ON BLOCK RELAXATION TECHNIQUES

by

D. Boley, B. L. Buzbee

and

S. V. Parter

Computer Sciences Technical Report #318

June 1978

UNIVERSITY OF WISCONSIN-MADISON
COMPUTER SCIENCES DEPARTMENT

ON BLOCK RELAXATION TECHNIQUES*

D. Boley[(1)], B. L. Buzbee[(2)]
and S. V. Parter[(3)]

Computer Sciences Technical Report #318

ABSTRACT

In connection with efforts to utilize the CRAY-1 computer efficiently, we present some methods of analysis of rates of convergence for block iterative methods applied to the model problem. One of the more interesting methods involves relaxing on $p \times p$ blocks of points. A Cholesky decomposition is used for that smaller problem. One of the basic methods of analysis is a modification of a method discussed earlier by Parter. This analysis easily extends to more general second order elliptic problems.

AMS (MOS) Subject Classification: 65F10

Key Words: Model Problem, Iterative Methods, Rates of Convergence.

Work Unit Number 7: Numerical Analysis

---

*Will also appear as MRC Technical Summary Report #1860

---

(1) Stanford University; Stanford, California.
(2) Los Alamos Scientific Laboratory; Los Alamos, New Mexico.
(3) University of Wisconsin; Madison, Wisconsin.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024; by the Los Alamos Scientific Laboratory under Contract No. W-7405-ENG-36; and by the Office of Naval Research under Contract No. N00014-76-C-0341.

MRC Technical Summary Report # 1960

ON BLOCK RELAXATION TECHNIQUES

D. Boley, B. L. Buzbee
and S. V. Parter

(Received April 10, 1978)

Sponsored by

U.S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

Office of Naval Research
Arlington, Va. 22217

Los Alamos Scientific
Laboratory
Los Alamos, N. M. 87545

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

ON BLOCK RELAXATION TECHNIQUES*

D. Boley[1], B. L. Buzbee[2]
and S. V. Parter[3]

Technical Summary Report #

ABSTRACT

In connection with efforts to utilize the CRAY-1 computer efficiently, we present some methods of analysis of rates of convergence for block iterative methods applied to the model problem. One of the more interesting methods involves relaxing an $p \times p$ blocks of points. A Cholesky decomposition is used for that smaller problem. One of the basic methods of analysis is a modification of a method discussed earlier by Parter. This analysis easily extends to more general second order elliptic problems.

AMS (MOS) Subject Classification: 65F10

Key Words: Model Problem, Iterative Methods, Rates of Convergence.

Work Unit Number 7: Numerical Analysis

*Will also appear as MRC Technical Summary Report #1860

(1) Stanford University; Stanford, California.
(2) Los Alamos Scientific Laboratory; Los Alamos, New Mexico.
(3) University of Wisconsin; Madison, Wisconsin.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024; by the Los Alamos Scientific Laboratory under Contract No. W-7405-ENG-36; and by the Office of Naval Research under Contract No. N00014-76-C-0341.

ON BLOCK RELAXATION TECHNIQUES

by

D. Boley, B. L. Buzbee
and S. V. Parter

Computer Sciences Department
University of Wisconsin, Madison

(Received April 10, 1978)

Sponsored by

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709

Los Alamos Scientific
Laboratory
Los Alamos, N.M. 87545

Office of Naval Research
Arlington, Va. 22217

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

ON BLOCK RELAXATION TECHNIQUES*

D. Boley(1), B. L. Buzbee(2)
and S. V. Parter(3)

Technical Summary Report #

ABSTRACT

In connection with efforts to utilize the CRAY-1 computer efficiently, we present some methods of analysis of rates of convergence for block iterative methods applied to the model problem. One of the more interesting methods involves relaxing an $p \times p$ blocks of points. A Cholesky decomposition is used for that smaller problem. One of the basic methods of analysis is a modification of a method discussed earlier by Parter. This analysis easily extends to more general second order elliptic problems.

AMS (MOS) Subject Classification: 65F10

Key Words: Model Problem, Iterative Methods, Rates of Convergence.

Work Unit Number 7: Numerical Analysis

*Will also appear as MRC Technical Summary Report #1860

(1) Stanford University; Stanford, California.
(2) Los Alamos Scientific Laboratory; Los Alamos, New Mexico.
(3) University of Wisconsin; Madison, Wisconsin.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024; by the Los Alamos Scientific Laboratory under Contract No. W-7405-ENG-36; and by the Office of Naval Research under Contract No. N00014-76-C-0341.

SIGNIFICANCE AND EXPLANATION

In the solution of very large systems of linear algebraic equations one is frequently led to consider "iterative methods". In these methods one chooses a first guess, say U^0, and then successively computes other guesses $U^1, U^2, \cdots$. The basic questions about such a method are (i) Does the method work? That is, assuming I can keep up the procedure, does U^k get near to the true answer U? (ii) How expensive is the method? That is, assuming I can estimate the cost per iteration, how many iterations are required to decrease the initial error $|U - U^0|$ by some fixed "small" factor, say δ. Both of these questions are answered by obtaining information on a quantity ρ which is associated with the iterative method. This report discusses some wavs of obtaining good asymptotic estimates for ρ <u>in an important class of problems</u> associated with second order elliptic partial differential equations.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

ON BLOCK RELAXATION TECHNIQUES*

D. Boley(1), B. L. Buzbee(2)
and S. V. Parter(3)

1. Introduction

Some 15-20 years ago there was a great interest in iterative methods for elliptic difference equations - see [13], [14], [15], [7], [9], [10]. More recently there has been a greater emphasis on direct methods for these sparse matrices - see [5], [6], [11], [12].

