296 Reg

AEC RESEARCH AND DEVELOPMENT REPORT

2/28

K-1737

MASTER

A STUDY OF NON-LINEAR DIRICHLET PROBLEMS

AUTHORS:

Dale W. Lick
J. Nelson Tunstall

UNION CARBIDE

UNION CARBIDE CORPORATION
NUCLEAR DIVISION

Operating the

OAK RIDGE GASEOUS DIFFUSION PLANT • OAK RIDGE NATIONAL LABORATORY
OAK RIDGE Y-12 PLANT • PADUCAH GASEOUS DIFFUSION PLANT

For the Atomic Energy Commission
Under U.S. Government Contract W7405 eng 26

DISCLAIMER

Printed in the United States of America. Available from Clearinghouse for Federal Scientific and Technical Information, National Bureau of Standards, U.S. Department of Commerce, Springfield, Virginia 22151
Price: Printed Copy $3.00; Microfiche $0.65

Report Date: February 27, 1968

Report Number: K-1737

Subject Category: Mathematics and Computers

A STUDY OF NON-LINEAR DIRICHLET PROBLEMS

by

Dale W. Lick
Applied Mathematics Department
Brookhaven National Laboratory
Upton, Long Island, New York

J. Nelson Tunstall
Computing Technology Center

UNION CARBIDE CORPORATION
NUCLEAR DIVISION
Oak Ridge Gaseous Diffusion Plant
Oak Ridge, Tennessee

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

TABLE OF CONTENTS

PAGE

ABSTRACT 5

PART I. A Non-Linear Dirichlet Problem

1. Introduction 7
2. Statement of the Problem 7
3. Theoretical Results. 8
4. Iteration Methods. 9
5. An Analysis. 13
6. Numerical Results. 20

PART II. A Dirichlet Problem for a System of Non-Linear Partial Differential Equations

1. Introduction 26
2. Statement of the Problem 26
3. Iteration Methods. 27
4. Other Results. 29
5. An Analysis. 31
6. Numerical Results. 32

ACKNOWLEDGEMENT 35

REFERENCES. 36

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

A STUDY OF NON-LINEAR DIRICHLET PROBLEMS

A B S T R A C T

Part I of this report considers the Dirichlet problem

$$\Delta u = f(u),\ u \geq 0,\ \text{in } R, \qquad u = \phi,\ \text{on } R', \tag{1}$$

where f is non-linear, R is a bounded domain with boundary R', ϕ is a nonnegative continuous function, and Δ is the Laplacian operator. It also discusses and analyzes the finite difference analogue of (1). Iteration methods and numerical calculations are included for the approximate solution of (1) with

$$f(u) = \frac{c_1 u}{1 + c_2 u},$$

where c_1 and c_2 are positive constants.

Part II of this report considers the finite difference analogue of the Dirichlet problem

$$\Delta u = f(u,v),\ u,\ v \geq 0,\ \text{in } R, \qquad \Delta v = g(u,v),\ u,\ v \geq 0,\ \text{in } R, \qquad u = \phi,\ v = \psi,\ \text{on } R', \tag{2}$$

where ϕ and ψ are nonnegative continuous functions and

$$f(u,v) = \frac{c_1 u}{1 + c_2 v + c_3 u}$$

and

$$g(u,v) = -\frac{c_4 u}{1 + c_2 v + c_3 u},$$

c_1, c_2, c_3, and c_4 being positive constants. The finite difference analogue of (2) is analyzed using the results of Part I and other known results, and iteration methods and numerical calculations are included for the solution of this problem.

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

Part I. A Non-Linear Dirichlet Problem

1. Introduction

In this part of the report we consider the elliptic boundary-value problem

$$\Delta u = f(u),\ u \geq 0,\ \text{in } R, \qquad u = \phi,\ \text{on } R', \tag{1}$$

where f is non-linear, R is a bounded domain with boundary R', ϕ is a nonnegative continuous function, and Δ is the Laplacian operator

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} .$$

The finite difference analogue of (1) is discussed and analyzed, and special attention is given to (1) with

$$f(u) = \frac{c_1 u}{1 + c_2 u} , \tag{2}$$

where c_1 and c_2 are positive constants.

We describe the finite difference analogue of (1) in Section 2. Section 3 contains an existence and uniqueness theorem for (1) and existence, uniqueness and convergence theorems for the finite difference problem. In Sections 4 and 5, we describe several iteration schemes for solving the approximate problem and give an analysis of certain of these. Section 6 contains results for a wide variety of numerical cases.

2. Statement of the Problem

We let R denote a fixed domain in the (x,y)-plane, and let $\overline{R}$ and R' denote the closure and boundary of R. Let P_o be the point with coordinates (x_o,y_o). The points $(x_o + h, y_o)$, $(x_o, y_o + h)$, $(x_o - h, y_o)$ and $(x_o, y_o - h)$ are called h-neighbors of P_o and are denoted $P_{o1}, \ldots, P_{o4}$. The lattice domain R_h is the set of points $P_1, \ldots, P_N$, situated in R, having coordinates that are multiples of h and such that $P_{i\nu}$, $i = 1, \ldots, N$; $\nu = 1, \ldots, 4$ belong to $\overline{R}$. Neighbors of points of R_h that are not themselves points of R_h form the boundary

R_h' of R_h. We denote the points of R_h', by $P_{N+1}, \ldots, P_M$, and the union of R_h' and R_h by $\overline{R}_h$. If R_h is fixed and ψ is a function on $\overline{R}_h$, we denote the value of ψ at point P_i (or at $P_{i\nu}$) by ψ_i (or $\psi_{i\nu}$). We define the operator L_h by

$$L_h[\psi_i] = \Delta_h[\psi_i] - f(\psi_i),$$

where

$$\Delta_h[\psi_i] = \frac{1}{h^2} \left\{ \sum_{\nu=1}^{4} \psi_{i\nu} - 4\psi_i \right\} .$$

The general boundary-value problem for some fixed R_h consists of determining a function u defined on $\overline{R}_h$ such that

$$\begin{aligned} &L_h[u_i] = 0, \; u_i \geq 0, \; i = 1, \ldots, N; \\ &u_j = \phi_j, \; \phi_j \geq 0, \; j = N+1, \ldots, M. \end{aligned} \tag{3}$$

We assume, of course, that h is sufficiently small so that at least one R_h exists.

