Taylor and J. F. O'Callaghan, 'A Numerical Solution of the Elastohydrodynamic lubrication problem using finite elements', *J. Mech. Eng. Sci.*, **14**, 229–237 (1972).
7. S. M. Rohde and K. P. Oh, 'A unified treatment of thick and thin film elastohydrodynamic problems by using higher order element methods', *Proc. R. Soc. Lond. A.*, **343**, 315–331 (1975).
8. C. Taylor and J. F. O'Callaghan, 'Discussion of the Swales–Dowson–Latham Paper', *Proc. Instn. Mech. Engrs. Elastohydrodynamic Lubrication Symposium*, London, 1972.
9. P. D. Swales, D. Dowson and J. L. Latham, 'Theoretical and experimental observations of the behavior of soft elastic materials under elastohydrodynamic conditions', in *Proc. Instn. Mech. Engrs. Elastohydrodynamic Lubrication Symposium*, London, 1972.
10. D. G. Christopherson, 'A new mathematical method for the solution of film lubrication problems', *Proc. Inst. Mech. Eng.*, **146**, 126–135 (1941).
11. C. W. Cryer, 'The method of Christopherson for solving free boundary problems for infinite journal bearings by means of finite differences', *Math. of Computation*, **25**, 435–443 (1971).
12. C. W. Cryer, 'The solution of a quadratic programming problem using systematic overrelaxation', *SIAM J. Control*, **9**, 385–392 (1971).
13. T. F. Conry and A. Seirig, 'A mathematical programming technique for the evaluation of load distribution and optimal modifications for gear systems', *ASME Trans. (Series B) J. Eng. Industry*, **95**, 1115–1122 (1973).
14. G. J. Habetler and M. M. Kostreva, 'On a direct algorithm for nonlinear complementarity problems', *SIAM J. Control Optimization*, **16**, 504–511 (1978).
15. C. E. Lemke, 'Recent results on complementarity problems', in J. B. Rosen, O. L. Mangasarian and K. Ritter (eds.), *Nonlinear Programming*, Academic Press, New York, 1970.
16. C. W. Cryer and M. A. H. Dempster, 'Equivalence of linear complementarity problems and linear programs in vector lattice Hilbert spaces', *SIAM J. Control and Optimization*, **18**, 76–89 (1980).
17. M. S. Bazaraa, J. J. Goode and M. Z. Nashed, 'A nonlinear complementarity problem in mathematical programming in Banach space', *Proc. Amer. Math. Soc.*, **35**, 165–170 (1972).
18. M. J. D. Powell, 'A Fortran subroutine for solving systems of nonlinear algebraic equations', in P. Rabinowitz (ed.), *Numerical methods for Nonlinear Algebraic Equations*, Gordon and Breach, London, 1970.
19. M. J. D. Powell, 'A hybrid method for nonlinear equations', in P. Rabinowitz (ed.), *Numerical Methods for Nonlinear Algebraic Equations*, Gordon and Breach, London, 1970.