

THE LIBRARY
THE COMPUTING CENTRE
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER B. C.

On the Solution of Block Tridiagonal Systems
Arising from Certain Finite-Difference Equations

by

J.M. Varah

Computer Science Department
University of British Columbia
Vancouver, British Columbia
Canada

March, 1972

Supported by National Research Council of Canada grant #A8240

1. Introduction

The approximation of a time-dependent system of linear partial differential equations in one space variable by an implicit finite-difference scheme leads to a linear algebraic system with a matrix of a definite band structure. This system must be solved at each time-step and thus it is worthwhile to solve this system as efficiently as possible. When only one equation is involved, this is fairly straightforward: for example approximating the heat equation

$$u_t = u_{xx}, \quad 0 \leq x \leq 1, \quad 0 \leq t \leq T,$$

$$u(x,0) = f(x), \quad u(0,t) = g_1(t), \quad u(1,t) = g_2(t)$$

by the Crank-Nicolson scheme

$$\left(I - \frac{k}{2} D_+ D_-\right) v_j^{m+1} = \left(I + \frac{k}{2} D_+ D_-\right) v_j^m$$

leads to a tridiagonal matrix equation, for which the direct solution using Gaussian elimination is well-known (see e.g. [3, pg 56]).

Here $v_j^m = v(j\Delta x, m\Delta t)$

$k = \Delta t = \text{time-step}$

$h = \Delta x = \text{space-step} \quad (h = 1/(n-1))$

D_+, D_- are the usual forward and backward spatial difference operations

However, when the problem in question is a system of differential equations, or if it is put in system form for solution, the structure of the accompanying matrix is more complicated. In particular, let us assume that the unknown grid functions $\underline{v}_j^m = ((v_1)_j^m, (v_2)_j^m, \dots, (v_p)_j^m)$ are numbered for the matrix equation so that components are

consecutive; then the matrix has a block-banded structure with $p \times p$ blocks instead of single elements.

For many of the basic schemes used for solving such systems, the matrix is in fact block-tridiagonal, i.e.

$$A = \begin{pmatrix} B_1 & C_1 & & \\ A_2 & B_2 & C_{n-1} & \\ & \ddots & \ddots & \\ A_n & & & B_n \end{pmatrix} \quad (1)$$

with each A_i , B_i , C_i square. For example, with the parabolic system (here $\underline{u} = (u_1, \dots, u_p)$)

$$\begin{aligned} \underline{u}_t &= P(x,t) \underline{u}_{xx}, & 0 \leq x \leq 1, 0 \leq t \leq T \\ \underline{u}(x,0) &= \underline{f}(x) & \\ \left. \begin{aligned} Q_1 \underline{u}_x + Q_0 \underline{u} &= g^I(t) \\ Q_2 \underline{u} &= g^{II}(t) \end{aligned} \right\} & \text{at } x=0 \\ \left. \begin{aligned} R_1 \underline{u}_x + R_0 \underline{u} &= g^{III}(t) \\ R_2 \underline{u} &= g^{IV}(t) \end{aligned} \right\} & \text{at } x=1 \end{aligned} \quad (2)$$

the Crank-Nicolson scheme

$$\left(I - \frac{k}{2} D_+ P(x_j - \frac{h}{2}, t_{m+1}) D_- \right) \underline{v}_j^{m+1} = \left(I + \frac{k}{2} D_+ P(x_j - \frac{h}{2}, t_m) D_- \right) \underline{v}_j^m$$

with any kind of discrete boundary conditions can be expressed in the form (1) with each block of order p .

Recently, Keller has proposed new schemes for a first order system of two-point boundary value problems [4], and for parabolic systems [5]; both of these can be expressed as block-tridiagonal systems (see Section 3). Other examples of block-tridiagonal schemes can be found in Richtmyer and Morton [6].