However, with the advent of "vector machines" and "parallel processors" we have found it necessary to return to a consideration of certain iterative methods.

The CRAY-1 computer can perform up to 250 million floating point operations per second [2]. Algorithms that execute with high arithmetic efficiency on this computer must "fit the architecture" of it and be carefully programmed in assembly language. Thus in using this computer, we seek computational modules that can be implemented efficiently and that can be used in solving diverse problems. The solution of banded positive definite linear systems is such a module, and the Cholesky decomposition algorithm for it can be implemented on the CRAY-1 such that its execution proceeds at the rate of about 100 million floating point operations per second. Since the vector registers of the CRAY-1 can hold at most 64 numbers, implementation of banded Cholesky is simplified if vector lengths do not exceed 64. Block Relaxation techniques for solving elliptic difference approximations require the solution of banded positive definite linear systems. These facts led us to investigate the convergence rate of block successive over-relaxation for the model problem using $p \times p$ blocks (preferably $p \leq 64$).

---

*Will also appear as MRC Technical Summary Report #1860

(1) Stanford University; Stanford, California.
(2) Los Alamos Scientific Laboratory; Los Alamos, New Mexico.
(3) University of Wisconsin; Madison, Wisconsin.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024; by the Los Alamos Scientific Laboratory under Contract No. W-7405-ENG-36; and by the Office of Naval Research under Contract No. N00014-76-C-0341.

In [9] one of us developed a fairly general theory for obtaining such estimates on the rates of convergence of iterative methods for elliptic difference equations. However, partly because of the generality of that work (variable coefficients, general domains, etc.) it is by no means a transparent discussion. On the other hand, in the case of the model problem it is relatively easy to develop this general approach. This is partly due to the strong estimates of [1] and [8].

In section 2 we describe the model problem and iterative methods for its solution. In section 3 we develop the general theory (for this special case). In section 4 we obtain the rates of convergence estimates for the $p \times p$ block method mentioned earlier. In section 5 we apply the theory to the multi-line methods (these methods have been studied earlier [9], [10]).

Finally, because it is worthwhile for the practical worker to have available many methods for getting information on rates of convergence (some work here - others there), in section 6 we return to multi-line methods. Using a completely different technique we obtain upper bounds (unfortunately: not sharp bounds) on the rates of convergence.

2. The Model Problem

Let

2.1) $$\Omega \equiv \{(x,y);\quad 0 < x,y < 1\} .$$

Let P be a fixed integer and set

$$h = \frac{1}{P+1} .$$

Consider the set of interior mesh points

2.2) $$\Omega(h) = \{(x_k,y_j) = (kh,jh)\}, \qquad 1 \le k,j \le P$$

as well as the boundary mesh points

2.3) $$\partial\Omega(h) \equiv \{(x_k,y_j);\quad (k = 0 \text{ or } P+1) \text{ or } (j = 0 \text{ or } P+1)\} .$$

Let $U = \{u_{kj}\}$ be a vector defined on the set of all grid points: $\Omega(h) \cup \partial\Omega(h)$ that is, u_{kj} is the value of U at (x_k,y_j). We call U a "grid vector". In differing circumstances we will choose different orderings of the components of U.

As usual, we define the discrete Laplace operator by

2.4a) $$(\Delta_h U)_{kj} = \frac{u_{k+1,j} - 2u_{kj} + u_{k-1,j}}{h^2} + \frac{u_{k,j+1} - 2u_{kj} + u_{k,j-1}}{h^2}, \quad 1 \le k,j \le P .$$

Note: While U is defined on the entire mesh region, $\Delta_h U$ is defined only on $\Omega(h)$, the interior. Also, we define the difference operators

2.4b) $$[\nabla_x U]_{k,j} = \frac{u_{k,j} - u_{k-1,j}}{h}, \qquad 1 \le j \le P, \quad 1 \le k \le P+1 ,$$

2.4c) $$[\nabla_y U]_{k,j} = \frac{u_{k,j} - u_{k,j-1}}{h}, \qquad 1 \le j \le P+1, \quad 1 \le k \le P .$$

The basic problem is: Given grid vectors F and G, find a grid vector U such that

2.5a) $$\Delta_h U = F, \text{ in } \Omega(h) ,$$

2.5b) $$U = G, \text{ on } \partial\Omega(h) .$$

After an ordering of the points (x_k,y_j) is determined we let A be the matrix representation of $-h^2\Delta_h$; symbolically, we write

2.6) $$A \sim -h^2\Delta_h .$$

As we have already remarked Δ_h maps vectors with $P^2 + 4P$ components into vectors with P^2 components. The matrix A actually is a square P^2 by P^2 matrix. The known boundary values, G, are put on the right-hand-side. In this way the difference equations (2.5a), (2.5b) take the form

