RELAXATION TECHNIQUES LEADING TO NUMERICAL SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS

Warren Doud

May 1969

RELAXATION TECHNIQUES LEADING TO NUMERICAL SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS

BY

Warren Doud

May, 1969

Approved:

Lewis E. Batson, PhD
Associate Professor of Mathematics

David Andrew, PhD
Professor of Mathematics

Ossie J. Huval, PhD
Professor of Mathematics

James Oliver, PhD
Dean of the Graduate School

UNIVERSITY OF SOUTHWESTERN LOUISIANA

RELAXATION TECHNIQUES LEADING TO NUMERICAL SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS

A Thesis
Presented to
The Department of Mathematics
The University of Southwestern Louisiana

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

---

by
Warren Doud
May, 1969

ACKNOWLEDGEMENT

This acknowledgement is a word of thanks to Dr. Lewis Batson, who has a contagious enthusiasm for scholarship, and who encouraged the author to discover the enjoyment to be found in finding buried treasure.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

It has been suggested that numerical analysis involves the development and evaluation of methods for computing required numerical results from given numerical data, making the subject a part of the modern study of information processing. In the language of information processing, then, the given data are the input information, the required results are the output information, and the problem solving method -- the method of computation -- is known as the algorithm.

Admittedly this description is oriented towards applications and focuses effort on the search for algorithms. This is not to derogate those who research the supporting theory of numerical analysis, since theory has an intrinsic interest of its own -- it is often beautiful mathematics. Furthermore, theory is very important because it contributes to the search for more and better algorithms.

Quite often, several algorithms are available for producing desired results in an application; and normally that is chosen which gives answers most quickly and accurately.

Numerous methods for the solution of ordinary and partial differential equations appear in mathematics literature. Notable among these are those which are

familiar to the undergraduate student of numerical analysis, namely, those of Euler, Keun, Runge, Kutta, et al. Of these the most popular algorithm seems presently to be the Runge-Kutta technique because of its speed and accuracy in digital computer applications.

Less widely used is the method which is the subject of this thesis, an algorithm first applied by Sir Richard Southwell in 1935[1], when he was professor of Engineering Science at Oxford, and which has since been developed and expanded upon by Southwell and his colleagues [1][6], by F. S. Shaw [21], by D. N. Allen [20] of London University, and by others also mentioned in the bibliography. The technique is known as the method of relaxation for the solution of simultaneous linear algebraic equations.

Designed originally to facilitate the analysis of positioned pin-jointed space frameworks[2], relaxation methods have been extended to numerous branches of applied mathematics. The main interest in these methods stems from their usefulness in making possible the approximate solution of practical physics problems.

Examples will first be given in this paper to

---

[1] D.N.deG. Allen, Relaxation Methods (New York: McGraw-Hill, 1954), Preface.

[2] F.S.Shaw, Relaxation Methods (New York:Dover Publications, 1953), Introduction.

illustrate the use of relaxation in approximating solutions to systems of simultaneous linear algebraic equations. Along with the examples will be given a theorem on the improvement of solutions under certain conditions.

Then this technique will be extended to the solution of partial differential equations which have been approximated by systems of finite difference equations.

In the last section, the tediousness of hand calculation will be apparent from examples worked in detail; and computer calculation will be demonstrated, first by the Southwell methods and then by a method of successive point relaxation apparently untried previously.

CHAPTER II

RELAXATION APPLIED TO SIMULTANEOUS LINEAR EQUATIONS

All applications of the relaxation method have an underlying computational process which can most easily be described in relation to the solution of a system of linear algebraic equations.

To illustrate in easy form the basic operation, the following example is used:

$$\text{(II-1)} \qquad \begin{aligned} 3x_1 + 2x_2 - x_3 &= 20 \\ x_1 + 6x_2 + 4x_3 &= 27 \\ 2x_1 - x_2 + 9x_3 &= 16 \end{aligned}$$

a set of equations especially designed for rapid convergence by relaxation. By conventional means the solution is found to be $x_1 = 5$, $x_2 = 3$, and $x_3 = 1$.

The relaxation process begins by making an initial guess of the values of the unknowns, and equations (II-1) are rewritten,

$$\text{(II-2)} \qquad \begin{aligned} 3x_1 + 2x_2 - x_3 - 20 &= r_1 \\ x_1 + 6x_2 + 4x_3 - 27 &= r_2 \\ 2x_1 - x_2 + 9x_3 - 16 &= r_3, \end{aligned}$$

where r_1, r_2, and r_3 are the respective values of each left member after an initial approximation has been substituted for each unknown. These values are called <u>residuals</u>, and we have:

RESIDUAL[3]...When an equation of a linear system of n equations in n unknowns has all terms on the left of the equal sign, then the value of the left member of the equation when approximations have been substituted for all of the unknowns is called the residual of the equation, denoted r_i, $(i = 1,2,\ldots,n)$.

It is apparent that if the solution values given above were substituted in (II-2) then $r_1 = r_2 = r_3 = 0$; and indeed this condition is made the goal of the relaxation process, namely, to modify the initial estimate with subsequent "guesses" in such a way that the residuals are forced to be zero, or as close to zero as approximate techniques will allow. As the pattern of these modifications develops, the procedure becomes much more of a reasoned process than inefficient trial and error.

RELAXATION PROCESS...a process for solving systems of linear equations in which an initial approximation to the solution is systematically altered with the purpose of reducing equation residuals is called a relaxation process.

The initial approximation is left to the solver, and it is just as well to choose $x_1 = x_2 = x_3 = 0$, leaving residuals of $r_1 = -20$, $r_2 = -27$, and $r_3 = -16$. When these are reduced to zero, the variables will coincidentally have

[3]Ibid., p. 7.

attained their correct values.

Each step in the calculation makes use of a multiple application of a basic unit operation and records in a table the operation used, the extent to which it is used, and the consequent effect on the residuals. In this example, the basic unit operations are $\Delta x_1 = 1$, $\Delta x_2 = 1$, and $\Delta x_3 = 1$, where the symbol Δ is used to denote an increment of addition. By reference to equations (II-2) and Fig. 1, the effect of these unit operations on the residuals is seen.

The reduction of residuals is achieved by applying multiples of the basic unit operations. Repeated use of these operations accomplishes the eventual complete reduction of residuals. When the residual highest in absolute value is relaxed in each step, the process is called the STANDARD RELAXATION PROCESS -- a relaxation process in which in each step the residual highest in absolute value is reduced is called the standard relaxation process.

At the start of the relaxation the residual highest in absolute value (also called the "largest" residual) is r_2=-27. The largest effect on r_2 occurs when Δx_2 is used. Reduction of r_2 from -27 to -3 uses the multiple $\Delta x_2 = 4$, at which time r_1 is increased by 8 to -12, and r_3 is decreased by 4 to -20 (see Fig. 2). The resulting new values of the residuals are also seen to be tabulated in Figure 2 with operation 2; and

operation number 3 is used to reduce the now largest residual r_3 by using basic unit operator Δx_3 two times.

Residual r_1 now deviates the most widely from the desired goal and Δx_1 is used 5 times in operation 4. In step 5, make note of the fact that the first four steps have gone too far in adjusting r_2. No change in procedure is required, however, and incrementing x_2 by -2 improves the picture.

OPERATION	Δr_1	Δr_2	Δr_3
$\Delta x_1 = 1$	3	1	2
$\Delta x_2 = 1$	2	6	-1
$\Delta x_3 = 1$	-1	4	9

Figure 1[4]

Op No	OPER.	r_1	r_2	r_3	R_T
1	All x =0	-20	-27	-16	63
2	$\Delta x_2 = 4$	-12	- 3	-20	35
3	$\Delta x_3 = 2$	-14	5	- 2	21
4	$\Delta x_1 = 5$	1	10	8	19
5	$\Delta x_2 = -2$	- 3	- 2	10	15
6	$\Delta x_3 = -1$	- 2	- 6	1	9
7	$\Delta x_2 = 1$	0	0	0	0

Figure 2[4]

Continuing to reduce the highest residuals leads to the solution in two additional steps. The system solution is determined by adding the incremental changes in each unknown. Hence, $x_1 = 5$, $x_2 = 4 + (-2) + 1 = 3$, and $x_3 = 2 + (-1) = 1$. The solution is checked in the conven-

[4]Ibid., pp. 7 - 25.

tional manner.

It may be seen that while individual residuals increased and decreased, apparently without pattern, the sum of the absolute values of the residuals R_T decreased with each step. This sum will be called the

RESIDUAL TOTAL -- the sum of the absolute values of the individual residuals is called the residual total, denoted

$$R_T = \sum_{i=1}^{n} |r_i| .$$

With each successive step, R_T decreased, indicating a very favorable improvement in the approximation to the solution with each step. The nature of this improvement is specified:

IMPROVEMENT -- a relaxation process step is said to be an improvement if R_T decreases with that step.