In the next section we give results for the general boundary-value problems (1) and (3), but the remainder of this part of the report will be for (3) with f given by (2).

3. Theoretical Results

The first author [6], [7] has extended the results of Bers [2] and the classical theory of Courant and Hilbert [3] to obtain the following theorems.

Theorem 1. Let f be defined, bounded and continuously differentiable for $u \geq 0$. If f satisfies

(i) $f(0) \leq 0$,

(ii) $f'(u) \geq 0$, $u \geq 0$,

(iii) $f'(0) > 0$,

then (1) has one and only one solution.

Theorem 2. If f satisfies the conditions of Theorem 1, then (3) has one and only one solution.

Theorem 3. If the analytic solution of (1) is in class $C^2(\bar{R})$, then the finite difference solution of (3) converges to the analytic solution as the mesh is refined.

Remark 1. The theorems hold under other conditions on f (e.g., if $f(0) < 0$, then (iii) can be eliminated).

Remark 2. The theorems can be extended to higher dimensions.

We observe that these theorems not only answer the questions of existence, uniqueness and convergence for f given by (2), but are also applicable, for example, to problems with f having the form

$$-\frac{c_1}{1 + c_2 u}$$

or

$$-c_1 e^{-c_2 u},$$

where c_1 and c_2 are positive constants.

4. Iteration Methods

In this section we describe five iteration schemes for solving (3). The effectiveness of these and other methods will be discussed in Sections 5 and 6. For simplicity, we let R be a rectangle, and the points of the lattice R_h are numbered, as in the following figure, with m points along a horizontal line and n points along a vertical line ($N = mn$).

	n	2n	mn
	n-1	2n-1	mn-1
	.	.	.
	.	. . .	.
	.	.	.
	2	n+2	m(n-1)+2
	1	n+1	m(n-1)+1

FIGURE 1. Numbering of the Points of R_h

The first two methods (denoted L and M) involve linearizing (3) by using previous iterates in f and then solving the resulting system of linear equations iteratively by a point successive overrelaxation (SOR) method. We use previous iterates for all u in f for method L, but only for the denominator u for Method M. The iteration equations for these methods are

$$4u_i^{(k+1)} = 4u_i^{(k)} + \omega\{u_{i4}^{(k+1)} + u_{i3}^{(k+1)} + u_{i2}^{(k)} + u_{i1}^{(k)} - h^2 f(u_i^{(k)}) - 4u_i^{(k)}\},$$

and

$$u_i^{(k+1)} = u_i^{(k)} + \omega\left\{\frac{u_{i4}^{(k+1)} + u_{i3}^{(k+1)} + u_{i2}^{(k)} + u_{i1}^{(k)}}{4 + \dfrac{c_1 h^2}{1+c_2 u_i^{(k)}}} - 4u_i^{(k)}\right\},$$

$$i = 1, \ldots, N;\ k \geq 0,$$

where k is the iteration number and ω, here and elsewhere, is the relaxation factor $0 < \omega < 2$.

The third iteration method (denoted N) is what might be called a "Newton-SOR" method (see [4]). Its defining equations are

$$u_i^{(k+1)} = u_i^{(k)} + \omega \left\{ \frac{4u_i^{(k)} - [u_{i4}^{(k+1)} + u_{i3}^{(k+1)} + u_{i2}^{(k)} + u_{i1}^{(k)} - h^2 f(u_i^{(k)})]}{4 + h^2 f'(u_i^{(k)})} \right\},$$

$$i = 1, \ldots, N; \; k \geq 0.$$

A fourth method (denoted P) is essentially the same as Method M, except we solve the linearized equations by a block SOR scheme instead of a point method. The system of equations for (3) can be written in matrix form (see [12])

$$(4) \qquad \begin{bmatrix} A_1^{(k+1)} & -I & & 0 \\ -I & A_2^{(k+1)} & -I & \\ & \ddots & \ddots & \ddots \\ & & -I & A_m^{(k+1)} \end{bmatrix} \begin{bmatrix} U_1^{(k+1)} \\ U_2^{(k+1)} \\ \vdots \\ U_m^{(k+1)} \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_m \end{bmatrix},$$

where the U_i's are vectors of values of u corresponding to the points of one of the m vertical lines of the lattice, the B_i's are vectors of constants corresponding to this partitioning and the boundary values, I is the identity matrix, and

$$A_i^{(k+1)} = \begin{bmatrix} \alpha_{i_1} & -1 & & 0 \\ -1 & \alpha_{i_2} & -1 & \\ & \ddots & \ddots & \ddots \\ & & -1 & \alpha_{i_n} \end{bmatrix}, \; i = 1, \ldots, m; \; k \geq 0,$$

with

$$\alpha_{i_j} = 4 + \frac{c_1 h^2}{1+c_2 u_p^{(k)}}, \quad p = n(i-1) + j, \; j = 1, \ldots, n.$$

Method P has iteration equations

$$A_1^{(k+1)} U_1^{(k+1)} = A_1^{(k+1)} U_1^{(k)} + \omega\{U_2^{(k)} + B_1 - A_1^{(k+1)} U_1^{(k)}\},$$

$$A_i^{(k+1)} U_i^{(k+1)} = A_i^{(k+1)} U_i^{(k)} + \omega\{U_{i-1}^{(k+1)} + U_{i+1}^{(k)} + B_i - A_i^{(k+1)} U_i^{(k)}\}, \tag{5}$$

$$i = 2, \ldots, m-1,$$

$$A_m^{(k+1)} U_m^{(k+1)} = A_m^{(k+1)} U_m^{(k)} + \omega\{U_{m-1}^{(k+1)} + B_m - A_m^{(k+1)} U_m^{(k)}\} \; k \geq 0.$$

In each iteration, m matrix equations of the form

$$A_i^{(k+1)} U_i^{(k+1)} = C_i$$

must be solved either directly or by some other means. The matrices $A_i^{(k+1)}$, here and in the next method, however, are positive definite (see [12]), and Gaussian elimination without pivoting is stable (see [13]) and gives a useful means for solving these equations.