2.7) $$AV = \tilde{F}$$

where the $\tilde{}$ over F is meant to indicate both the result of ordering the components of $-h^2F$ and the necessary modifications of F required by the G terms. In any case, <u>every vector V with P^2 components may be thought of as a grid vector which also satisfies</u>

2.8) $$V = 0 \quad \text{on} \quad \partial\Omega(h) .$$

An iterative method for the solution of (2.7) is determined by a "splitting"

2.9a) $$A = M - N .$$

Equation (2.7) is then

2.9b) $$MV = NV + \tilde{F} .$$

After choosing a first guess V^0, one obtains $V^1, V^2, \ldots, V^k, \ldots$ from

2.10) $$MV^{k+1} = NV^k + \tilde{F} .$$

Let

2.11) $$\rho = \max\{|\lambda|; \det(\lambda M - N) = 0\} .$$

It is well known that the iterates V^k converge to the unique solution V of (2.7) if and only if (independently of V^0)

2.12) $$\rho < 1 .$$

The problem studied in this report is: find the asymptotic behaviour of ρ as $h \to 0$.

<u>Remark</u>: Of course, for every λ which is a generalized eigenvalue, (i.e. $\det(\lambda M - N) = 0$) there is a vector $U \neq 0$ such that

2.13) $$\lambda MU = NU .$$

3. A General Approach

We make some assumptions about the splitting (2.9a).

A.1) $$M = M^* \text{ and is positive definite}$$

A.2) $$\rho = \max_{x \neq 0} \frac{\langle Nx, x \rangle}{\langle Mx, x \rangle}$$

where

$$\langle x, y \rangle = x^T \bar{y} = \Sigma\, x_{kj} \bar{y}_{kj} \ .$$

Note: Since $A = A^*$, $M = M^*$ then $N = N^*$; and, as is well-known [4] the generalized eigenvalues are all real and

$$\rho = \max_{x \neq 0} \frac{|\langle Nx, x \rangle|}{\langle Mx, x \rangle} \ .$$

Thus, the force of the assumption (A.2) is that $\max|\lambda|$ occurs for a positive eigenvalue $\lambda = \rho$.

A.3) There is a positive constant N_0, independent of h, such that

$$||N||_\infty \leq N_0 \ .$$

Here

$$||N||_\infty = \sup\{|(NU)_{kj}|;\ |u_{kj}| \leq 1\} \ .$$

Finally we come to the main new concept.

A.4) There are positive constants q, K, independent of h, such that: if U is a grid vector satisfying

(i) $$U = 0 \quad \text{on} \quad \partial\Omega(h)$$

and

(ii) $$|\nabla_x U| + |\nabla_y U| \leq B$$

for some constant B, then

3.1) $$\langle NU, U \rangle = q \langle U, U \rangle + E$$

where

3.1a) $$|E| \leq KB/h \ .$$

<u>Remark</u>: As one might imagine, the determination of q and the verification of (A.4) is the important technical aspect of this analysis when applied to any particular case. However, as we shall see in sections 4 and 5, it is not too difficult.

Lemma 3.1: Suppose the splitting (2.9a) satisfies (A.1) and (A.2). Then the method is convergent. That is;

3.2) $$\rho < 1 .$$

Proof: Let U be the eigenvector associated with ρ. Then $\langle NU,U\rangle \geq 0$. Since $M = A + N$ and A is positive definite, we have

$$0 \leq \rho = \frac{\langle NU,U\rangle}{\langle MU,U\rangle} = \frac{\langle NU,U\rangle}{\langle AU,U\rangle + \langle NU,U\rangle} < 1 .$$

The basic result of this section is

Theorem 3.1: Suppose the splitting (2.9a) satisfies the conditions (A.1), (A.2), (A.3) and (A.4). Then

3.3) $$\rho = 1 - \frac{2\pi^2}{q} h^2 + O(h^3) .$$

Proof: Let U be the grid vector

3.4) $$u_{kj} = (\sin k\pi h)(\sin j\pi h) .$$

Then U satisfies conditions (i), (ii) of (A.4). In particular, because of (i) we may speak of $\langle NU,U\rangle$ and $\langle U,U\rangle$. The constant B of (ii) is 2π. The following facts are well known (see [13] particularly page 202).

3.5) $$h^2\langle U,U\rangle = \frac{1}{4}\left(\frac{P}{P+1}\right)^2$$

3.6) $$\frac{h^2\langle AU,U\rangle}{h^2\langle U,U\rangle} = 4(1 - \cos\pi h) = 2\pi^2 h^2 [1 - \frac{1}{12}(\pi h)^2 + O(h^4)] .$$

For all V which are zero on $\partial\Omega(h)$ and $V \neq 0$,

3.7) $$\frac{-h^2\langle\Delta_h V,V\rangle}{h^2\langle V,V\rangle} = \frac{1}{h^2}\left[\frac{h^2\langle AV,V\rangle}{h^2\langle V,V\rangle}\right] \geq 2\pi^2[1 - \frac{1}{12}(\pi h)^2 + O(h^4)] .$$