Now, improvement does not necessarily imply that a solution is guaranteed. A slightly stronger condition is indicated, thus:

CONVERGENCE -- a relaxation process is said to be convergent if R_T approaches zero as a limit as the process is carried out, that is, as the steps are continued.

This leads to:

SOLUTION -- a set of values for the variables in a system of simultaneous linear equations, obtained by adding the incremental changes in the variable to the initial approximations, is a solution if the individual residuals are all zero when these values are substituted in the equations.

Now the example (II-1) was chosen specifically for its properties which facilitate rapid convergence. It was possible to order equations (II-1) so that each main diagonal coefficient was greater than or equal to the sum of the absolute values of the remaining coefficients in the same column. (The matrix of coefficients is said to be diagonally dominant if $|a_{ii}| \geq \sum_{i=1}^{n} |a_{ij}|$ for all j, with inequality for at least one j.)[5] The conjecture for a general statement suggested by this consideration leads to the following:

THEOREM: Let $\sum_{j=1}^{n} a_{ij}x_j = c_i$, (i= 1,2,...,n) be a system which can be ordered so that,

$$|a_{kk}| - \left(\sum_{i=1}^{k-1} |a_{ik}| + \sum_{i=k+1}^{n} |a_{ik}|\right) > 0,$$

for all k = 1,2,...n.

Then any step of the standard relaxation process

[5] George E. Forsythe, Cleve B. Moler, Computer Solution of Linear Algebraic Systems (Englewood Cliffs:Prentice-Hall, 1967), p.11.

performed on the system will be an <u>improvement</u>.

PROOF: Assume that relaxation step m - 1 has just been performed in the given system (where integers m are used to number the individual steps and where m = 0 implies that no relaxation has been performed). Let the residuals be denoted by r_i for all i.

Then $R_{T_{m-1}} = \sum_{i=1}^{n} |r_i|$. Assume that the residual highest in absolute value is r_k for k an element of $\{1,2,\ldots,n\}$.

Since $|a_{kk}| > \sum_{i=1}^{k-1} |a_{ik}| + \sum_{i=k+1}^{n} |a_{ik}|$ implies that $|a_{kk}| > |a_{ik}|$ for any k, then the basic unit operation effective in reducing r_k to zero is $\Delta x_k = \frac{-r_k}{a_{kk}}$. And for $i \neq k$, the residuals will become $r_i + a_{ik}\frac{-r_k}{a_{kk}}$.

Then,

$$R_{T_m} = \sum_{i=1}^{k-1} \left| r_i + a_{ik}\left(\frac{-r_k}{a_{kk}}\right) \right| + \sum_{i=k+1}^{n} \left| r_i + a_{ik}\left(\frac{-r_k}{a_{kk}}\right) \right|$$

$$\leq \sum_{i=1}^{k-1} |r_i| + \sum_{i=k+1}^{n} |r_i| + \sum_{i=1}^{k-1} \left| a_{ik}\left(\frac{-r_k}{a_{kk}}\right) \right| + \sum_{i=k+1}^{n} \left| a_{ik}\left(\frac{-r_k}{a_{kk}}\right) \right|$$

$$= \sum_{i=1}^{k-1} |r_i| + \sum_{i=k+1}^{n} |r_i| + \left|\frac{-r_{k-}}{a_{kk}}\right| \left(\sum_{i=1}^{k-1} |a_{ik}| + \sum_{i=k+1}^{n} |a_{ik}| \right)$$

$$< \sum_{i=1}^{k-1} |r_i| + \sum_{i=k+1}^{n} |r_i| + \left|\frac{-r_{k-}}{a_{kk}}\right| |a_{kk}|$$

$$= \sum_{i=1}^{k-1} |r_i| + \sum_{i=k+1}^{n} |r_i| + |-r_k|$$

$$= \sum_{i=1}^{n} |r_i| = R_{T_{m-1}}$$

Hence, $R_{T_m} < R_{T_{m-1}}$ => step m is an improvement. Q.E.D.

While convergence is not assured by the above conclusion, every case tested converged rapidly, including the following which was designed to have small improvement at each step:

$$\begin{aligned} 6x_1 - 4x_2 - 3x_3 + 14 &= r_1 \\ -2x_1 - 8x_2 - 8x_3 + 134 &= r_2 \\ -3x_1 - 3x_2 + 12x_3 = 6 &= r_3. \end{aligned}$$

In figure 4 note that not only is improvement seen in each step but convergence is rapid, (Fig. 3 is the relaxation operation table). Rapid convergence was also experienced in these experiments when rational coefficients were used with diagonal elements very nearly equal in absolute value to the sum of the absolute values of the remaining column elements.

OPERATION	Δr_1	Δr_2	Δr_3
$\Delta x_1 = 1$	6	-2	-3
$\Delta x_2 = 1$	-4	-8	-3
$\Delta x_3 = 1$	-3	-8	12

Figure 3

m	OPERATION	r_1	r_2	r_3	R_{T_m}
0	All x =0	14	134	6	154
1	$\Delta x_2 = 16$	-50	6	-42	98
2	$\Delta x_1 = 8$	- 2	- 10	166	78
3	$\Delta x_3 = 5$	-17	- 50	- 6	73
4	$\Delta x_2 = -6$	7	- 2	12	21
5	$\Delta x_3 = -1$	10	6	0	16
6	$\Delta x_1 = -1$	4	8	3	15
7	$\Delta x_2 = 1$	0	0	0	0

Figure 4

Some systems of linear equations are so conditioned that the standard relaxation process will diverge unless special adjustments are made in the structure of the process itself.[6] There are a number of refinements of the basic relaxation process which are discussed in detail in Shaw [21] and in Allen [20]. These include over-relaxation, block-relaxation, and other topics which are useful when equations

[6] Allen, op. cit., pp. 21-23.

are ill-conditioned for rapid convergence. However, the finite-difference approximations discussed in the next chapter are systems well suited for application of the standard relaxation process, as will be seen.

CHAPTER III

FINITE DIFFERENCE APPROXIMATIONS TO ORDINARY AND PARTIAL DIFFERENTIAL EQUATIONS

This section is largely a precis of thorough treatments in Shaw [21] and Allen [20]; and it is included here for clarity and continuity.

A general solution to a differential equation provides a function which can be evaluated at any point in its domain. A relaxational solution is, of course, numerical, and consists of values of the wanted function determined at certain equally spaced points of subdivision along the range of integration.

Suppose the general ordinary differential equation is given as

$$\frac{d^2\phi}{dx^2} + f(x) = 0, \tag{III-1}$$

and that the solution is sought over the domain $a \leq x \leq b$. The domain is divided into n intervals of equal length; and n is chosen to provide enough points for a reasonably accurate solution without overburdening one with excess computation. The uniform interval length, denoted h, is then given by

$$h = \frac{b - a}{n} .$$

The boundary conditions for the specific problem will provide the values of ϕ at $x = a$ and $x = b$. The Taylor series

expansion about the typical point of subdivision x_i is used to obtain the finite difference approximation for ϕ''_i . The expansion about x_i is

$$\phi(x) = \phi(x_i) + (x-x_i)\phi'(x_i) + \frac{(x-x_i)^2\phi''(x_i)}{2!} + \ldots$$

Evaluating the series for $x = x_i + h$ gives

$$\phi(x_i+h) = \phi(x_i) + h\phi'(x_i) + \frac{h^2}{2!}\phi''(x_i) + \frac{h^3}{3!}\phi'''(x_i) + \ldots$$

Evaluation for $x = x_i - h$ gives

$$\phi(x_i-h) = \phi(x_i) - h\phi'(x_i) + \frac{h^2}{2!}\phi''(x_i) - \frac{h^3}{3!}\phi'''(x_i) + \ldots$$

Adding the two series:

$$\phi(x_i+h) + \phi(x_i-h) = 2\phi(x_i) + h^2\phi''(x_i) + \frac{2h^4}{4!}\phi''''(x_i) + \ldots$$

or, approximately,

$$h^2\phi''(x_i) = \phi(x_i+h) + \phi(x_i-h) - 2\phi(x_i) \qquad \text{(III-2)}$$

in which the error is $\frac{2h^4}{4!}\phi''''(x_i) + \ldots$, which decreases as h is made smaller.

Now, according to relation (III-2), satisfaction of (III-1) required that $\phi(x_i+h) + \phi(x_i-h) - 2\phi(x_i) + h^2f(x_i)=0$. This equation is typical of those which connect every set of ϕ values at every set of three successive points of subdivision(disregarding the endpoints as points of subdivision). The effect is, then, that of replacing the differential equa-

tion (III-1) with a set of n-1 algebraic equations which can be satisfied by n-1 values of ϕ at n-1 points.