There are two common methods for solving the system (1); treating it as a band matrix, or using its block-tridiagonal structure. Either way, for an accurate solution we need to ensure that the intermediate quantities do not become too large (we call this numerical stability). In the case of a band matrix, this can be done by partial pivoting; however this increases the amount of work necessary so it is useful to know when pivoting is not required. Sufficient conditions for this (namely A diagonally dominant or A symmetric positive definite) were first given by Wilkinson [8] and are also discussed in Wendroff [7].

In Section 2 we discuss the same problem for a general block-tridiagonal system, and in Section 3 compare the work involved in the band and block methods for the particular schemes mentioned above.

2. Solution of General Block-Tridiagonal Systems

The direct solution of (1) by block-Gaussian elimination without pivoting is well-known (see for example Isaacson and Keller [3, pg. 59]); the block-triangular decomposition can be expressed as

$$\begin{pmatrix} B_1 & C_1 & & & \\ A_2 & B_2 & C_2 & & \\ & \ddots & \ddots & \ddots & \\ & & A_n & B_n & C_{n-1} \end{pmatrix} = \begin{pmatrix} I & & & & \\ L_2 & I & & & \\ & L_3 & I & & \\ & & \ddots & I & \\ & & & L_n & I \end{pmatrix} \begin{pmatrix} U_1 & C_1 & & & \\ U_2 & & C_2 & & \\ & & & \ddots & \\ & & & & C_{n-1} \\ & & & & U_n \end{pmatrix} \quad (3)$$

or $A = LU$. The recurrence for the L_i, U_i is

$$\left. \begin{aligned} U_1 &= B_1 \\ L_i &= A_i U_{i-1}^{-1} \\ U_i &= B_i - L_i C_{i-1} \end{aligned} \right\} i = 2, \dots, n. \quad (4)$$

For numerical stability, we need a bound on the pivotal growth, i.e. $\max(\|L_i\|, \|U_i\|) \leq K$ for some norm. In the scalar case, this is guaranteed if A is diagonally dominant; we wish to show that a similar result holds here.

Definition 1: The matrix A of (1) is block diagonally dominant with respect to the matrix norm $\|\cdot\|$ if

$$\|B_i^{-1}\|(\|A_i\| + \|C_i\|) \leq 1, \quad i=1, \dots, n. \quad (5)$$

This concept has been discussed often in the literature; see for example Feingold and Varga [1].

A necessary condition for this of course is that the diagonal blocks B_i be nonsingular, and we assume this in what follows. We also assume $C_i \neq 0$, for if some $C_i = 0$ the system would be decoupled.

Theorem 2.1: If A is block diagonally dominant, then the block-triangular decomposition (3) is numerically stable, and in fact,

$$\|L_i\| \leq \frac{\|A_i\|}{\|C_{i-1}\|}, \quad \|U_i\| \leq \|B_i\| + \|A_i\|.$$

Proof: From (4), it's clear we need only a bound for $\|U_i^{-1}\|$. We show by induction that $\|U_i^{-1}\| \leq \frac{1}{\|C_i\|}$. This is clear for $i=1$ from (4) and (5). Assume for $i-1$; then

$$\begin{aligned} U_i &= B_i - A_i U_{i-1}^{-1} C_{i-1} \\ &= B_i (I - B_i^{-1} A_i U_{i-1}^{-1} C_{i-1}) \\ &= B_i (I - R) \end{aligned}$$

$$\text{Now } \|R\| \leq 1 - \|B_i^{-1}\| \|C_i\|$$

from (5) and the induction hypothesis. Thus

$$\|U_i^{-1}\| \leq \frac{\|B_i^{-1}\|}{\|B_i^{-1}\| \|C_i\|} = \frac{1}{\|C_i\|}$$

Now the above bounds on $\|L_i\|, \|U_i\|$ follow easily from (4). QED

If we assume A is block diagonally dominant by columns rather than rows, a similar analysis gives $\|L_i\| \leq 1, \|U_i\| \leq \|B_i\| + \|C_i\|$, which is the obvious generalization of Theorem 5, pg 56 of [3].