The last method (denoted Q) is what we might descriptively call a "block overrelaxed quasilinearization" method (see [1]) or, for this problem, a "block Newton-SOR" method. In this method we linearize by using previous iterates for all u_i in f in (3), subtract $h^2 f'(u_i^{(k)}) u_i^{(k+1)}$ from the left side and $h^2 f'(u_i^{(k)}) u_i^{(k)}$ from the right side of each equation in (3), and solve by a block SOR scheme as in Method P. The system of equations for (3) can then be written as in (4), except the B_i's now have elements that are sums of constants, $h^2 f(u_i^{(k)})$ and $h^2 f'(u_i^{(k)}) u_i^{(k)}$, and the $A_i^{(k+1)}$'s have α_{i_j}'s given by

$$\alpha_{i_j} = 4 + h^2 f'(u_p^{(k)}).$$

The iteration equations are the same as those in (5).

5. An Analysis

For an analysis of (3), we shall restrict ourselves to a one-dimensional analogue of (3). Although we only consider this reduced problem, everything can be extended to (3), and the salient features we desire to consider are more easily seen in this one-dimensional setting. The one-dimensional analogue is given by

$$u_{i2} + u_{i4} - 2u_i = h^2 f(u_i), \quad i = 1, \ldots, n,$$

$$u_o = \phi(P_o), \quad u_{n+1} = \phi(P_{n+1}). \tag{6}$$

In linear problems one uses the point-Jacobi iteration method to study the Gauss-Seidal and SOR iteration methods. We shall do likewise in this section. In particular, if we apply a point-Jacobi scheme to the linearized equations in Method L, we obtain the following result.

Theorem 4. If (6) is linerized by using previous iterates for all u's in f, then the point-Jacobi iteration method diverges for

$$c_1 > \frac{2(1+c_2\alpha)^2}{h^2},$$

where

$$\alpha = \max[\phi(P_o), \phi(P_{n+1})].$$

Proof. Suppose $u = (u_1, \ldots, u_n)^T$ is the solution of (6) (whose existence is guaranteed by Theorem 2) and $u^{(k)} = (u_1^{(k)}, \ldots, u_n^{(k)})^T$ is the kth iterate in the point-Jacobi scheme for the linearized problem. Let $\varepsilon^{(k)} = u - u^{(k)}$. Matrix equations corresponding to the point-Jacobi method can be written for (6) and the linearized form of (6); they are

$$\begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_{n-1} \\ u_n \end{bmatrix} = \begin{bmatrix} 0 & 1/2 & & & 0 \\ 1/2 & 0 & 1/2 & & \\ & \cdot & \cdot & \cdot & \\ & & 1/2 & 0 & 1/2 \\ 0 & & & 1/2 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_{n-1} \\ u_n \end{bmatrix} - \frac{1}{2} \begin{bmatrix} \frac{c_1 u_1 h^2}{1+c_2 u_1} - u_o \\ \frac{c_1 u_2 h^2}{1+c_2 u_2} \\ \cdot \\ \cdot \\ \cdot \\ \frac{c_1 u_{n-1} h^2}{1+c_2 u_{n-1}} \\ \frac{c_1 u_n h^2}{1+c_2 u_n} - u_{n+1} \end{bmatrix}$$

and

$$\begin{bmatrix} u_1^{(k)} \\ u_2^{(k)} \\ \cdot \\ \cdot \\ \cdot \\ u_{n-1}^{(k)} \\ u_n^{(k)} \end{bmatrix} = \begin{bmatrix} 0 & 1/2 & & & 0 \\ 1/2 & 0 & 1/2 & & \\ & \cdot & \cdot & \cdot & \\ & & 1/2 & 0 & 1/2 \\ 0 & & & 1/2 & 0 \end{bmatrix} \begin{bmatrix} u_1^{(k-1)} \\ u_2^{(k-1)} \\ \cdot \\ \cdot \\ \cdot \\ u_{n-1}^{(k-1)} \\ u_n^{(k-1)} \end{bmatrix} - \frac{1}{2} \begin{bmatrix} \frac{c_1 u_1^{(k-1)} h^2}{1+c_2 u_1^{(k-1)}} - u_o \\ \frac{c_1 u_2^{(k-1)} h^2}{1+c_2 u_2^{(k-1)}} \\ \cdot \\ \cdot \\ \cdot \\ \frac{c_1 u_{n-1}^{(k-1)} h^2}{1+c_2 u_{n-1}^{(k-1)}} \\ \frac{c_1 u_n^{(k-1)} h^2}{1+c_2 u_n^{(k-1)}} - u_{n+1} \end{bmatrix}$$

Subtracting one of these matrix equations from the other gives

$$\varepsilon^{(k)} = M_k \, \varepsilon^{(k-1)}, \quad k > 0,$$

where

$$M_k = \begin{bmatrix} \alpha_1 & 1/2 & & & 0 \\ 1/2 & \alpha_2 & 1/2 & & \\ & \ddots & \ddots & \ddots & \\ & & 1/2 & \alpha_{n-1} & 1/2 \\ 0 & & & 1/2 & \alpha_n \end{bmatrix}$$

and

$$\alpha_i = - \frac{c_1 h^2}{2(1+c_2 u_i)(1+c_2 u_i^{(k-1)})} .$$

Let $\rho(M)$ denote the spectral radius of matrix M (i.e., $\rho(M) = \max_i |\lambda_i|$, where the λ_i's are the eigenvalues of M).

Now

$$\rho(M_k) \geq \max_i |\mu_i|,$$

where the μ_i's are eigenvalues of any of the principal submatrices of M_k.