In two dimensions, equation (III-1) can be extended to Poisson's equation:

$$\frac{\delta^2\phi}{\delta x^2} + \frac{\delta^2\phi}{\delta y^2} + f(x,y) = 0. \qquad \text{(III-3)}$$

Whereas in one dimension the range of integration was divided up by points of subdivision into subintervals of equal length, in two dimensions an area of integration is subdivided by a uniform mesh of points (x,y) such that $a \leq x \leq b$ and $c \leq y \leq d$. Usually $|a-b| = |c-d|$, providing a particular space or region which is a square. In the region thus established, point (x_i, y_i) is the typical point, separated vertically and horizontally from neighboring points in the mesh by the distance $h = \frac{b-a}{n} = \frac{d-c}{n}$. Intervals [a,b] and [c,d] are of the same length; and each is divided into n subdivisions forming the square mesh.[7]

(III-4) depicts typical node (x_i, y_i) of the net. The finite difference approximation for $\frac{\delta^2\phi}{\delta x^2}$ is given by (III-2) as

$$\frac{\delta^2\phi}{\delta x^2} = \frac{\phi(x_i+h, y_i) + \phi(x_i-h, y_i) - 2\phi(x_i, y_i)}{h^2} \qquad \text{(III-4)}$$

Similarly, the finite difference approximation to

[7] Ibid., pp. 54-59.

$\frac{\delta^2\phi}{\delta y^2}$ is

$$\frac{\delta^2\phi}{\delta y^2} = \frac{\phi(x_i,y_i+h) + \phi(x_i,y_i-h) - 2\phi(x_i,y_i)}{h^2}$$

Hence, equation (III-3) becomes, in approximation, a system of $(n-1)^2$ equations in $(n+1)^2$ unknowns, with typical equation-

$$\phi(x_i+h,y_i)+\phi(x_i-h,y_i)+\phi(x_i,y_i+h)+\phi(x_i,y_i-h)-4\phi(x_i,y_i)$$

$$+ h^2 f(x_i,y_i) = 0$$

In two dimensions boundary points have their values established by boundary conditions, hence no equations for these points are considered. The values of the function are already known at these points, so they are not subjected to relaxation. But (x_i,y_i), where (x_i,y_i) is a boundary point, is used in the initial computation of residuals; after this these points have no effect on the relaxation process.

Chapter IV continues with the discussion of relaxation applied to the systems produced by the methods of this section. Finite-difference approximations to other frequently used second, third, and fourth order differential equations may be found in Shaw [21] and Allen [20].

CHAPTER IV

RELAXATION METHODS FOR SOLVING PARTIAL DIFFERENTIAL EQUATIONS OF THE SECOND ORDER

The solution to specific equation

$$\frac{\delta^2\phi}{\delta x^2} + \frac{\delta^2\phi}{\delta y^2} = -(2x^2 + y)$$

will be approximated by relaxation on the square region $0 \leq x \leq 4$, $0 \leq y \leq 4$. Boundary conditions are specified such that $\phi(x_i, y_i) = 0$ on the border of the square region, that is, wherever $x_i = 0,4$ or $y_i = 0,4$.

The region is subdivided into squares of length $h = 1$, the mesh being kept coarse deliberately for ease of description.

The general approximation equation is written in residual form by reference to Chapter III, as follows:

$$\text{(IV-1)} \qquad \phi(x_i+1, y_i) + \phi(x_i-1, y_i) + \phi(x_i, y_i+1) + \phi(x_i, y_i-1) - 4\phi(x_i, y_i) + 2x_i^2 + y_i = r_i \,.$$

Refer to Figure 5. In this demonstration the nodes are numbered consecutively, as shown, to make possible a simplified notation for the finite-difference system. Then the system of equations with $h = 1$, of which (IV-1) is the pattern, contains nine equations in only nine unknowns, as follows: (boundary values included for calculating residuals)

$$\phi_8 + \phi_2 + \phi_6 + \phi_{12} - 4\phi_7 + (2x^2+y)_7 = r_7$$

$$\phi_9 + \phi_3 + \phi_7 + \underline{\phi_{13}} - 4\phi_8 + (2x^2+y)_8 = r_8$$

$$\phi_{10} + \phi_4 + \phi_8 + \phi_{14} - 4\phi_9 + (2x^2+y)_9 = r_9$$

$$\phi_{17} + \underline{\phi_{13}} + \phi_7 + \phi_{11} - 4\phi_{12} + (2x^2+y)_{12} = r_{12}$$

(IV-2)
$$\phi_{18} + \phi_{14} + \phi_8 + \phi_{12} - \underline{4\phi_{13}} + (2x^2+y)_{13} = r_{13}$$

$$\phi_{19} + \phi_{15} + \phi_9 + \underline{\phi_{13}} - 4\phi_{14} + (2x^2+y)_{14} = r_{14}$$

$$\phi_{22} + \phi_{18} + \phi_{12} + \phi_{16} - 4\phi_{17} + (2x^2+y)_{17} = r_{17}$$

$$\phi_{23} + \phi_{19} + \underline{\phi_{13}} + \phi_{17} - 4\phi_{18} + (2x^2+y)_{18} = r_{18}$$

$$\phi_{24} + \phi_{20} + \phi_{14} + \phi_{18} - 4\phi_{19} + (2x^2+y)_{19} = r_{19},$$

where $(2x^2+y)_k$ is the value of f(x,y) at the node numbered k.

The relaxation operator effecting the biggest change in r_k is $\Delta\phi_k = 1$, since r_k will change by 4 units for each increment of $\Delta\phi_k$. No other operator would change r_k by more than one unit. The effect of this incremental change is observed by reference to equations (IV-2) and Figure 6. Attention is focused on a single node, the point numbered 13. Residual r_{13} at this point is reduced when necessary by application of the $\Delta\phi_{13}$ operator. Among equations (IV-2), ϕ_{13} appears underlined.

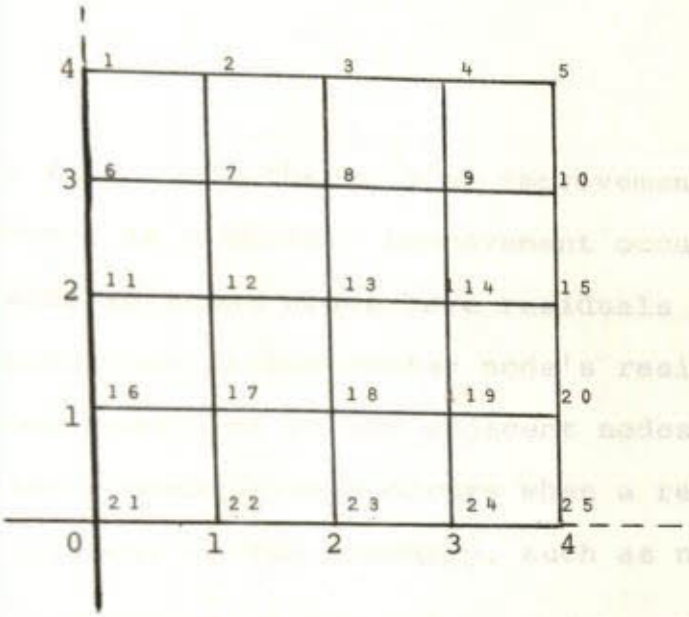

Figure 5

The result of $\Delta\phi_{13} = 1$ is to reduce r_{13} by four units. At the same time r_8, r_{12}, r_{14}, and r_{18} are increased by one unit each because ϕ_{13} appears in each of the equations involving those residuals. All other residuals are unaffected. Thus, Figure 6 can be used as a relaxation operator in two dimensions,[8] applicable at any node because relaxation at a point affects only the four neighboring points.

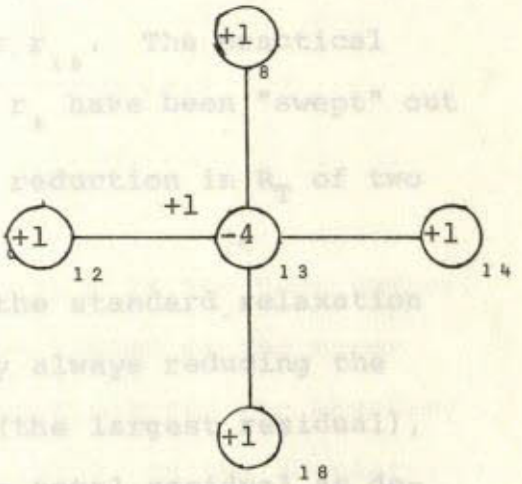

Figure 6

It must be noted that while $\Delta\phi_{13} = 1$ reduces the residual for one equation, the residual total R_T may remain the

[8] Shaw, op. cit., pp. 45-51.

same, in which case there is no improvement in the residual picture as a whole. Improvement occurs in interior points when adjacent nodes have residuals opposite in sign. Then a reduction in the center node's residual would also cause some reduction in the adjacent nodes of opposite sign.