We can in fact be more general than this by including block-diagonal scaling via DAE,

$$D = \begin{pmatrix} D_1 & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & D_n \end{pmatrix}, \quad E = \begin{pmatrix} E_1 & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & E_n \end{pmatrix}$$

If $A = LU$ is the block-triangular factorization (3), then the corresponding factorization for DAE is $(DL)(UE)$, and we can apply Theorem 2.1 to this matrix. Thus we have (as in the scalar case):

Corollary: The block-triangular decomposition (3) is numerically stable if DAE is block diagonally dominant for some block diagonal D, E with $\|D\|, \|D^{-1}\|, \|E\|, \|E^{-1}\|$ bounded.

One fruitful choice of D, E is $D_i = B_i^{-1}, E_i = e_i I$.

This leads to the following:

Theorem 2.2: Let A be as in (1) and define $\alpha_i = (\|B_i^{-1} C_i\| \|B_{i+1}^{-1} A_{i+1}\|)^{\frac{1}{2}}$ (assume $\alpha_i \neq 0$). Then the block tridiagonal factorization (3) is numerically stable if the matrix

$$S = \begin{pmatrix} 1 & & & & & & & & & & \\ & \alpha_1 & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & \cdot & & & & & & & \\ & & & & \cdot & & & & & & \\ & & & & & \cdot & & & & & \\ & & & & & & \alpha_{n-1} & & & & \\ & & & & & & & 1 & & & \\ & & & & & & & & & & \end{pmatrix}$$

is positive semidefinite.

Before proving this, we should remark that this theorem includes the result of Richtmyer and Morton [6, pg 275ff] as a special case, namely with $\alpha_i = \alpha$ and the condition $\alpha \leq \frac{1}{2}$.

Proof: With $D_i = B_i^{-1}$, $E_i = e_i I$, DAE is block diagonally dominant if the following set of inequalities holds:

$$\begin{aligned} e_2 \|B_1^{-1} C_1\| &\leq e_1 \\ e_{i-1} \|B_i^{-1} A_i\| + e_{i+1} \|B_i^{-1} C_i\| &\leq e_i, \quad i=2, \dots, n-1 \\ e_{n-1} \|B_n^{-1} A_n\| &\leq e_n. \end{aligned}$$

With $\delta_i = \|B_i^{-1} C_i\|$, $\gamma_i = \|B_i^{-1} A_i\|$, this holds if there is a positive vector \underline{e} so that

$$P \underline{e} \equiv \begin{pmatrix} 1 & -\delta_1 & & & \\ -\gamma_2 & 1 & -\delta_2 & & \\ & \cdot & \cdot & \cdot & \\ & & & \cdot & -\delta_{n-1} \\ & & & -\gamma_n & 1 \end{pmatrix} \underline{e} \geq 0.$$

But this holds whenever it holds for the symmetrized matrix

$$DPD^{-1} = \begin{pmatrix} 1 & -\alpha_1 & & & \\ -\alpha_1 & 1 & -\alpha_2 & & \\ & \cdot & \cdot & \cdot & \\ & & & \cdot & -\alpha_{n-1} \\ & & & \alpha_{n-1} & 1 \end{pmatrix} \equiv I - R$$

Now if $(I-R)$ is positive definite, we know from its M-matrix properties (see for example Householder [2, pg 58]) that there is a positive vector \underline{e} so that $(I-R) \underline{e} > 0$. And if it is singular but positive semi-definite, we take for \underline{e} the null vector of $(I-R)$; it must have all

8.

positive components as it is the Perron vector of R (R is irreducible because no $\alpha_i = 0$). So a sufficient condition for numerical stability is $(I-R)$ positive semidefinite. But since R 's eigenvalues occur in $(+,-)$ pairs, this holds if and only if $I+R=S$ is positive semidefinite. QED

3. Solution of Systems: Block-Tridiagonal vs Bandsolve

We consider two kinds of systems: the general block-tridiagonal matrix (1) and a specialized form arising from the solution of first-order systems. The general form (1) arises for example in the solution of (2) via the Crank-Nicolson scheme. For a constant coefficient problem (i.e. $P=\text{constant}$), we have $A_i=C_i = -\frac{\lambda}{2}P$, $B_i=I+\lambda P$ where $\lambda = k/h^2$, and since P is positive definite, we can apply Theorem 2.2 to show the numerical stability of the block-tridiagonal factorization. For variable coefficients, we will certainly have this stability for h small enough. Treating the system as a band matrix, we may have to row pivot the first p equations (these are the boundary conditions) but this is not necessary with the rest because of the positive definiteness.