Since $M = A + N$

$$\rho \geq \frac{\langle NU,U\rangle}{\langle MU,U\rangle} = \frac{h^2\langle NU,U\rangle}{h^2\langle AU,U\rangle + h^2\langle NU,U\rangle} .$$

Applying (A.4) we have

$$h^2\langle NU,U\rangle = q[h^2\langle U,U\rangle] + h^2 E .$$

And, using (3.5) we have

$$h^2\langle NU,U\rangle = [q + O(h)][h^2\langle U,U\rangle] .$$

Thus

$$\rho \geq \frac{1}{1 + \dfrac{h^2\langle AU,U\rangle}{[q + O(h)][h^2\langle U,U\rangle]}} .$$

Using (3.6) we obtain

3.8) $$\rho \geq 1 - \frac{2\pi^2 h^2}{q} + O(h^3) .$$

In order to obtain the reverse inequality we require some basic estimates of [1] and [8]. These are

Lemma 3.2: Let V be a grid vector which is zero on $\partial\Omega(h)$. Then

3.9) $$|v_{kj}| \leq \frac{1}{4}\sqrt{1 + \pi}\,\{h^2\langle\Delta_h V,\Delta_h V\rangle\}^{1/2} .$$

Proof: See lemma 8, page 304 of [8].

Lemma 3.3: Let V be a grid vector which is zero on $\partial\Omega(h)$. Then

3.10a) $$|\nabla_x V| \leq \max|\Delta_h V| ,$$

3.10b) $$|\nabla_y V| \leq \max|\Delta_h V| .$$

Proof: This result is contained in Theorem 5, page 488 of [1].

For convenience of notation, for every grid vector V, restricted to the interior $\Omega(h)$, we write

3.11) $$||V||_g = \{h^2\langle V,V\rangle\}^{1/2} .$$

Returning to the proof of the theorem, let U be the eigenvector associated with ρ and normalized so that

3.12) $$||U||_g = 1 .$$

Then

$$\rho MU = NU$$

$$\rho AU = \rho(M - N)U = (1 - \rho)NU .$$

That is

3.13a) $$-\Delta_h U = \mu N U$$

where

3.13b) $$\mu = (1 - \rho)/\rho h^2 \ .$$

From lemma 3.1 and (3.8) we see that

3.14a) $$0 < \mu$$

and

3.14b) $$\limsup_{h\to 0} \mu \le \frac{2\pi^2}{q} \ .$$

Moreover, the theorem will be proven if we show that

$$\mu = \frac{2\pi^2}{q} + O(h) \ .$$

We write (3.13a) as

$$-\Delta_h U = \psi$$

where, if h is small enough

$$\|\psi\|_g \le \frac{4\pi^2}{q} N_0 = N_1 \ .$$

Applying lemma 3.2 we see that

$$|u_{kj}| \le \frac{1}{4}\sqrt{1+\pi}\, N_1 \ .$$

Thus

$$|\psi_{kj}| \le \frac{4\pi^2}{q} N_0 [\frac{1}{4}\sqrt{1+\pi}\, N_1] = N_2 \ .$$

Applying lemma 3.3 we have (ii) of (A.4) with $B = 2N_2$. Hence, using (A.4) we have

$$h^2 \langle NU,U\rangle = q[h^2\langle U,U\rangle] + h^2 E \ .$$

Or, making use of (3.12)

3.15) $$h^2\langle NU,U\rangle = [q + O(h)][h^2\langle U,U\rangle] \ .$$

From (3.13a) we have

$$-h^2\langle\Delta_h U,U\rangle = \mu h^2\langle NU,U\rangle$$

$$= \mu[q + O(h)][h^2\langle U,U\rangle] \ .$$

Hence, from (3.7)

$$2\pi^2[1 + O(h^2)] \le \mu[q + O(h)] \ .$$

Thus, combining this result with (3.14b), the theorem is proven.

4. $p \times p$ Blocks

Let p be a fixed integer and assume that

4.1) $$P = pQ .$$

Of course, as $P \to \infty$ (i.e. $h \to 0$) $Q \to \infty$ and vice-versa.

The interior grid vector is arranged into sub grid vectors $\{U_{rs}\}$ of p^2 entrees as follows

4.2) $$U_{rs} = \{u_{(r-1)p+\sigma,(s-1)p+\mu};\ 1 \le \sigma,\mu \le p\}, \quad 1 \le r,s \le Q .$$

Within U, the U_{rs} are ordered as follows

4.3) $$U = \{U_{11}, U_{21}, \cdots U_{Q1}, U_{12}, U_{22}, \cdots U_{Q2}, \cdots U_{1Q}, \cdots U_{QQ}\}^T .$$

That is; we start at the bottom row of $p \times p$ blocks and count off from left to right; then to the next (second) row of $p \times p$ blocks - again from left to right, etc. Within each block (U_{rs}) the subgrid vector is ordered in the same manner. To be specific, let $G(r,s,\mu)$, $\mu = 1,2,\ldots,p$ be the p vector of grid values associated with the μ^{th} horizontal line within the (r,s) block. That is