Thus

$$\rho(M_k) \geq \max |\alpha_i| \geq \frac{c_1 h^2}{2(1+c_2 \alpha)^2} .$$

If

$$\frac{c_1 h^2}{2(1+c_2 \alpha)^2} > 1$$

or

$$c_1 > \frac{2(1+c_2\alpha)^2}{h^2},$$

then

$$\rho(M_k) > 1, \quad k > 0.$$

Suppose the method converges for

$$c_1 > \frac{2(1+c_2\alpha)^2}{h^2}.$$

Then M_k converges to a matrix M as $k \to \infty$ and $\rho(M) > 1$. But $\rho(M) > 1$ implies divergence, and so leads to a contradiction of the assumption of convergence. Hence the point-Jacobi method diverges for

$$c_1 > \frac{2(1+c_2\alpha)^2}{h^2}.$$

A similar argument gives the following theorem.

Theorem 5. If (6) is linearized by using previous iterates for all u's in f, then all iteration methods diverge for

$$c_1 > \frac{4(1+c_2\alpha)^2}{h^2}.$$

Proof. Suppose $u = (u_1, \ldots, u_n)^T$ is the solution of (6) and $u^{(k)} = (u_1^{(k)}, \ldots, u_n^{(k)})^T$ is the kth iterate in any iteration scheme for the linearized problem. Let $\varepsilon^{(k)} = u - u^{(k)}$. Subtracting the corresponding equations for u in (6) and the ones for $u^{(k)}$ in (6), we get the matrix equation

$$\begin{bmatrix} 2 & -1 & & 0 \\ -1 & 2 & -1 & \\ & \ddots & \ddots & \ddots \\ & & -1 & 2 & -1 \\ 0 & & & -1 & 2 \end{bmatrix} \varepsilon^{(k)} = \begin{bmatrix} \beta_1 & & & 0 \\ & \beta_2 & & \\ & & \ddots & \\ & & & \beta_{n-1} \\ 0 & & & & \beta_n \end{bmatrix} \varepsilon^{(k-1)}, \quad k > 0,$$

where

$$\beta_i = \frac{c_1 h^2}{(1+c_2 u_i)(1+c_2 u_i^{(k-1)})} .$$

This equation can be written in the form

$$M_k^{-1} \varepsilon^{(k)} = \varepsilon^{(k-1)}$$

where

$$M_k^{-1} = \begin{bmatrix} -2\beta_1^{-1} & \beta_1^{-1} & & 0 \\ \beta_2^{-1} & -2\beta_2^{-1} & \beta_2^{-1} & \\ & \cdot \quad \cdot \quad \cdot & & \\ & \beta_{n-1}^{-1} & -2\beta_{n-1}^{-1} & \beta_{n-1}^{-1} \\ 0 & & \beta_n^{-1} & -2\beta_n^{-1} \end{bmatrix} .$$

Now M_k^{-1} is irreducibly diagonally dominant, and hence non-singular (see [12]). The eigenvalues of M_k are just the reciprocals of the eigenvalues of M_k^{-1}.

But

$$\rho(M_k^{-1}) \leq 4 \max_{1 \leq i \leq n} |\beta_i^{-1}| .$$

Therefore,

$$\rho(M_k) \geq \frac{1}{4 \max_{1 \leq i \leq n} |\beta_i^{-1}|} \geq \frac{c_1 h^2}{4(1+c_2\alpha)^2} .$$

If

$$\frac{c_1 h^2}{4(1+c_2\alpha)^2} > 1$$

or

$$c_1 > \frac{4(1+c_2\alpha)^2}{h^2},$$

then

$$\rho(M_k) > 1,$$

which implies divergence as in the above proof.

Remark. These theorems also hold for $c_2 = 0$, which corresponds to the linear problem $\Delta u - c_1 u = 0$. Observe then that the finite difference analogue for this linear problem has a divergence criterion $c_1 > 4/h^2$ for the linearization scheme suggested by Theorems 4 and 5. This is illustrated in an example in Table 2 of the next section.

Theorems 4 and 5 indicate that iteration schemes with complete linearization using previous iterates in f fail for certain values of c_1 and c_2. Examples are given in the next section to illustrate this. A similar analysis for the linearization of Method M for (6) shows that one can expect convergence in those cases where Method L fails.

We now explore the point-Jacobi iteration method for the linearization of Method M for (6), different than the one just mentioned above. The point-Jacobi method here has matrix form

$$u^{(k+1)} = J_k u^{(k)} + J_k b,$$

where

$$J_k = \begin{bmatrix} 0 & \gamma_1 & & 0 \\ \gamma_2 & 0 & \gamma_2 & \\ & \ddots & \ddots & \ddots \\ & \gamma_{n-1} & 0 & \gamma_{n-1} \\ 0 & & \gamma_n & 0 \end{bmatrix}$$

and

$$\gamma_i = \left(2 + \frac{c_1 h^2}{1+c_2 u_i^{(k)}} \right)^{-1} .$$

To get some feeling for what happens for different values of c_1, c_2, and h, we suppose the iterates $u^{(k)}$ are near the solution of u of (6), and treat $J_k b$ essentially as a constant vector. In this case, J_k corresponds to the iteration matrix. Studying J_k, we observe that for $c_1 >> c_2$ and small, $c_1 = c_2$ and small, and $c_2 >> c_1$, J_k approximates a matrix T of Toeplitz form

$$T = \begin{bmatrix} 0 & \beta & & \\ \beta & 0 & \beta & \\ & \ddots & \ddots & \ddots \\ & \beta & 0 & \beta \\ & & \beta & 0 \end{bmatrix},$$

with $\beta = 1/2$. The matrix T has eigenvalues

$$\lambda_\nu = 2\beta \cos\frac{\nu\pi}{n+1}, \quad \nu = 1, \ldots, n,$$

where n is the order of T (see [5]). For $\beta = 1/2$, the spectral radii of T for $h = 0.1$ and $h = 0.02$, the values of h considered in the next section, are .95 and .99. In the case $c_1 >> c_2$ and large, J_k approximates T with very small β for $h = 0.1$ and $\beta = 1/6$ for $h = 0.02$. We also note that as $h \to 0$, J_k approximates T with $\beta = 1/2$ for all fixed values of c_1 and c_2, and so $\rho(T) \to 1$, since its order is increasing as $h \to 0$.