So we compare operations of bandsolve and block-tridiagonal for general systems (1) without pivoting (the blocks are $p \times p$ and there are n such blocks in each row and column).

(a) bandsolve:

(i) LU decomposition: each set of p equations requires

$$\begin{aligned} & (2p)(2p-1) + (2p-1)(2p-2) + \dots + (p+1)p \\ & = \frac{p}{3} (7p^2 - 1) \text{ multiplications} \end{aligned}$$

(ii) backsolution: both forward and backward substitutions require

$$(p+1)+\dots+2p = \frac{p}{2} (3p+1) \text{ operations per } p \text{ equations}$$

This gives a total of $np^2(\frac{7}{3}p+3)$ multiplications (carrying only $O(np^2)$ terms).

(b) blocksolve: (see(4))

(i) block LU decomposition: for $i=2, \dots, n$, we have

$$\left. \begin{array}{l} \text{LU decomposition of } U_{i-1}: p^3/3 \text{ operations} \\ \text{solution of } L_i(U_{i-1}) = A_i: p^3 \text{ operations} \\ \text{formation of } U_i = B_i - L_i C_{i-1}: p^3 \text{ operations} \end{array} \right\} \text{ total } \frac{7np^3}{3}$$

(ii) backsolution:

$$Ly = f \text{ takes } np^2$$

$$Ux = y \text{ takes } 2np^2 \text{ (using the saved LU decomposition of each } U_i)$$

This gives a total of $np^2(\frac{7}{3}p + 3)$, the same as for bandsolve.

Now consider the more specialized matrix form:

$$(6) \quad \begin{array}{c} \boxed{F_0} \\ \boxed{F_{-1}} + \boxed{G_{-1}} \\ \text{---} \boxed{F_2} \text{---} \boxed{G_2} \\ \boxed{G_0} \end{array}$$

Here all blocks are $p \times p$ except F_0 ($q \times p$) and G_0 ($p - q \times p$). When split up so all blocks are $p \times p$ (as indicated by the dotted lines), it becomes a block-tridiagonal system.

This form arises in solving the general first order system

$$\underline{u}'(x) = K(x) \underline{u}(x) + \underline{f}(x), \quad 0 \leq x \leq 1$$

with separated boundary conditions

$$F_0 \underline{u}(0) = \underline{\alpha}$$

$$G_0 \underline{u}(1) = \underline{\beta}$$

by the midpoint rule:

$$\frac{v_j - v_{j-1}}{h_j} = K(x_{j-\frac{1}{2}}) \left(\frac{v_j + v_{j-1}}{2} \right) + f(x_{j-\frac{1}{2}}) \quad (7)$$

See Keller [4] for details. Here F_0 has q rows (assumed linearly independent) and G_0 has $p-q$ rows.

This matrix also comes up in solving the parabolic system (2) by Keller's box scheme [5]. Since the problem is converted to a first order system (with unknowns u, u_x), the matrix blocks are of order $2p$ now, and $q=p$.

We again wish to compare the solution of (6) by band and block methods. First we examine the stability or pivoting problem for the midpoint rule and box scheme. For the midpoint rule, from (7) we have

$$F_j = -I - \frac{h_j}{2} K(x_{j-\frac{1}{2}}), \quad G_j = I - \frac{h_j}{2} K(x_{j-\frac{1}{2}}).$$