4.4) $$G(r,s,\mu) = \begin{bmatrix} u_{(r-1)p+1,(s-1)p+\mu} \\ u_{(r-1)p+2,(s-1)p+\mu} \\ \vdots \\ u_{(r-1)p+\sigma,(s-1)p+\mu} \\ \vdots \\ u_{rp,(s-1)p+\mu} \end{bmatrix} ,$$

then

4.5) $$U_{rs} = \begin{bmatrix} G(r,s,1) \\ \vdots \\ G(r,s,\mu) \\ \vdots \\ G(r,s,p) \end{bmatrix} .$$

The discrete Poisson equation (2.5a), (2.5b) takes the form

4.6) $$TU_{rs} - A_{-1}U_{r-1,s} - A_1U_{r+1,s} - B_{-1}U_{r,s-1} - B_1U_{r,s+1} = F_{rs}$$

where $T, A_{-1}, A_1, B_{-1}, B_1$ are $p^2 \times p^2$ matrices. Each is block tridiagonal of "pseudo order" p and each block is a $p \times p$ matrix. Specifically

$$T = [-I_p, R_p, -I_p] \quad \text{"block tridiagonal"}$$

$$R_p = [-1,4,-1] \quad \text{tridiagonal} .$$

If $E_{\alpha\beta}$ is the $p \times p$ matrix with "1" in the (α,β) position and zero elsewhere, then

4.7a) $$A_{-1} = \text{diagonal}[E_{1p}, E_{1p}, \ldots, E_{1p}] ,$$

4.7b) $$A_1 = \text{diagonal}[E_{p1}, E_{p1}, \ldots, E_{p1}] .$$

Notice that

$$E_{1p} = \begin{bmatrix} & 1 \\ \bigcirc & \end{bmatrix}, \quad E_{p1} = E_{1p}^T .$$

The matrix B_{-1} is the "block" E_{1p} while B_1 is the "block" E_{p1}. That is

4.8) $$B_{-1} = \begin{bmatrix} & I_p \\ \bigcirc & \end{bmatrix}, \quad B_1 = \begin{bmatrix} & \bigcirc \\ I_p & \end{bmatrix} .$$

We rewrite (4.6) as

4.9) $$MU = NU + F$$

where $M = \text{diagonal}[T,T,\ldots,T]$ and N is made up of A_{-1}, A_1, B_{-1}, B_1.

We see at a glance that M is positive definite and <u>(A.1) is satisfied</u>. Furthermore, we are dealing with a "block" five point star, that is, the equations have the same block structure as the original problem. Therefore, our splitting has "block property A". Thus, we have the basic result, if λ is an eigenvalue of $\det\{\lambda M - N\} = 0$, <u>so is $-\lambda$</u> (see [14], [15]). <u>Therefore (A.2) is satisfied.</u>

Now, N only includes the coefficients in Δ_h which relate points in the (r,s) block with neighboring points in the four blocks $(r+1,s)$, $(r-1,s)$, $(r,s+1)$, $(r,s-1)$. We see that each row of N <u>not corresponding to a corner point</u> of the (r,s) block has at most one "1" and all other entrees are "0". The rows corresponding to corners lead to exactly two "1"'s. Thus

4.10) $$\|N\|_\infty = 2 ,$$

and <u>(A.3) is satisfied with $N_0 = 2$.</u>

Finally we turn our attention to the determination of q and the verification of (A.4).

Lemma 4.1: Suppose U is a grid vector which satisfies (i), (ii) of (A.4). Then

4.11a) $$\langle NU,U\rangle = \frac{4}{p}\langle U,U\rangle + E$$

where

4.11b) $$|E| \leq \frac{16B^2}{h} .$$

That is, (A.4) is satisfied with

4.12a) $$q = \frac{4}{p}$$

and

4.12b) $$K = 16B^2 .$$

Proof: We have

4.13) $$\langle NU,U\rangle = \sum_{r,s} U^T_{rs}(NU)_{rs} .$$

Consider a term

4.14) $$U^T_{rs}(NU)_{rs} = U^T_{rs}A_{-1}U_{r-1,s} + U^T_{rs}A_1U_{r+1,s} + U^T_{rs}B_{-1}U_{r,s-1} + U^T_{rs}B_1U_{r,s+1} .$$

It is easy to see that

4.15) $$U^T_{rs}A_{-1}U_{r-1,s} = \sum_{\mu=1}^{p} u_{(r-1)p,(s-1)p+\mu} \cdot u_{(r-1)p+1,(s-1)p+\mu} .$$

Fix σ, $1 \leq \sigma \leq p$. We use (ii) to write

$$\begin{aligned} F_\mu &= u_{(r-1)p,(s-1)p+\mu} \cdot u_{(r-1)p+1,(s-1)p+\mu} \\ &= [u_{(r-1)p+\sigma,(s-1)p+\mu} + \sigma h\beta_1][u_{(r-1)p+\sigma,(s-1)p+\mu} + \sigma h\beta_2] \end{aligned}$$

where

$$|\beta_j| \leq B, \qquad j = 1,2 .$$

Thus

$$F_\mu = [u_{(r-1)p+\sigma,(s-1)p+\mu}]^2 + 2\varepsilon_1 h + \varepsilon_2 h^2$$