Using the formula for obtaining the optimum overrelaxation factor, ω_b, for linear problems (see [12]),

$$\omega_b = \frac{2}{1 + \sqrt{1 - \rho^2(J)}},$$

where $\rho(J)$ is the spectral radius of the point-Jacobi iteration matrix, and the above approximations, we summarize our findings or "predictions" in the table below.

TABLE 1. Predictions.

Values of c_1 and c_2	h = 0.1		h = 0.02	
	Approx. $\rho(J_k)$	Approx. ω_b	Approx. $\rho(J_k)$	Approx. ω_b
$c_1 >> c_2$, small	.95	1.5	.99	1.9
$c_1 = c_2$, small	.95	1.5	.99	1.9
$c_2 >> c_1$	.95	1.5	.99	1.9
$c_1 >> c_2$, large	.01	1.0	.33	1.1

6. Numerical Results

Many numerical cases were solved for (3) with f given by (2), R the unit square, $\phi = x^2 + 2y^2$, and h = 0.1 or 0.02. These cases are summarized in the tables below. Included in these tables are the mesh size, h, the approximate optimum relaxation factor, ω_b, the number of iterations for convergence, I, and the computer execution time (in minutes), T, for each method for each case. All of these were run on an IBM 360, Model 50.

TABLE 2. Results for Methods L, M, N and Q with h = 0.1

h = 0.1		Method L			Method M			Method N			Method Q		
c_1	c_2	ω_b	I	T	ω_b	I	T	ω_b	I	T	ω_b	I	T
10^{-1}	10^{-1}	1.5	17	.013	1.5	17	.013	1.5	17	.016	1.4	12	.019
10^{4}	10^{-1}	DIVERGED			1.0	3	.008	1.0	3	.009	1.0	3	.010
10^{4}	10^{2}	DIVERGED			1.5	13	.012	1.3	8	.011	1.2	8	.015
10^{4}	10^{4}	1.5	17	.013	1.5	21	.014	1.5	19	.017	1.4	14	.022
1	1	1.5	17	.014	1.5	17	.014	1.5	17	.016	1.4	12	.020
10^{2}	10^{2}	1.5	17	.014	1.5	18	.014	1.5	17	.016	1.4	13	.021
10^{2}	10^{4}	1.5	17	.014	1.5	17	.013	1.5	17	.016	1.4	13	.020
10^{2}	10^{6}	1.5	17	.014	1.5	17	.013	1.5	17	.016	1.4	13	.022
10^{6}	10^{6}	1.5	17	.013	1.6	24	.015	1.5	21	.018	1.4	16	.023
10^{6}	10^{2}	DIVERGED			1.0	4	.008	1.0	3	.009	1.0	3	.010
10^{-1}	10^{4}	1.5	17	.013	1.5	17	.014	1.5	17	.016	1.4	12	.019
10^{-4}	10^{-4}	1.5	17	.014	1.5	17	.013	1.5	17	.015	1.4	12	.020
10^{-4}	1	1.5	17	.013	1.5	17	.013	1.5	17	.016	1.4	12	.019
1	10^{-4}	1.5	17	.014	1.5	17	.014	1.5	17	.016	1.4	12	.020
$4x10^{2}$	0	DIVERGED			1.1	5	.009	1.1	5	.009	1.0	5	.012

TABLE 3. Results for Method P with h = 0.1

h = 0.1		Method P		
c_1	c_2	ω_b	I	T
10^4	10^{-1}	1.0	3	.009
10^4	10^4	1.5	16	.021
10^2	10^6	1.4	13	.019
1	1	1.4	12	.018

TABLE 4. Results for Methods L, M, N, and Q with h = 0.02.

h = 0.02		Method L			Method M			Method N			Method Q		
c_1	c_2	ω_b	I	T	ω_b	I	T	ω_b	I	T	ω_b	I	T
10^4	10^{-4}	DIVERGED			1.1	9	.216	1.1	9	.257	1.0	7	.335
1	1	1.9	110	1.298	1.9	110	1.201	1.9	110	1.724	1.8	102	3.106
10^2	10^6	1.9	110	1.301	1.9	110	1.208	1.9	110	1.735	1.8	104	3.492
10^{-4}	10^{-4}	1.9	110	1.350	1.9	110	1.206	1.9	110	1.719	1.8	103	3.136
1	10^{-4}	1.9	110	1.303	1.9	110	1.207	1.9	110	1.715	1.8	98	2.979
10^{-4}	1	1.9	110	1.297	1.9	110	1.202	1.9	110	1.732	1.8	103	3.136

TABLE 5. Results for Method P with h = 0.02.

h = 0.02		Method P		
c_1	c_2	ω_b	I	T
10^4	10^{-1}	1.0	8	.323
10^2	10^6	1.8	104	2.609
1	1	1.8	102	2.582

We now indicate some conclusions from these numerical calculations. We first observe that the "predictions" in Table 1 of Section 5 are indeed accurate for all cases studied. When $c_1 >> c_2$ and large, the divergence indicated in Theorems 4 and 5, and the remark following these theorems, is obtained for all such numerical examples. In those cases where Method L converges, Method L and M have almost identical iteration characteristics (a comparison of linear methods). Also, Methods M and N parallel each other except for the slightly increased time factor in Method N (a comparison of a linear with Newton-type method). The block methods, Methods P and Q, are iteratively faster than the point methods, Methods L, M, and N, as expected (see [12]), but require more time; this is due to the added time needed to solve the block equations by Gaussian elimination.

Methods M, N, P, and Q all effectively solved (3) with f given by (2). However, Method M and N are simpler and less time consuming, so one might consider them slightly better than the other methods. Clearly, Method L could not be considered a good "general" method even though it effectively handled many cases. Other interesting observations can be obtained from the tables, but we shall leave these to the reader's interest.

Direct methods for solving the equations in (3) for Methods L, M, and Q by a "band matrix inversion" technique suggested in a paper by Martin and Wilkinson [11] were tried for h = 0.1. A table of results for such calculations is given below.