Thus for $\max_j h_j = h$ small enough these are nonsingular. However as a band matrix there is no guarantee that the diagonal pivots will be large, and pivoting may be necessary; for example the first column of F_0 may be all zero necessitating an initial row interchange using a row of F_1 . This of course increases the bandwidth and the work involved. We include below a calculation of the work required for bandsolve using pivoting. Similarly, for the block-tridiagonal decomposition, the diagonal blocks B_i may be singular. Thus B_1 will be singular if the q rows of F_0 and the first $(p-q)$ rows of F_1 are

dependent. But if so, we can interchange rows of F_1 (and G_1) so B_1 is nonsingular; moreover this preserves the matrix form (6). We can apply a similar "pivoting" for each such diagonal block as we proceed, and it seems clear that this will be numerically stable for h small enough. Indeed, if we assume the boundary matrix F_0 is scaled so $F_0 = (H_0 \ I)$, then for h small enough

$$U_1 = B_1 = \left(\begin{array}{c|c} H_0 & I \\ \hline I & 0 \end{array} \right) + O(h);$$

and since

$$A_2 = \left(\begin{array}{c|c} 0 & I \\ \hline 0 & 0 \end{array} \right) + O(h), \text{ we have } U_2 = \left(\begin{array}{c|c} H_0 & I \\ \hline I & 0 \end{array} \right) + O(h),$$

so the diagonal blocks U_i stay well-conditioned.

Now consider the box scheme (see Keller [5]). For simplicity, we will only consider it applied to the heat equation $u_t = u_{xx}$. This can be written as a first-order system:

$$\begin{aligned} u_x &= v \\ v_x &= u_t \end{aligned}$$

and then differenced as

$$\begin{aligned} D_{-x} u_j^n &= \frac{1}{2}(v_j^n + v_{j-1}^n) \\ D_{-x}(v_j^n + v_j^{n-1}) &= D_{-t}(u_j^n + u_{j-1}^n) \end{aligned} \quad (8)$$

This leads to an implicit equation for (u_i, v_i) on the n -th time line with matrix

$$B = \left(\begin{array}{cc|cc|ccc} \alpha_0 & \alpha_1 & 0 & 0 & & & \\ 1 & h/2 & -1 & h/2 & & & \\ \hline h/k & 1 & h/k & -1 & & & \\ 0 & 0 & 1 & h/2 & & & \\ \hline & & & & & & \\ & & & & & & \\ & & & & & & \end{array} \right)$$

Clearly we cannot guarantee the stability of bandsolve without pivoting, and similarly the block tridiagonal factorization may be unstable ($\alpha_1 = 0$ implies $\|U_1^{-1}\| = O(\frac{1}{h})$), and in fact then each $\|U_i\| = O(\frac{1}{h})$, $i=2,3,\dots,n$). However if we interchange the equations in (8) we obtain the matrix

$$B' = \begin{pmatrix} \alpha_0 & \alpha_1 & 0 & 0 & & \\ \frac{h}{k} & 1 & \frac{h}{k} & -1 & & \\ \hline 1 & \frac{h}{2} & -1 & \frac{h}{2} & & \\ 0 & 0 & \frac{h}{k} & 1 & & \\ \hline & & & & \ddots & \\ & & & & & \ddots \end{pmatrix}$$

Treating this as a band matrix, it is easy to see the bandsolve decomposition is numerically stable with a slight restriction on h and k . Indeed, because of the sign pattern, if our LU decomposition proceeds so that the third pivot B'_{33} is < 0 independent of h , then from that point on the pivots will slowly increase in magnitude (by $O(h)$ on each step), giving a stable decomposition. But this third pivot is, omitting $O(h)$ terms, $(-2 + \frac{1}{h\alpha_1}) / (1 - \frac{h}{k\alpha_0})$. If $\alpha_0\alpha_1 \leq 0$, we have $-2 \leq B'_{33} \leq -1$ with no restriction on h and k ; if $\alpha_0\alpha_1 > 0$, we have $-3 \leq B'_{33} \leq -2$ for $\frac{k}{h} \leq \frac{\alpha_1}{2\alpha_0}$ and $-1 \leq B'_{33} \leq -\frac{1}{2}$ for $\frac{k}{h} \geq \frac{3\alpha_1}{\alpha_0}$.