Improvement always occurs when a reduction is made on a node adjacent to the boundary, such as node 9. With $\Delta\phi_9=1$, r_9 is reduced four units, and r_8 and r_{14} are increased one unit each. But since $\phi_4 = \phi_{10} = 0$, by the boundary conditions, it is never necessary to relax r_4 or r_{10}. The practical result is that half of the units of r_9 have been "swept" out of the picture, with the consequent reduction in R_T of two units.

Experiments have shown that the standard relaxation process, continued in this manner by always reducing the residual highest in absolute value (the largest residual), can be used to remove as much of the total residual as desired, leading to a solution which is as accurate as desired.

The "hand" version of the relaxation solution is carried out on a diagram of the problem region itself (Fig.7). The nodes are left unnumbered to keep from cluttering the diagram.

The initial approximations to the ϕ_k are placed at the upper left of the nodes. In this example $\phi_k = 0$ for

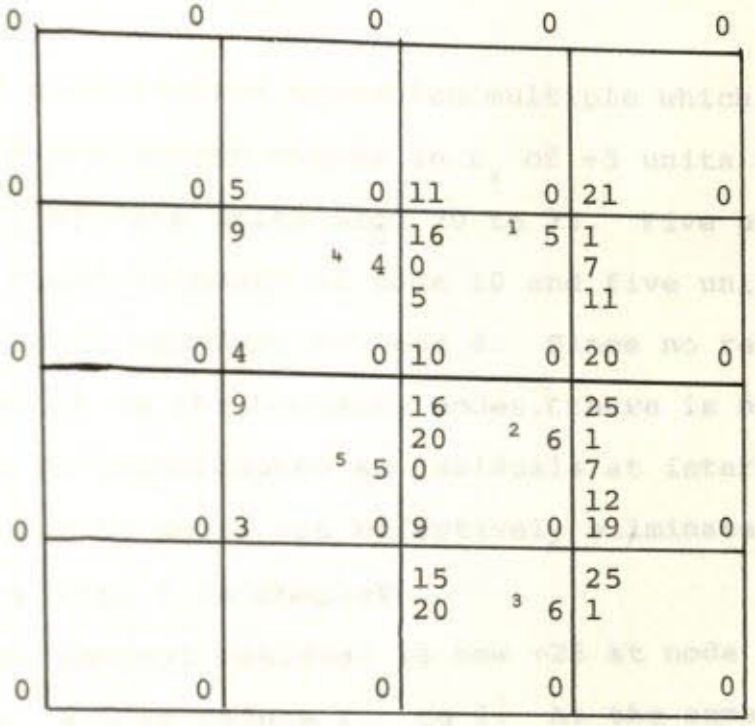

Figure 7
Steps 1 - 5

all k is the first approximation, where k is the node number.

Then the r_k are calculated and placed at the upper right of each node. The initial conditions for the boundary are entered as permanent values for the ϕ_k on the boundary. So the numbers above the line at the nodes in Figure 7 represent the condition before relaxation begins. The small numbers shown indicate the relaxation step numbers as described below.

The largest residual is seen to be 21 at node 9. The increment $\Delta\phi_9 = 5$, indicated with a 5 at the lower left of node 9, results in a decrease in r_9 by 20 units to a value of 1. The new residual is placed at the lower right

of node 9 opposite the operation multiple which produced it. There is a concurrent change in r_8 of +5 units from 11 to 16, and in r_{14} of five units from 20 to 25. Five units are moved onto the right boundary at node 10 and five units are moved onto the upper boundary at node 4. Since no relaxation need be carried on at the boundary nodes, there is no way these units can be reintroduced as residuals at interior nodes. Hence, these 10 units are effectively eliminated from R_T, and relaxation step 1 is complete.

The largest residual is now +25 at node 14. Step 2 uses $\Delta\phi_{14} = 6$ to reduce r_{14} to 1. At the same time, r_{19} is increased to 25, r_{13} goes to 16 and r_9 goes to 7. Not only are five units of the original r_{14} "swept over the right boundary", but one unit of r_9 from the first point relaxation is also eliminated, a salutary effect indeed.

After five steps of the relaxation, crowding of the diagram occurs; so the status quo is reëntered on another blank diagram (Figure 8). Subsequent frames, each detailing five point-relaxations, are seen following this page, ending with step 47.

At step 47 all residuals have been reduced to less than 1 except for r_{13} which is 1. It should be obvious that the process does continue to improve as long as it is carried on and that an even greater refinement could be made in the