where

$$|\varepsilon_j| \leq (pB)^2, \qquad j = 1,2 .$$

Therefore, we may replace F_μ by the average over σ, $1 \le \sigma \le p$. Thus

$$U_{rs}^T A_{-1} U_{r-1,s} = \sum_{\mu=1}^{p} F_\mu = \frac{1}{p} \sum_{\mu=1}^{p} \sum_{\sigma=1}^{p} [u_{(r-1)p+\sigma,(s-1)p+\mu}]^2 + E_1$$

where

$$|E_1| \le 2(pB)^2 h(1+h) .$$

Each of the other terms in the right-hand-side of (4.14) may be treated in a similar manner. We obtain

$$U_{rs}^T (NU)_{rs} = \frac{4}{p} U_{rs}^T U_{rs} + E_2$$

where

$$|E_2| \le 16(pB)^2 h .$$

Finally, using (4.13) we see that

$$\langle U, NU \rangle = \frac{4}{p} \langle U, U \rangle + E$$

where

$$|E| \le 16(pB)^2 Q^2 h \le (16B^2) \frac{1}{h} .$$

Corollary: If one considers the $p \times p$ block Jacobi iterative method described by (4.6)-(4.9) then

$$\rho = 1 - \left(\frac{\pi^2}{2} p\right) h^2 + O(h^3) .$$

Proof: Apply Theorem 3.1.

We close this section with a consideration of the successive over-relaxation iterative method based on this splitting.

Let a parameter ω be chosen. Then the successive over relaxation method based on (4.6) is given by

$$\frac{1}{\omega} TU_{r,s}^{k+1} = A_{-1} U_{r-1,s}^{k+1} + B_{-1} U_{r,s-1}^{k+1} + A_1 U_{r+1,s}^{k} + B_1 U_{r,s+1}^{k} + \left(\frac{1}{\omega} - 1\right) U_{rs}^{k} + \tilde{F}_{rs} .$$

Because the basic splitting satisfies block property A the number $\rho(\omega)$ which is the related spectral radius satisfies the equation (see [15])

$$(\rho(\omega) + \omega - 1)^2 = \omega^2 \rho^2 \rho(\omega) .$$

Thus, having determined ρ, we know $\rho(\omega)$.

5. p Line Method: I

Again, let p be a fixed integer and assume that (4.1) holds.

The interior grid vector is arranged into subgrid vectors $\{U_j\}$ of pP entrees as follows

5.1) $$U_j = \{u_{\sigma,(j-1)p+\mu}; \quad 1 \le \sigma \le P, \quad 1 \le \mu \le p\} .$$

That is, U_j consists of the values associated with the j^{th} block of p lines. We now have

5.2) $$U = \{U_1, U_2, \ldots, U_Q\}^T .$$

Within each U_j the ordering is the same. That is, let $G(j,\mu)$ be the P vector associated with the μ^{th} horizontal line within the j^{th} block of p horizontal lines, i.e.

$$G(j,\mu) = \begin{bmatrix} u_{1,(j-1)p+\mu} \\ u_{2,(j-1)p+\mu} \\ \vdots \\ u_{\sigma,(j-1)p+\mu} \\ \vdots \\ u_{P,(j-1)p+\mu} \end{bmatrix},$$

then

$$U_j = \begin{bmatrix} G(j,1) \\ G(j,2) \\ \vdots \\ G(j,p) \end{bmatrix} .$$

The discrete Poisson equation (2.5a), (2.5b) now takes the form

5.3) $$TU_j = RU_{j-1} + R^T U_{j+1} + \tilde{F}_j$$

where T and R are pP by pP matrices. In fact, T is block tridiagonal with

5.4a) $$T = [-I_P, T_P, -I_P]_p \qquad T_P = [-1,4,-1]_P$$

and

5.4b) $$R = \begin{bmatrix} & I_p \\ O & \end{bmatrix} .$$

This decomposition is used to make the splitting

5.5) $$A = M - N$$

where

5.6a) $$M = \text{diagonal}(T)$$

5.6b) $$N = [R, 0, R^T] .$$

It is immediately clear that $M = M^*$ and is positive definite since each T is positive definite. Once more, this splitting satisfies block property A (see [9], [10]). Thus (A.1) and (A.2) are satisfied. From the structure of N we see that (A.3) is satisfied with $N_0 = 2$.

Once more we seek to determine an appropriate q and verify (A.4).

<u>Lemma 5.1</u>: Suppose U is a grid vector which satisfies (i) and (ii) of (A.4). Then

5.7a) $$\langle NU, U \rangle = \frac{2}{p} \langle U, U \rangle + E$$

where

5.7b) $$|E| \le 8B^2 p .$$

That is, (A.4) is satisfied with

5.8a) $$q = 2/p$$

and

5.8b) $$K = 8B^2 p .$$

<u>Proof</u>: We have

5.9) $$\langle NU, U \rangle = \sum_{j=1}^{Q} U_j^T (NU)_j .$$

Consider a term

5.10) $$U_j^T (NU)_j = U_j^T R U_{j-1} + U_j^T R^T U_{j+1} .$$

Now

$$U_j^T R U_{j-1} = \sum_{\sigma=1}^{p} u_{\sigma,(j-1)p+1} \cdot u_{\sigma,(j-1)p} .$$