Treating B' as a block tridiagonal matrix, for numerical stability we again need to ensure $\frac{k}{h} \neq \frac{\alpha_1}{\alpha_0}$ so U_i is nonsingular. In fact if we carry through the algorithm, we see that

$$U_i = \begin{pmatrix} \frac{h}{ke_i} - 1 & -\frac{1}{e_i} \\ \frac{h}{k} & 1 \end{pmatrix} + O(h) \quad , \quad i=2,3,\dots,$$

where $e_2 = \frac{\alpha_0}{\alpha_1} - \frac{h}{k}$, $e_{i+1} = e_i - 2\frac{h}{k}e_i^2$, $i=2,3,\dots$. For stability, we need only to keep the e_i away from zero. Note that $e_{i+1} < e_i$ and thus we have a stable decomposition for all h and k if $\alpha_0\alpha_1 \leq 0$. If $\alpha_0\alpha_1 > 0$, it is stable if $e_2 < 0$ (for example if $\frac{k}{h} \leq \frac{1}{2} \frac{\alpha_1}{\alpha_0}$); if $\frac{\alpha_1}{\alpha_0} \ll 1$ so this would be unduly restrictive, we have stability if we choose $\frac{k}{h}$ so that $e_2 > 0$, $e_3 < 0$.

Although we have only discussed the box scheme for the heat equation, it is clear that similar considerations could be applied to the box scheme for more general parabolic equations and systems.

Now we consider the operation counts for the matrix (6) using both bandsolve and the block tridiagonal factorization. For this, we make use of the zero patterns, but make no assumptions about the nonzero elements for the sake of generality. Thus the operation times could be cut somewhat for special nonzero elements; for example in the box scheme one might take special account of the ones appearing (this is done by Keller [5]).

(a) bandsolve for (6) (no pivoting)

(i) LU decomposition: each set of p rows takes

$$\begin{aligned}
& \bar{p}(p+q-1) + (p-1)(p+q-2) + \dots + (p-q+1)p \\
& \quad + (2p-q)(p-1) + (2p-q-1)(p-2) + \dots + (p+1)q \\
= & \frac{np}{6} (5p^2 + 3pq - 3q^2 - 3p + 6q - 2) \quad \text{for all } np \text{ rows.}
\end{aligned}$$

(ii). backsolution:

$Ly = f$ takes for each set of p rows (except first)

$$[(p+1)+\dots+(p+q)] + [(q+1)+\dots+p]$$

$Ux = y$ takes for each set of p rows (except last)

$$[(p+(p-1)+\dots+(p-q+1))] + [(2p-q)+\dots+(p+1)]$$

giving a total of $np(2p+1)$ multiplications.

(b) bandsolve for (6) (with pivoting)

Now L will have the same form, but U could look like the upper triangular part of a general block-tridiagonal matrix.

(i) LU decomposition: each set of p rows takes

$$\begin{aligned}
& 2p(p+q-1) + (2p-1)(p+q-2) + \dots + (p+1)q \\
= & \frac{np}{6} (5p^2 + 9pq - 3p + 3q - 2) \quad \text{for all } np \text{ rows.}
\end{aligned}$$

(ii) backsolution:

$Ly = f$ same as (a), i.e.

$$n[(p+1)+\dots+(p+q)+(q+1)+\dots+p] = \frac{np}{2} (p+2q+1)$$

$Ux = y$ same as for general block tridiagonal, i.e.

$$\frac{np}{2} (3p+1)$$

giving a total of $np(2p+q+1)$ multiplications for the backsolution.

(c) blocksolve for (6)

(i) block LU decomposition: for $i=2, \dots, n$, we have

$$\begin{array}{l}
 \text{LU decomposition of } U_{i-1}: p^3/3 \text{ operations} \\
 \text{solution of } L_i(U_{i-1}) = A_i: qp^2 \text{ operations} \\
 \text{formation of } U_i = B_i - L_i C_{i-1}: pq(p-q) \text{ opns.}
 \end{array}
 \left. \vphantom{\begin{array}{l} \text{LU decomposition of } U_{i-1}: p^3/3 \text{ operations} \\ \text{solution of } L_i(U_{i-1}) = A_i: qp^2 \text{ operations} \\ \text{formation of } U_i = B_i - L_i C_{i-1}: pq(p-q) \text{ opns.} \end{array}} \right\} \begin{array}{l} \text{Total:} \\ np \left(\frac{p^2}{3} + 2pq - q^2 \right) \end{array}$$

(Note that L_i is zero in its last $(p-q)$ rows and C_i is zero in its first q rows.)