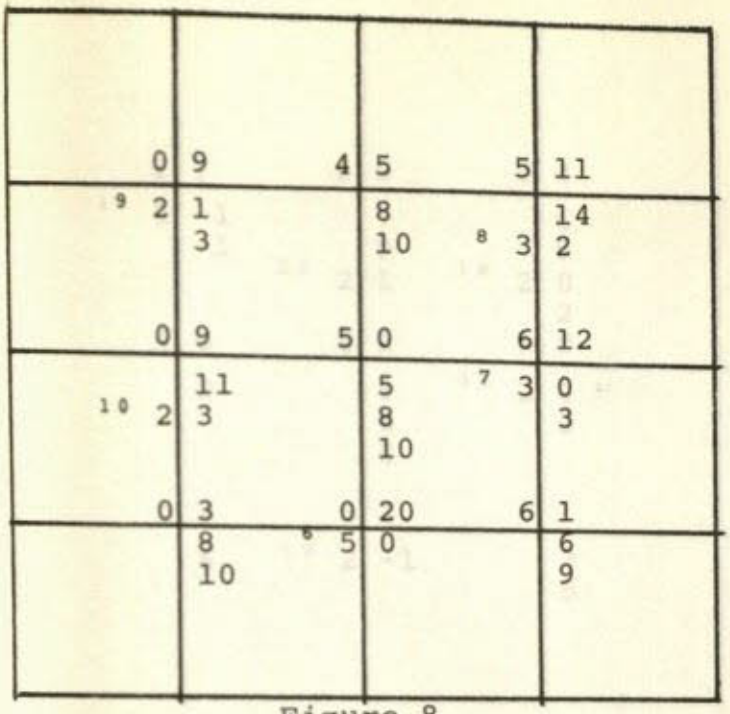

Figure 8
Steps 6 - 10

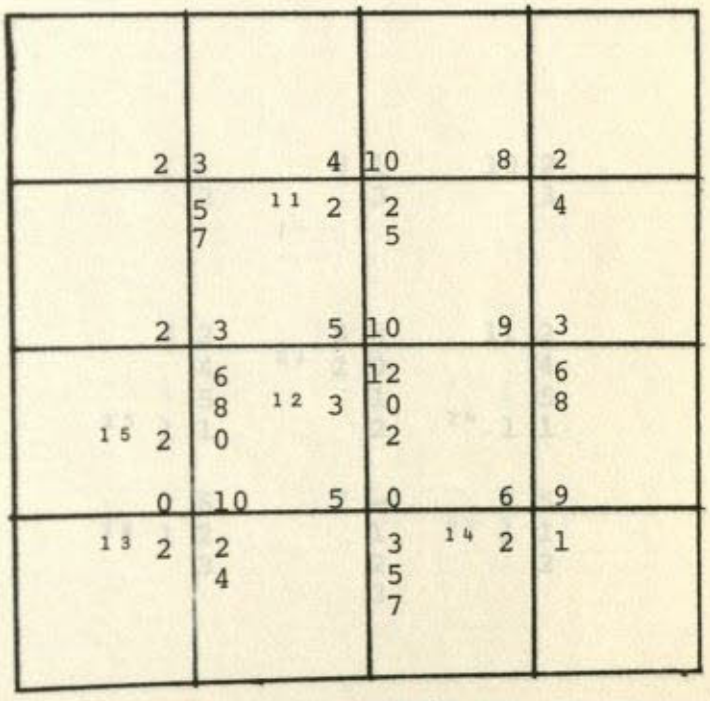

Figure 9
Steps 11 - 15

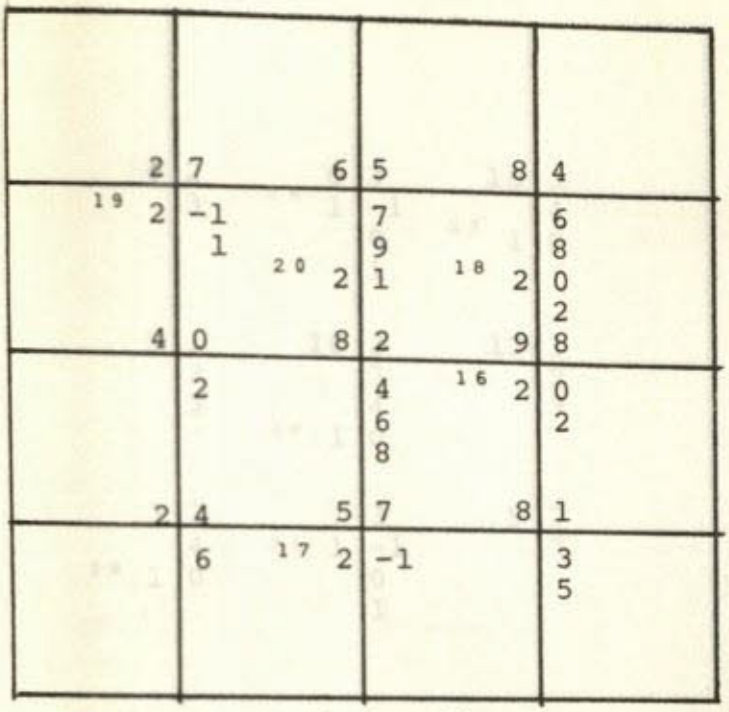

Figure 10
Steps 16 - 20

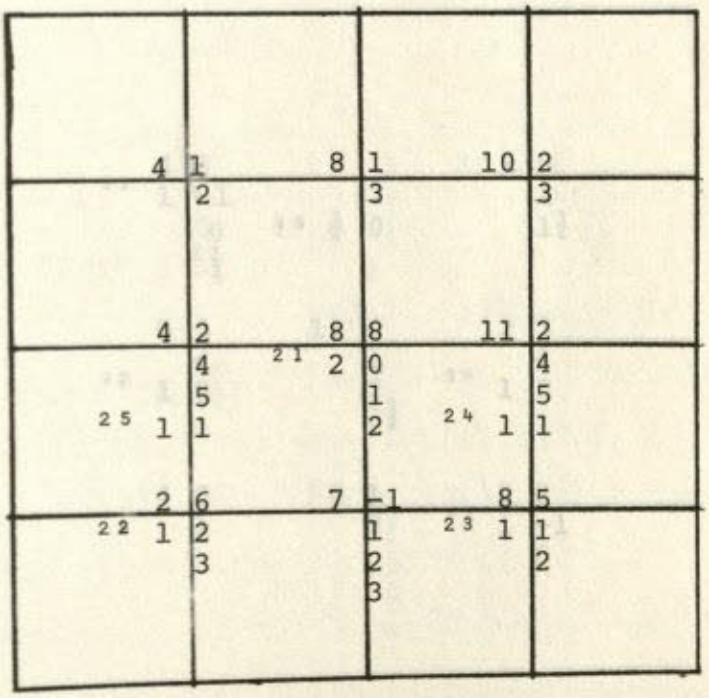

Figure 11
Steps 21 - 25

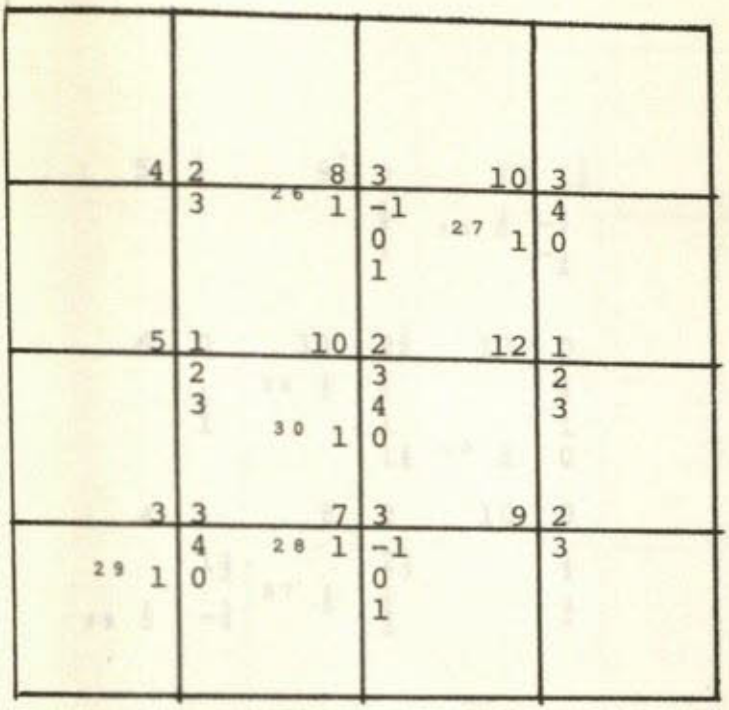

Figure 12
Steps 26 - 30

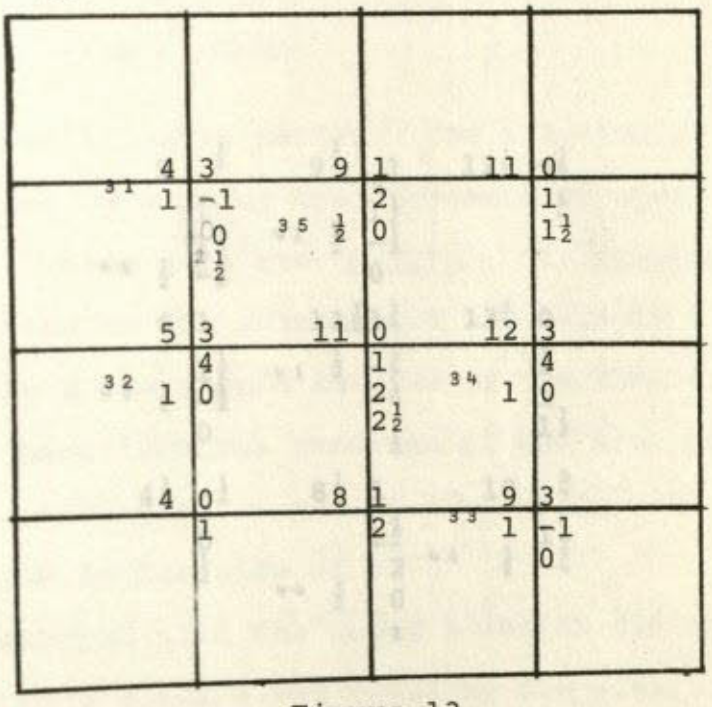

Figure 13
Steps 31 - 35

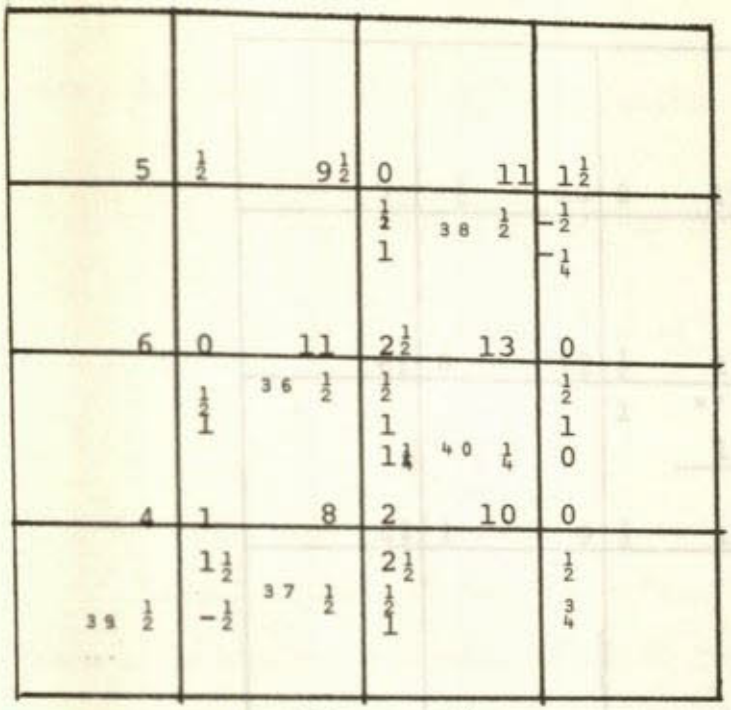

Figure 14
Steps 36 - 40

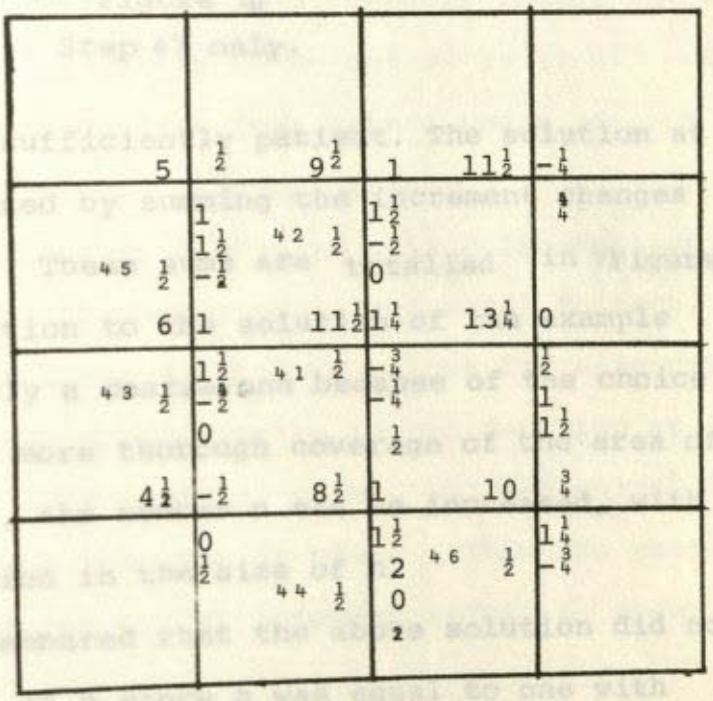

Figure 15
Steps 41 - 46

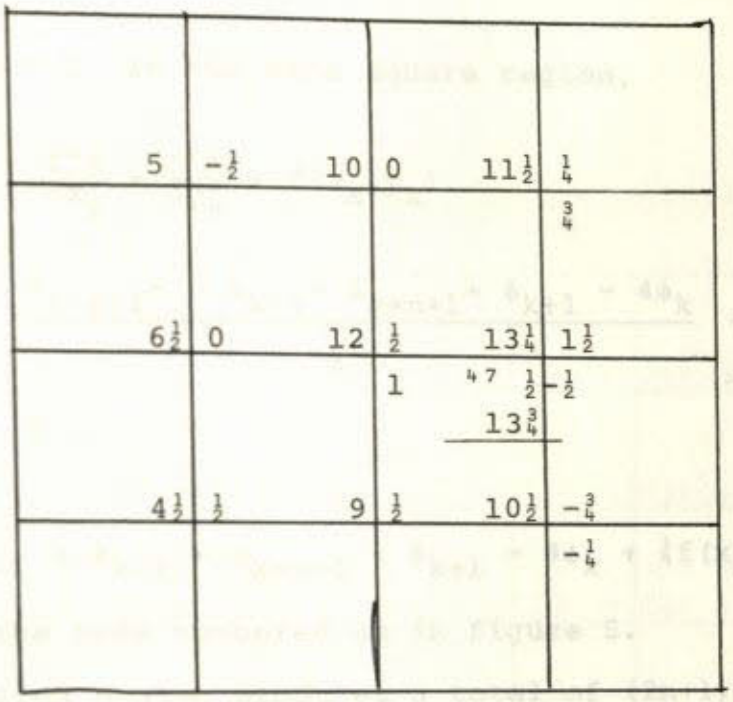

Figure 16
Step 47 only.

solution if one were sufficiently patient. The solution at this stage is determined by summing the increment changes at each of the nodes. These sums are totalled in Figure 16.

This approximation to the solution of the example equation is regrettably a coarse one because of the choice of mesh size. If more thorough coverage of the area of integration is wanted, the number n can be increased, with the consequent reduction in the size of h.

It must be remembered that the above solution did not involve consideration of h since h was equal to one with n = 4. But doubling n results in halving the value of h.

Hence, for $h = \frac{1}{2}$ in the same square region,

$$\frac{\delta^2\phi}{\delta x_k^2} + \frac{\delta^2\phi}{\delta y_k^2} + f(x_k, y_k)$$

$$= \frac{\phi_{k-n-1} + \phi_{k-1} + \phi_{k+n+1} + \phi_{k+1} - 4\phi_k}{\frac{1}{4}} + f(x_k, y_k)$$

$$= 0$$

or,

$$\phi_{k-n-1} + \phi_{k-1} + \phi_{k+n+1} + \phi_{k+1} - 4\phi_k + \tfrac{1}{4}f(x_k, y_k) = 0,$$