TABLE 6. Results for Direct Methods.

h = 0.1		Direct L			Direct M			Direct Q		
c_1	c_2	ω_b	I	T	ω_b	I	T	ω_b	I	T
10^{-1}	10^{-1}	1.0	3	.028	1.0	2	.036	1.0	2	.036
10^{4}	10^{-1}	DIVERGED			1.0	3	.053	1.0	3	.054
10^{4}	10^{2}	DIVERGED			1.7	11	.170	1.0	7	.116
10^{4}	10^{4}	1.0	3	.029	1.1	9	.140	1.0	7	.121
1	1	1.0	3	.029	1.0	3	.055	1.0	3	.052
10^{2}	10^{2}	1.0	3	.029	1.1	5	.080	1.0	4	.067
10^{2}	10^{4}	1.0	2	.026	1.0	4	.069	1.0	3	.052
10^{2}	10^{6}	1.0	2	.026	1.0	3	.051	1.0	3	.053
10^{6}	10^{6}	1.0	3	.029	1.0	15	.239	1.0	9	.143
10^{6}	10^{2}	DIVERGED			1.0	4	.069	1.0	3	.052
10^{-1}	10^{4}	1.0	2	.026	1.0	3	.053	1.0	3	.052
10^{-4}	10^{-4}	1.0	2	.025	1.0	2	.038	1.0	2	.039
10^{-4}	1	1.0	2	.027	1.0	2	.038	1.0	2	.037
1	10^{-4}	1.0	3	.028	1.0	2	.039	1.0	2	.037

The direct methods are iteratively faster, but overall took a great deal more time (a factor of 2 or more) than the point and block methods. Furthermore, direct methods become less and less practical as the mesh is refined. Observe that the divergence criterion of Theorems 4 and 5 is again illustrated for the linearized equations of Method L.

The equations in (3) are quadratic in u_i. One might therefore try to solve them iteratively using the quadratic formula. This is a successful method of solution, but double precision is necessary, and this method is more time-consuming than the other methods.

There were numerical instabilities for large values of the overrelaxation factor, ω, in those cases where $c_1 >> c_2$ and large.

All calculations above used a zero initial guess and the convergence criterion

$$\frac{\max_i |u_i^{(k+1)} - u_i^{(k)}|}{\max_i |u_i^{(k+1)}|} < \varepsilon ,$$

where $\varepsilon = .002$ for $h = 0.1$ and $\varepsilon = .00002$ for $h = 0.02$.

Part II. A Dirichlet Problem for a System of Non-Linear Partial Differential Equations

1. Introduction.

In this part of the report, we consider the non-linear boundary-value problem

$$
\begin{aligned}
&\Delta u = f(u,v),\ u,\ v \geq 0,\ \text{in } R,\\
&\Delta v = g(u,v),\ u,\ v \geq 0,\ \text{in } R,\\
&u = \phi,\ v = \psi,\ \text{on } R',
\end{aligned}
\tag{7}
$$

where R, R', and Δ are as in Part I, ϕ and ψ are non-negative continuous functions and

$$
f(u,v) = \frac{c_1 u}{1+c_2 v+c_3 u} \tag{8}
$$

and

$$
g(u,v) = -\frac{c_4 u}{1+c_2 v+c_3 u}\ , \tag{9}
$$

where c_i, $i=1, \ldots, 4$, are positive constants.

We describe the finite-difference analogue of (7) in Section 2. Section 3 contains iterative methods for solving this approximate problem. Other related results and an analysis of these and the results of Part I are given in Sections 4 and 5. In Section 6 there are numerical calculations to illustrate the iteration methods and a variety of examples of the approximate problem for (7).

2. Statement of the Problem.

For convenience, we shall use the notation of Section 2 of Part I. We define operators M_h and N_h by

$$M_h(\xi_i,\eta_i) = \Delta_h[\xi_i] - f(\xi_i,\eta_i)$$

and

$$N_h(\xi_i,\eta_i) = \Delta_h[\eta_i] - g(\xi_i,\eta_i).$$

The boundary-value problem for some fixed R_h consists of determining functions u and v defined on $\overline{R}_h$ such that

$$M_h[u_i,v_i] = 0,\ u_i,\ v_i \geq 0,\ i = 1,\ \ldots,N;$$

$$N_h[u_i,v_i] = 0,\ u_i,\ v_i \geq 0,\ i = 1,\ \ldots,N;$$

(10)

$$u_j = \phi_j,\ j = N+1,\ \ldots,M;$$

$$v_j = \psi_j,\ j = N+1,\ \ldots,M.$$

3. Iteration Methods.

The iteration methods of this part of the report are the same as Methods L, M, N, and Q of Part I, except they are extended to handle the system instead of just the single equation. We shall denote these comparable methods by L', M', N', and Q'.

Method L'

We linearize (10) at the i-th point by taking all previous iterates for u_i and v_i in f, and the new iterate $u_i^{(k+1)}$ and previous iterate $v_i^{(k)}$ in g. We first solve the equation with the operator M_k at the i-th point, and then solve the equation with the operator N_k at the i-th point, using the iteration equations of Method L for both. The iteration equations for this method are

$$4u_i^{(k+1)} = 4u_i^{(k)} + \omega\{u_{i4}^{(k+1)} + u_{i3}^{(k+1)} + u_{i2}^{(k)} + u_{i1}^{(k)} - h^2 f(u_i^{(k)},v_i^{(k)}) - 4u_i^{(k)}\},$$

$$4v_i^{(k+1)} = 4v_i^{(k)}+\omega\{v_{i4}^{(k+1)}+v_{i3}^{(k+1)}+v_{i2}^{(k)}+v_{i1}^{(k)}-h^2g(u_i^{(k+1)},v_i^{(k)})-4v_i^{(k)}\} ,$$

$$i=1,\ldots,N;\ k \geq 0 .$$

Method M'