Fix μ, $1 \le \mu \le p$. Then

$$[u_{\sigma,(j-1)p+1}][u_{\sigma,(j-1)p}] = [u_{\sigma,(j-1)p+\mu} + \varepsilon_1][u_{\sigma,(j-1)p+\mu} + \varepsilon_2]$$

where

$$|\varepsilon_j| \le Bph \; .$$

Therefore, averaging once more, we have

$$U_j^T R U_{j-1} = \frac{1}{p}\left[\sum_{\mu=1}^{p} \sum_{\sigma=1}^{p} [u_{\sigma,(p-1)j+\mu}]^2 + 4hp^2B^2\right] .$$

Thus, as in section 4

$$\langle U, NU \rangle = \frac{2}{p}\langle U, U\rangle + E$$

where

$$|E| \le 8B^2 p \; .$$

Corollary: If we consider the p line iterative method described by (5.3)-(5.6b), then

$$\rho = 1 - p\pi^2 h^2 + O(h^3) \; .$$

Proof: Apply theorem 3.1.

Remark: A careful look at this section shows that $K = 8B^2 ph$ and hence we easily obtain

$$\rho = 1 - p\pi^2 h^2 + O(h^4) \; .$$

6. p-line Method II

In this section we approach the p-line method of section 5 with another method of analysis. The results obtained are weaker, but the approach may well have applications in cases where the analysis of section 3 does not apply.

Lemma 6.1: Let $u_n(x)$ denote the Chebychev polynomial of the second kind of order n, i.e., $u_0(x) = 1$, $u_1(x) = 2x$ and $u_{n+1}(x) = 2xu_n(x) - u_{n-1}(x)$.

If $x \geq 1$, then

6.1a) $$u_n(x) \geq u_{n-1}(x) + 1, \qquad n \geq 1$$

6.1b) $$u_n(x) \geq n + 1, \qquad n \geq 1$$

and

6.1c) $$\frac{d}{dx} u_n(x) \geq \frac{d}{dx} u_{n-1}(x) + 2n .$$

Proof: Apply induction.

Corollary: $u_n'(x) \geq 2$, $n \geq 1$ and $x \geq 1$.

Lemma 6.2: Let $B = [-I, 2S, -I]_m$ where S and I are $n \times n$. Let

$$U_j(S) = u_j(S) ,$$

that is, $U_j(S)$ is an $n \times n$ matrix obtained by evaluating u_j, the Chebychev polynomial, at S. If $U_m(S)$ is nonsingular, then

6.2) $$B_{ij}^{-1} = \begin{cases} U_m^{-1}(S)U_{j-1}(S)U_{m-i}(S), & j \leq i , \\ U_m^{-1}(S)U_{i-1}(S)U_{m-j}(S), & i \leq j . \end{cases}$$

Proof: See Theorem 1 of [3].

The quantity of interest, ρ, is the spectral radius of $M^{-1}N$. Since

$$M^{-1}N = [T^{-1}R, O, T^{-1}R^T]$$

(see section 5) we may apply Lemma 6.2 to obtain T^{-1} and hence $M^{-1}N$. We find that

$$M^{-1}N = [D, O, E]$$

where

6.3a)
$$D = \begin{bmatrix} & U_p^{-1}U_{p-1} \\ & U_p^{-1}U_{p-2} \\ \bigcirc & \vdots \\ & U_p^{-1}U_0 \end{bmatrix} ,$$

6.3b)
$$E = \begin{bmatrix} U_p^{-1}U_0 & \\ U_p^{-1}U_1 & \\ \vdots & \bigcirc \\ U_p^{-1}U_{p-1} & \end{bmatrix} ,$$

and

6.3c)
$$U_j = U_j(\tfrac{1}{2} T_P) .$$

If Q denotes the unitary matrix which diagonalizes T_P (and hence $\frac{1}{2} T_P$), then

6.4a)
$$\begin{cases} Q^{-1}U_j(\frac{1}{2} T_P)Q = U_j(\frac{1}{2} Q^{-1}T_PQ) \\ \qquad = \mathrm{diag}\{u_j(\lambda_r)\}, \quad r = 1,2,\ldots,P , \end{cases}$$

and

6.4b)
$$\begin{cases} Q^{-1}U_p^{-1}(\frac{1}{2} T_P)Q = U_p^{-1}(\frac{1}{2} Q^{-1}T_PQ) \\ \qquad = \mathrm{diag}\{u_p^{-1}(\lambda_r)\}, \quad r = 1,2,\ldots,P , \end{cases}$$

where

6.4c)
$$\lambda_r = 2 - \cos(r\pi h) > 1$$

is the r'th eigenvalue of $\frac{1}{2} T_P$.