where k is the node numbered as in figure 5.

Doubling n also produces a total of $(2n+1)^2 = 81$ nodes instead of the original 25. With the boundary values equal to zero, the number of equations to be solved is $(n-1)^2 = 49$, in $(n-1)^2 - 4 = 49$ unknowns (the four corner nodes are not included in any equations of interior nodes).

It is not necessary in producing this refinement of the solution to begin with all values of ϕ_k set equal to zero. The work done on the coarser net led to values for nine points which should be a better approximation to the function at those points than any other initial guess. So the solution is begun with this previous solution in place. Then the residuals are recalculated for all 49 interior nodes, as shown in Figure 17.

As action swirls around the nodes with solutions already in place, the relaxation at adjacent points has the

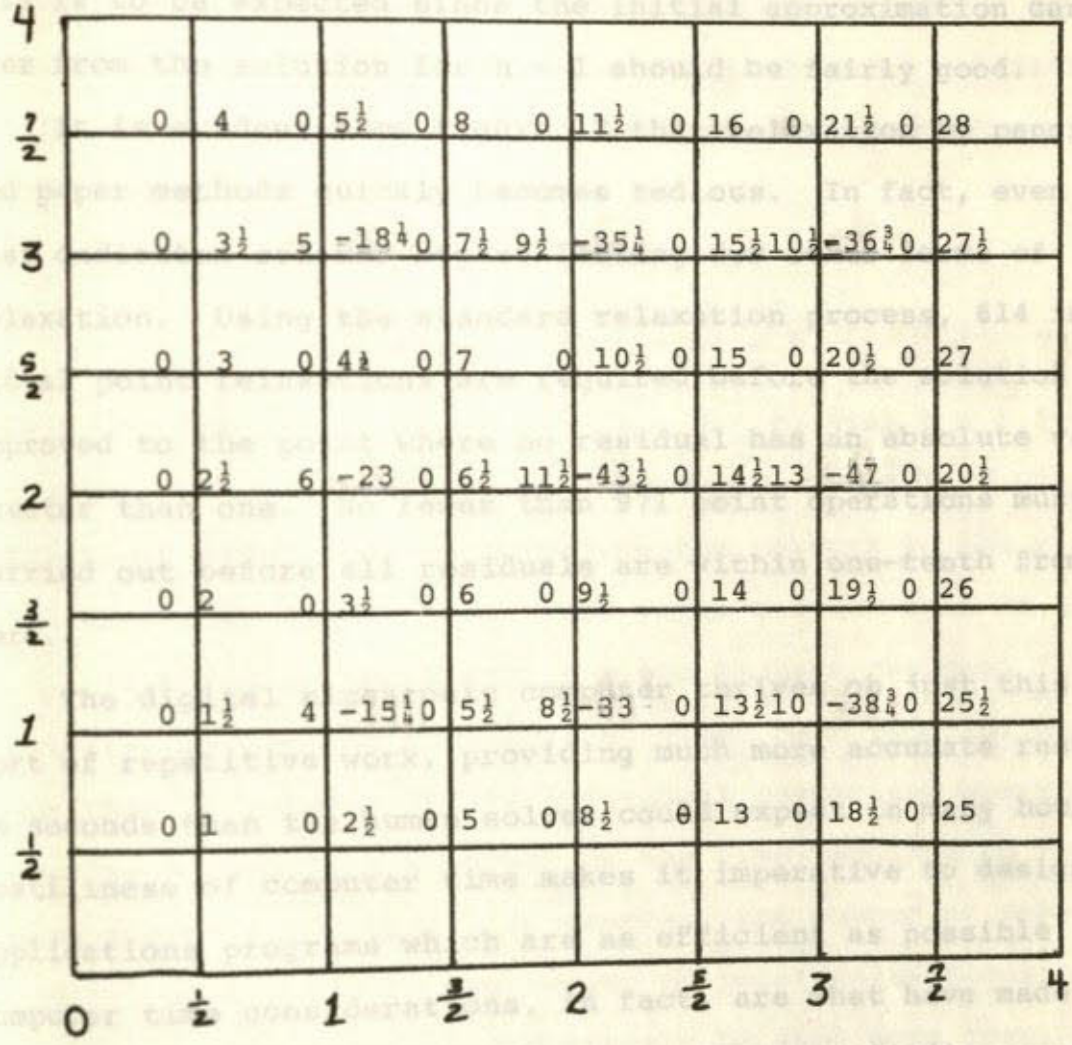
4
7/2
3
5/2
2
3/2
1
1/2
0
1/2
1
3/2
2
5/2
3
7/2
4
0 4 0 5½ 0 8 0 11½ 0 16 0 21½ 0 28
0 3½ 5 -18¼ 0 7½ 9½ -35¼ 0 15½ 10½ -36¾ 0 27½
0 3 0 4½ 0 7 0 10½ 0 15 0 20½ 0 27
0 2½ 6 -23 0 6½ 11½ -43½ 0 14½ 13 -47 0 20½
0 2 0 3½ 0 6 0 9½ 0 14 0 19½ 0 26
0 1½ 4 -15¼ 0 5½ 8½ -33 0 13½ 10 -38¾ 0 25½
0 1 0 2½ 0 5 0 8½ 0 13 0 18½ 0 25

Figure 17

effect also of reducing residuals here rather than increasing them. This means that changes in the functional value at these nodes, when called for, may be relatively very small. This is to be expected since the initial approximation carried over from the solution for h = 1 should be fairly good.