We linearize the system (10) as in Method L' except we replace $u_i^{(k)}$ by $u_i^{(k+1)}$ in the numerator of f. The iteration equations for v are the same as in Method L' and the iteration equations for u are given by

$$u_i^{(k+1)} = u_i^{(k)} + \omega\left\{\frac{u_{i4}^{(k+1)}+u_{i3}^{(k+1)}+u_{i2}^{(k)}+u_{i1}^{(k)}}{4+\dfrac{c_1h^2}{1+c_2v_i^{(k)}+c_3u_i^{(k)}}} - u_i^{(k)}\right\} ,$$

$$i=1,\ldots,N;\ k \geq 0.$$

Method N'

Method N' is the same as Method N except we replace the single equation in u with two equations in u and v and solve the u-equation at the i-th point and then the v-equation there. Its defining equations are

$$u_i^{(k+1)} = u_i^{(k)} - \omega\,\frac{4u_i^{(k)}-[u_{i4}^{(k+1)}+u_{i3}^{(k+1)}+u_{i2}^{(k)}+u_{i1}^{(k)}-h^2f(u_i^{(k)},v_i^{(k)})]}{4+h^2\frac{\partial f}{\partial u}(u_i^{(k)},v_i^{(k)})}$$

and

$$v_i^{(k+1)} = v_i^{(k)} - \omega\,\frac{4v_i^{(k)}-[v_{i4}^{(k+1)}+v_{i3}^{(k+1)}+v_{i2}^{(k)}+v_{i1}^{(k)}-h^2g(u_i^{(k+1)},v_i^{(k)})]}{4+h^2\frac{\partial g}{\partial v}(u_i^{(k+1)},v_i^{(k)})} ,$$

$$i=1,\ldots,N;\ k \geq 0.$$

Method Q'

Method Q' is analogous to Method Q except we have two block equations to solve at each step, one corresponding to a vertical line of values for u and the other to a vertical line of values of v. The equations are just those indicated in (5) with $f(u_i^{(k)})$ and $f'(u_i^{(k)})$ replaced by $f(u_i^{(k)},v_i^{(k)})$ and $\frac{\partial f}{\partial u}(u_i^{(k)},u_i^{(k)})$ for the equations in u, and the u_i, $f(u_i^{(k)})$ and $f'(u_i^{(k)})$ replaced by v_i, $g(u_i^{(k+1)}, v_i^{(k)})$ and $\frac{\partial g}{\partial v}(u_i^{(k+1)}, v_i^{(k)})$ for the equations in v. Again, we solve the u-equation and then the v-equation before proceeding to the next vertical line of values.

4. Other Results.

Lick and Coleman [8] have considered problem (1) with f given by

$$f(u) = -\frac{c_1}{1+c_2u}. \tag{11}$$

Included in their paper are existence, uniqueness, and convergence theorems and numerical results for the finite-difference analogue of this problem. We summarize their numerical results in the table below for Methods L (Methods L and M are equivalent here), N, and P. These tables contain the mesh size, h, the approximate optimum relaxation factor, ω_b, and the number of iterations for convergence, I. The convergence criterion used was

$$\max_i \left| u_i^{(k+1)} - u_i^{(k)} \right| < \varepsilon ,$$

where $\varepsilon = 10^{-4}$. The boundary condition was given by $\phi = xy$ on R', where R is the unit square.

In those cases where c_1 was large, there was numerical instability for Methods L and P when a large ω was used (for $h = 0.1$, $\omega > 1.6$; for $h = 0.02$, $\omega > 1.9$).

TABLE 7. Results of Lick and Coleman for h = 0.1.

h = 0.1		Method L		Method N		Method P	
c_1	c_2	ω_b	I	ω_b	I	ω_b	I
10^{-4}	10^{-4}	1.6	23	1.6	23	1.4	16
10^{-4}	1	1.6	23	1.6	23	1.4	16
1	10^{-4}	1.6	23	1.6	23	1.4	16
1	1	1.5	23	1.5	23	1.4	15
10^4	1	1.4	30	1.5	25	1.2	23
1	10^4	1.5	23	1.6	24	1.4	16
10^4	10^4	1.5	32	1.5	26	1.4	27

TABLE 8. Results of Lick and Coleman for h = 0.02.

h = 0.02		Method L		Method N		Method P	
c_1	c_2	ω_b	I	ω_b	I	ω_b	I
10^{-4}	10^{-4}	1.9	104	1.9	104	1.8	82
10^{-4}	1	1.9	104	1.9	104	1.8	82
1	10^{-4}	1.9	104	1.9	104	1.8	83
1	1	1.9	104	1.9	104	1.8	80
10^4	1	1.8	116	1.9	108	1.7	98
1	10^4	1.9	105	1.9	106	1.8	82
10^4	10^4	1.8	211	1.9	118	1.8	111

5. An Analysis.

We now try to analyze (10) using the results of Part I and the results of Lick and Coleman.

We first observe that f(u,v) can be written in the form

$$f(u,v) = \frac{c_1' u}{1+c_2' u} ,$$

where

$$c_1' = \frac{c_1}{1+c_2 v}$$

and

$$c_2' = \frac{c_3}{1+c_2 v} .$$

Then f(u,v) can be considered to be of form (2). We see then that c_1' and c_2' vary as c_1 and c_3 in (10). Similarly, g(u,v) can be written as

$$g(u,v) = - \frac{c_3'}{1+c_4' v} ,$$

where

$$c_3' = \frac{c_4 u}{1+c_3 u}$$

and

$$c_4' = \frac{c_2}{1+c_3 u} .$$

Thus g(u,v) can be thought of as in form (11). We see then that c_3' and c_4' vary as $c_4 u$ and c_2. The above discussion gives us a way of examining the system in (7) in terms of the single equations in (1) and (11).

Lick and Coleman's results of Tables 7 and 8 (corresponding to values of c_3' and c_4') indicate that the second equation in (10) should cause no difficulties for the numerical solution by Method L' or N'. However, the analysis of Section 5 of Part I (corresponding to c_1' and c_2') indicates that Method L' should fail for $c_1 >> c_3$ and large (see Table 9), while the other methods (M', N', and Q') should be successful. Also, the values in the tables coupled with the discussion above leads one to expect rapid convergence and ω_b near 1.5 for the point methods and 1.4 for the block method for most cases when h = 0.1.