Let

$$\hat{Q} = \mathrm{diag}\{Q,Q,\ldots,Q\}$$

we see that

6.5)
$$\hat{Q}^{-1}M^{-1}N\hat{Q} = [Q^{-1}DQ,O,Q^{-1}EQ] .$$

Thus, applying the Gershgorin circle theorem

6.6) $$\rho \le B_p \equiv \max_{\substack{r \\ 1\le j\le p}} \{u_p^{-1}(\lambda_r)[u_{p-j}(\lambda_r) + u_{j-1}(\lambda_r)]\} .$$

Lemma 6.3: If $x \ge 1$, then

6.7) $$u_0(x) + u_{p-1}(x) \ge u_{j-1}(x) + u_{p-j}(x), \qquad j = 1,2,\ldots,p .$$

Proof: For $x \ge 1$ and $i \ge 2$, from the basic recursion formula and (6.1b) we have

$$u_i(x) - u_{i-1}(x) \ge u_{i-1}(x) - u_{i-2}(x) .$$

Thus, if $n > m > 0$ we have

$$u_n(x) - u_{n-1}(x) \ge u_m(x) - u_{m-1}(x) ,$$

or

6.8) $$u_n(x) + u_{m-1}(x) \ge u_m(x) + u_{n-1}(x) .$$

Of course (6.7) is true for $j = 1$ and $j = p$. Assume that (6.7) is true for a value of $j = \sigma$, $1 \le \sigma \le p - 2$. Then

6.9) $$u_0(x) + u_{p-1}(x) \ge u_{\sigma-1}(x) + u_{p-\sigma}(x) .$$

We may assume $\sigma < p - \sigma$. Then applying (6.8) with $n = p - \sigma$ and $m = \sigma$ we find

$$u_{p-\sigma}(x) + u_{\sigma-1}(x) \ge u_{p-(\sigma+1)}(x) + u_\sigma(x) .$$

That is

$$u_{p-\sigma}(x) + u_{\sigma-1}(x) \ge u_j(x) + u_{p-(j+1)}(x) .$$

Substitution of this result into (6.9) gives (6.7) for the larger value of j and the lemma is proven.

Theorem 6.1: With B_p defined by (6.6) we have

6.10) $$\rho(M^{-1}N) \le B_p = 1 - \frac{p}{2}\pi^2h^2 + O(h^4) .$$

Proof: From Lemma 6.3 we see that

$$B_p = \max_{\lambda_r} \frac{1 + u_{p-1}(\lambda_r)}{u_p(\lambda_r)} ,$$

where the λ_r are given by (6.4c). It is not difficult to see that

$$\hat{B}_p(\lambda) = \frac{1 + u_{p-1}(\lambda)}{u_p(\lambda)}$$

is a monotone non-decreasing function of λ for $\lambda \geq 1$. Thus

$$B_p = \frac{1 + u_{p-1}(2 - \cos\pi h)}{u_p(2 - \cos\pi h)} .$$

Expansion of $\hat{B}_p(\lambda)$ about $\lambda = 1$ yields (6.10).

Final Remark: Since it is better to slightly overestimate the relaxation parameter in successive overrelaxation than to underestimate it, it might appear that the estimate of this section is preferable to that of section 5 for coarse meshs. However, numerical experiments contradict this idea.

REFERENCES

[1] Brandt, Achi, "Estimates for difference quotients of solutions of Poisson type difference equations", Math. of Comp., 20 (1966), pp. 473-499.

[2] Buzbee, B. L., Golub, G. H., and Howell, J. A., "Vectorization for the CRAY-1 of some methods for solving elliptic difference equations", High Speed Computer and Algorithm Organization, Ed. David J. Kuck, Duncan H. Lawrie, and Ahmed H. Sameh, pp. 255-272, Academic Press, New York 1977.

[3] Fischer, C. F. and Usmani, R. A., "Properties of some tridiagonal matrices and their application to boundary value problems", SINUM 6 (1969), pp. 127-142.

[4] Franklin, J. L., Matrix Theory, Prentice-Hall, Englewood Cliffs, N.J., 1968.

[5] George, Alan, "Nested dissection of a regular finite element mesh", SINUM 10 (1973), pp. 345-363.

[6] ____________________, "Numerical experiments using dissection to solve n by n Grid problems", SINUM 14 (1977), pp. 161-179.

[7] Keller, H., "On some iterative methods for solving elliptic difference equations", Quart. Appl. Math. 16 (1958), pp. 209-226.

[8] Nitsche, J. and Nitsche, J. C. C., "Error estimates for the numerical solution of elliptic difference equations", Archive for Rat. Mech. and Anal. (1960), pp. 293-306.

[9] Parter, S. V., "On estimating the 'Rates of Convergence' of iterative methods for elliptic difference equations", Trans. of the A.M.S. 114 (1965), pp. 320-354.

[10] _____________, "Multi-line iterative methods for elliptic difference equations and fundamental frequencies", Numerische Math. 3 (1961), pp. 305-319.

[11] Rose, D. J., "A graph-theoretic study of the numerical solution of sparse positive definite systems of linear equations", Graph Theory and Computing, R. C. Read ed. Academic Press, New York 1972.

[12] Rose, D. J. and Willoughby, R. A. - Editors, Sparse Matrices and their Applications, Plenum Press, N.Y. 1972.

[13] Varga, R. S., "Matrix Iterative Analysis", Prentice-Hall, Englewood Cliffs, N.J., 1962.

[14] Young, David M., _Iterative Solution of Large Linear Systems_, Academic Press, New York, 1971.

[15] __________________, "Iterative methods for solving partial difference equations of elliptic type", _Trans. Amer. Math. Soc._ 76 (1954), pp. 92-111.