It is evident from figure 17 that relaxation by pencil and paper methods quickly becomes tedious. In fact, even the most dedicated relaxer begins looking for other forms of relaxation. Using the standard relaxation process, 614 individual point relaxations are required before the solution is improved to the point where no residual has an absolute value greater than one. No fewer than 971 point operations must be carried out before all residuals are within one-tenth from zero.

The digital electronic computer thrives on just this sort of repetitive work, providing much more accurate results in seconds than the human solver could expect in many hours. Costiliness of computer time makes it imperative to design applications programs which are as efficient as possible. Computer time considerations, in fact, are what have made relaxation methods less desirable than others in finding approximate solutions to partial differential equations.

A major consideration in carrying out the standard relaxation process on a computer is the time used simply in locating the largest residual at each step. The human solver

can locate this node in a relatively short elapsed time, but the computer must perform a brute force search by comparing every node with the other nodes. The time required for this search is appreciably longer than is the time required to perform the point relaxation itself. For each point relaxation n comparison operations must be executed to locate the next largest residual.

Using again the example equation

$$\frac{\delta^2\phi}{\delta x^2} + \frac{\delta^2\phi}{\delta y^2} + (2x^2 + y) = 0,$$

solving on the region $1 \leq x \leq 8$, $1 \leq y \leq 8$, with $h = 1$, $(n=6)$, and using the computer for the arithmetic, a total of 614 point relaxations were required to reduce the network to the place where no residual had an absolute value greater than or equal to 1. Involved in the search for the highest residual (always considered as highest in absolute value, bear in mind), were no fewer than 36 x 614 = 21,104 individual comparison operations employing a significant amount of the time required to reach the approximate solution. In general, the number of comparisons is $(n-1)^2$ multiplied by the number of point-relaxations required. The algorithm would obviously be much more feasible if this major barrier to efficiency could be eliminated.

Experiment has indicated that apparently relaxation need not be applied to the residual highest in absolute value at each step in order for improvement to take place, even

though this is advocated by Southwell, Allen, Shaw, and others. Further, it seems points may be relaxed in successive order without respect to the size of the residual. While efficiency is impaired in moving the residuals over the boundaries, that is to say, while R_T does not decrease as rapidly as in the standard process, nevertheless approximate solutions are obtained which are as accurate,without using the time-consuming comparisons in the computer program. Working a simpler example by hand readily illustrates that the successive-point technique takes more point-relaxations and so is less efficient on pencil and paper. But this extra work is much more than offset in computer solutions by the saving of time in elimination of comparisons. In the example just above, a total of 756 relaxations were required in the successive point process instead of the 614 of the Southwell process. But, the 21,104 comparisons were not needed in the former method. Comparisons do not take as much time as relaxations on the computer; and a rough estimate of time savings based on speed of individual arithmetic operations in the RCA SPECTRA-70 computer is that a savings of 75% was effected using the points-in-succession algorithm. Figure 18 shows the condition of the network with residuals reduced to less than one. Figure 19 shows the network with the largest residual reduced to less than 1. Both figures compare the Southwell method and the

SOUTHWELL METHOD

Point Relaxations 614

Comparisons 22,140

k	r_k	ϕ_k
1	0.656	55.815
2	0.655	82.347
3	0.633	93.917
4	0.684	95.006
5	0.000	85.317
6	0.340	58.922
7	0.699	68.571
8	0.649	106.308
9	0.344	123.949
10	0.715	125.475
11	0.702	110.339
12	0.644	72.712
13	0.260	61.857
14	0.984	99.016
15	0.719	117.440
16	0.329	119.322
17	0.641	103.553
18	0.978	66.232
19	0.255	47.103
20	0.881	77.443
21	0.691	93.191
22	0.000	95.151
23	0.651	81.961
24	0.000	51.640
25	0.549	30.365
26	0.000	51.344
27	0.951	62.420
28	0.329	64.130
29	0.606	55.151
30	0.679	34.368
31	0.268	14.561
32	0.274	25.149
33	0.839	30.965
34	0.329	32.129
35	0.637	27.751
36	0.681	17.359

Figure 18

SUCCESSIVE-POINT METHOD

Point Relaxations 756

Comparisons 0

k	r_k	ϕ_k
1	0.135	56.100
2	0.152	82.728
3	0.137	94.328
4	0.099	95.451
5	0.049	85.625
6	0.000	59.206
7	0.354	68.806
8	0.439	106.636
9	0.246	124.270
10	0.335	125.948
11	0.198	110.895
12	0.049	73.199
13	0.540	61.843
14	0.691	99.180
15	0.683	117.596
16	0.548	119.510
17	0.335	104.006
18	0,099	66.744
19	0.644	46.927
20	0.839	77.336
21	0.840	93.106
22	0.682	95.039
23	0.426	82.208
24	0.137	51.871
25	0.622	30.171
26	0.828	50.972
27	0.839	62.291
28	0.690	64.015
29	0.439	55.344
30	0.152	34.669
31	0.452	14.409
32	0.622	24.918
33	0.643	30.911
34	0.539	32.077
35	0.353	27.920
36	0.135	17.614

Figure 19

successive-point method.

Appendix A shows the hand calculation of the example for the coarser net of nine nodes used at the beginning of the chapter but employing successive-point relaxation. Notice that 57 calculations were used rather than the 47 in the previous calculation in reducing the system to roughly the same level of approximation.

In successive-point relaxation, as stated, each point is relaxed in succession until all points have been relaxed. The process is then repeated from the first point until the residual total is as small as desired. The concluding paragraphs of this thesis describe and prove a theorem which certifies the validity of the successive-point relaxation process.

First, two definitions:

RELAXATION CYCLE -- the successive-point relaxation process applied so that each node in the mesh is relaxed exactly once is called a relaxation cycle.

CYCLICAL OPERATION -- the operation involving the successive reduction of all r_k using $\Delta\phi_k = r_k/4$ for all k is called a cyclical operation, denoted σ.

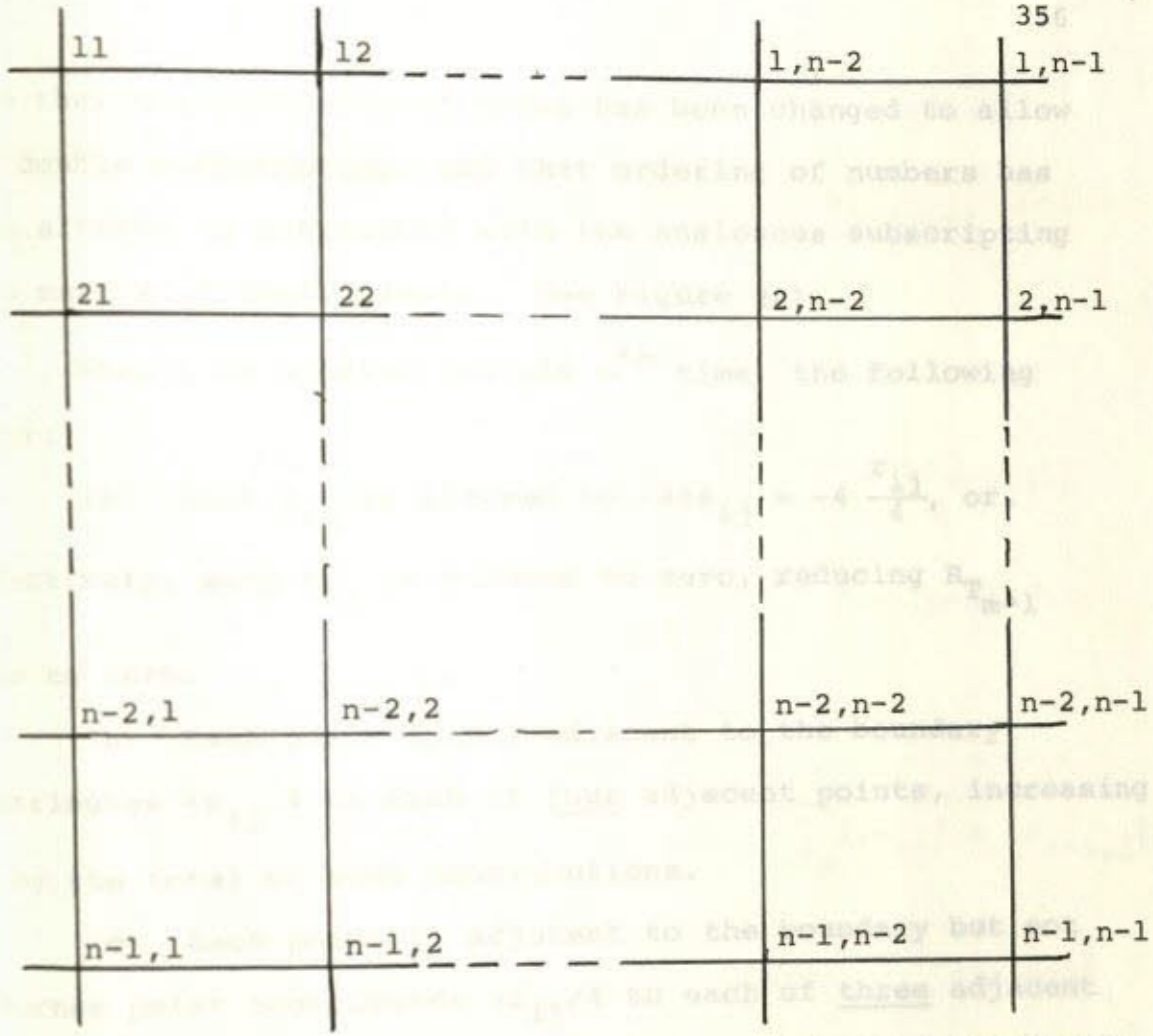

Figure 20

THEOREM -- The cyclical operation σ, applied to a system of finite-difference equations which is an approximation to Poisson's equation, is an improvement.