6. Numerical Results.

Many numerical cases were solved for (10) with f given by (8) and g by (9), R the unit square, $\phi = x^2+2y^2$, $\psi = xy$, and h = 0.1. These cases are summarized in the table below. Included in this table are the mesh size, h, the approximate optimum relaxation factor, ω_b, the number of iterations for convergence, I, and the computer execution time (in minutes), T, for each method for each case. All of these were run on an IBM 360, Model 50.

The following calculations had convergence criteria

$$\frac{\max_i \left| u_i^{(k+1)} - u_i^{(k)} \right|}{\max_i \left| u_i^{(k+1)} \right|} < \varepsilon_u$$

and

$$\frac{\max_i \left| v_i^{(k+1)} - v_i^{(k)} \right|}{\max_i \left| v_i^{(k+1)} \right|} < \varepsilon_v ,$$

with $\varepsilon_u = \varepsilon_v = .002$.

TABLE 9. Results for Methods L', M', N', and Q' for h = 0.1.

h = 0.1				Method L'			Method M'			Method N'			Method Q'		
c_1	c_2	c_3	c_4	ω_b	I	T	ω_b	I	T	ω_b	I	T	ω_b	I	T
1	1	10^5	10^5	1.5	17	.027	1.5	17	.026	1.5	17	.032	1.4	12	.038
1	10^5	10^5	10^5	1.5	17	.026	1.5	17	.027	1.5	17	.031	1.4	12	.040
10^5	10^5	10^5	10^5	1.5	17	.027	1.5	18	.027	1.5	19	.033	1.4	14	.045
10^{-4}	1	10^{-4}	1	1.5	18	.027	1.5	18	.028	1.5	18	.034	1.4	13	.044
10^{-4}	1	1	1	1.5	17	.029	1.5	17	.027	1.5	17	.034	1.4	12	.041
1	1	10^{-4}	1	1.5	18	.029	1.5	18	.028	1.5	18	.033	1.4	13	.044
1	1	1	1	1.5	17	.029	1.5	17	.028	1.5	17	.036	1.4	12	.040
1	1	1	10^5	1.3	28	.036	1.3	28	.035	1.5	19	.033	1.5	13	.041
1	10^5	1	1	1.5	17	.028	1.5	17	.029	1.5	19	.037	1.4	13	.043
10^5	1	1	1	DIVERGED			1.5	17	.027	1.5	17	.033	1.4	13	.044
10^5	1	1	10^5	DIVERGED			1.4	20	.030	1.5	18	.034	1.4	12	.040
10^5	10^5	1	1	DIVERGED			1.5	21	.031	1.5	19	.033	1.4	14	.044

All cases behaved as expected from the analysis of Section 5. For the cases where Method L' converged, Methods L' and M' were almost identical. Methods M' and N' again essentially parallel each other except for the slight increase in time for Method N'. Method Q' (block method) was iteratively faster, but again took more time. Some numerical instability was observed and is summarized in the following table.

TABLE 10. Numerical Instability of the System.

Case \ Method	L'	M'	N'	Q'
$c_1 = c_2 = c_3 = 1$, $c_4 = 10^5$	$\omega \geq 1.6$	$\omega > 1.5$	None	None
$c_1 = 10^5$, $c_2 = c_3 = c_4 = 1$	DIVERGED	$\omega \geq 1.7$	$\omega \geq 1.7$	$\omega \geq 1.8$
$c_1 = c_4 = 10^5$, $c_2 = c_3 = 1$	DIVERGED	$\omega \geq 1.8$	$\omega \geq 1.8$	$\omega \geq 1.8$

From simplicity, time, and effectiveness considerations, we conclude that Methods M' and N' are better than the other methods considered in this report.

Systems like (7) frequently arise in various physical problems. If the Dirichlet conditions of (7) are replaced by mixed boundary conditions, (7) describes a system that arises in corrosion-diffusion problems. In particular, examples of such problems and their numerical solutions can be found in [9] and [10].

ACKNOWLEDGEMENT

A portion of the work for this report was completed at the University of Tennessee and the Applied Mathematics Department, Brookhaven National Laboratory.

REFERENCES

1. Bellman, R. E., and Kalabo, R. E., Quasilinearization and Nonlinear Boundary-Value Problems, American Elsevier Co., New York (1965).
2. Bers, L., "On Mildly Nonlinear Partial Differential Equations of Elliptic Type", Jour. Res. Nat. Bur. Stand., Vol. 51, No. 5, 229-236 (1953).
3. Courant, R., and Hilbert, D., Methods of Mathematical Physics, Vol. 2, Interscience, New York (1962).
4. Greenspan, D., Introductory Numerical Analysis of Elliptic Boundary-Value Problems, Harper and Row, New York (1965).
5. Grenander, U., and Szegö, G., Toeplitz Forms and Their Applications, University of California Press, Berkeley (1958).
6. Lick, D. W., "Existence-Uniqueness Theorems for Nonlinear Dirichlet Problems", to appear, Duke Math. Jour.
7. Lick, D. W., "A Note on a Nonlinear Dirichlet Problem", to appear.
8. Lick, D. W., and Coleman, P. R., "The Approximate Solutions of a Quasilinear Dirichlet Problem", to appear.
9. Lick, D. W., and Tunstall, J. N., "A Numerical Solution of a System of Ordinary and Nonlinear Partial Differential Equations", Oak Ridge Gaseous Diffusion Plant, Oak Ridge, Tennessee, Oct. 30, 1967, (K-1724), (unclassified).
10. Lick, D. W., and Tunstall, J. N., "The Solution of a Diffusion-Convective Corrosion Problem", to appear.
11. Martin, R. S., and Wilkinson, J. H., "Symmetric Decomposition of Positive Definite Band Matrices", Numer. Math., 7, 355-361 (1965).
12. Varga, R. S., Matrix Iterative Analysis, Prentice-Hall, New Jersey (1962).
13. Wilkinson, J. H., "Error Analysis of Direct Methods of Matrix Inversion", J. ACM, 8, 281-320 (1961).