PROOF -- Assume that cyclical operation number m-1 has been performed (m=0 implies that no relaxation has taken place), and that

$$R_{T_{m-1}} = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} |r_{ij}| .$$

(Note that the numbering of nodes has been changed to allow for double subscripting, and that ordering of numbers has been altered to correspond with the analogous subscripting of a matrix of coefficients. See Figure 20).

When σ is applied for the m^{th} time, the following occur:

(a) Each r_{ij} is altered by $-4\Delta\phi_{ij} = -4\,\frac{r_{ij}}{4}$, or, effectively, each r_{ij} is reduced to zero, reducing $R_{T_{m-1}}$ also to zero.

(b) Each point ij not adjacent to the boundary contributes $+r_{ij}/4$ to each of <u>four</u> adjacent points, increasing R_T by the total of such contributions.

(c) Each point ij adjacent to the boundary but not a corner point contributes $+r_{ij}/4$ to each of <u>three</u> adjacent interior points, the fourth point being a boundary point and not considered in computing the R_T.

(d) Each corner point ij adjacent to the boundary contributes $+r_{ij}/4$ to each of two adjacent interior points, the other two being boundary points.

Hence, under σ,

$$R_{T_m} = R_{T_{m-1}} - R_{T_{m-1}} + 4 \sum_{i=2}^{n-2} \sum_{j=2}^{n-2} \frac{r_{ij}}{4} +$$

(expression continued next page)

$$+ 3\sum_{i=2}^{n-2}\left(\frac{r_{i1}}{4} + \frac{r_{i,n-1}}{4}\right) + 3\sum_{j=2}^{n-2}\left(\frac{r_{1j}}{4} + \frac{r_{n-1,j}}{4}\right)$$

$$+ 2\left(\frac{r_{11}}{4} + \frac{r_{n-1,1}}{4} + \frac{r_{1,n-1}}{4} + \frac{r_{n-1,n-1}}{4}\right)$$

$$= \sum_{i=2}^{n-2}\sum_{j=2}^{n-2} r_{ij}$$

$$+ \frac{3}{4}\left(\sum_{i=2}^{n-2}(r_{i1} + r_{i,n-1}) + \sum_{j=2}^{n-2}(r_{1j} + r_{n-1,j})\right)$$

$$+ \frac{1}{2}(r_{11} + r_{n-1,1} + r_{1,n-1} + r_{n-1,n-1})$$

$$< \sum_{i=2}^{n-2}\sum_{j=2}^{n-2}|r_{ij}|$$

$$+ \sum_{i=2}^{n-2}(|r_{i1}| + |r_{i,n-1}|) + \sum_{j=2}^{n-2}(|r_{1j}| + |r_{n-1,j}|)$$

$$+ |r_{11}| + |r_{n-1,1}| + |r_{1,n-1}| + |r_{n-1,n-1}|$$

$$= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}|r_{ij}|$$

$$= R_{T_{m-1}}.$$

Hence, σ is an improvement. Q.E.D.

0	5 0	11 0	21
1 1	1 4 4 3	12 0 3	24
0	4	10	20
2 1	5 1 2 5 3 5	11 14 2	23
0	3	9	19
3 1	4 0	10 13	

Steps 1 - 5

Calculation of solution by method of successive-point relaxation.

1	4 3	3 0	24
10 1	0 2	9 7 6 10	0 7
1	5 3	2 0	23
11 1	6 2 3	5 12 8 7 14	29 1 8
1	0 0	13 0	19
12 1	3 6 3 5 1	1 8 9 9 7	22 29 1

Steps 6 - 12

$5\frac{1}{4}$	0 $9\frac{3}{4}$	$\frac{1}{2}$ 11	0
$6\frac{1}{4}$	$\frac{3}{4}$ $12\frac{1}{4}$	0 $13\frac{1}{4}$	$\frac{1}{2}$
$4\frac{3}{4}$	0 $8\frac{3}{4}$	$10\frac{1}{4}$	0

Approximate Solution with

$|r_i| < 1$ for all i.

57 steps required.

BIBLIOGRAPHY

1. Southwell, Richard V. Relaxation Methods in Engineering Science. London: Oxford University Press, 1940.

2. Fox, L. A Short Account of Relaxation Methods. London: Quarterly Journal of Mechanics and Applied Mathematics, 1948, 1, 253-280.

3. Bickley, W. G. Finite-Difference Formulae for the Square Lattice. Quarterly Journal Mechanics and Applied Mathematics, 1948, 1, pp. 35-42.

4. Kunz, Kaiser Numerical Analysis. New York: McGraw--Hill, 1957.

5. Young, David Iterative Methods for Solving Partial Differential Equations of the Elliptic Type. Transactions of the American Mathematical Society, Jan.,1954; 76:92-111.

6. Christopherson, D.G., and Southwell, Richard V. Relaxation Methods Applied to Engineering Problems, III: Problems Involving the Two Independent Variables, 1938, Proceedings of the Royal Society,168:317-350.

20. Allen, D. N. deG. Relaxation Methods, New York: McGraw-Hill, 1954.

21. Shaw, F. S., Relaxation Methods, New York: Dover Publications, 1953.

22. Forsythe, George E., and Moler, Cleve B. Computer Solution of Linear Algebraic Systems. Englewood Cliffs, New Jersey: Prentice-Hall, 1967, p. 11.

ABSTRACT

Doud, Warren D., Bachelor of Science, University of Southwestern Louisiana, Lafayette, Louisiana.
Master of Science, May, 1969.
Major: Mathematics
Thesis Title: Relaxation Techniques Leading to Numerical Solutions of Partial Differential Equations
Pages in Thesis 43 .

This thesis develops the topic of relaxation methods in an expository manner to the place where a theorem can be introduced which specifies the conditions sufficient for improvement to take place in the methods applied to systems of linear algebraic equations. Then, after tracing the development of techniques for approximating differential equations by systems of linear equations of the finite-difference form, solutions by Richard Southwell's methods and by an innovative new method of successive-point relaxation are compared, using a computer for arithmetical calculations. Finally, a theorem is introduced which validates the new method in a manner parallel to that of the first theorem.

AUTHOR'S AUTOBIOGRAPHY

Warren Doud was born in Chicago, Illinois, on July 11, 1934. Graduating from Southwest High School in Kansas City, Missouri, he attended Wheaton College, Wheaton, Illinois, for one year, Kansas City Conservatory of Music, Kansas City, Missouri, for one year. He attended Prairie Bible Institute, Three Hills, Alberta, Canada, from 1953-1957, graduating with a diploma in Bible, English, and Public Speaking. At this time he also received a certificate of Christian Education from the Evangelical Teacher's Training Association.

He was then employed for two years with the Bendix Corporation as a quality control inspector, after which he moved to New Iberia, Louisiana. From 1960 to 1964 he was a life underwriter with Metropolitan Life Insurance Co., and with the Jefferson Standard Life Insurance Co. He then entered undergraduate mathematics at University of Southwestern Louisiana, Lafayette, Louisiana, where he received the Bachelor of Science in May, 1967. He was elected to Pi Mu Epsilon, Kappa Mu Epsilon, and Pi Delta Phi (French) honor fraternities. He was a member of Association for Computing Machinery and for two years was a director and teacher in the ACM Student Chapter's program for teaching FORTRAN computer programming to special area high school groups.

During his undergraduate schooling he was employed full-time with Milwhite, Inc., and with Baroid Division, National Lead. His extra-curricular activities included Trustee, Treasurer, Adult Class Teacher, Music Director, in the New Iberia Bible Church; secretary and president of the Dodson School Parent Teacher's Association.

He entered graduate school at University of Southwestern Louisiana by taking graduate courses during his last semester as an undergraduate in Spring of 1967, at which time he was also awarded a teaching assistantship in mathematics.

He is a candidate for the degree Master of Science.