(ii) backsolution:

$Ly = f$ takes npq operations

$Ux = y$ takes $np(2p-q)$ operations.

We can summarize these operation counts, keeping only terms $O(npq)$:

(a) bandsolve (no pivoting)	- $np \left(\frac{5}{6}p^2 + \frac{1}{2}pq - \frac{1}{2}q^2 + \frac{3}{2}p + q \right)$
(b) bandsolve (with pivoting)	- $np \left(\frac{5}{6}p^2 + \frac{3}{2}pq + \frac{3}{2}p + \frac{3}{2}q \right)$
(c) blocksolve	- $np \left(\frac{1}{3}p^2 + 2pq - q^2 + 2p \right)$.

It is clear that (b) is always slower than (c), but the interesting comparison is (a) vs (c): (c) is better for $r = q/p$ near zero

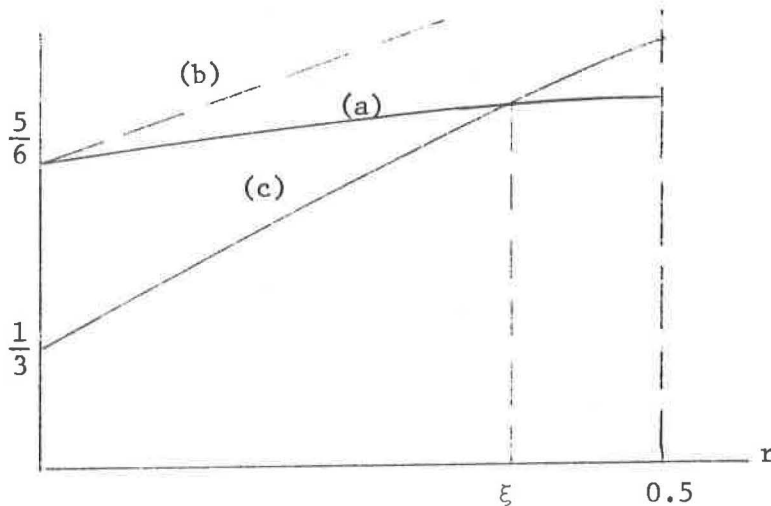
(e.g. $q=1$), but (a) is better for r near $1/2$. (We need only consider $0 < r \leq 1/2$ since for $r > 1/2$ it is more efficient to turn the matrix around.)

To get a graphical comparison, we can drop the terms in npq, np^2 , and divide by np^3 ; then each count is only a function of r :

$$(a) \frac{5}{6} + \frac{1}{2}r - \frac{1}{2}r^2$$

$$(b) \frac{5}{6} + \frac{3}{2}r$$

$$(c) \frac{1}{3} + 2r - r^2$$



Using just these terms, we see that (a) is better for $r < \xi$, (c) better for $r > \xi$, $\xi \approx 0.38$, with the difference at $r = \frac{1}{2}$ amounting to about 12%. Actually the graph is biased in favor of (c) especially for r small, because the term in np^2 has been dropped. For an additional comparison, we give the operation counts from the table for $p=8$:

$\frac{\text{count}}{np} \backslash q$	1	2	3	4
(a)	70	73	76	77
(b)	78	92	105	119
(c)	52	65	76	85

References

1. D.G. Feingold and R.S. Varga, Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem, Pac. J. Math (12)(1962), 1241-1250.
2. A.S. Householder, The Theory of Matrices in Numerical Analysis Blaisdell, Waltham, Mass., 1964.
3. E.A. Isaacson and H.B. Keller, Analysis of Numerical Methods, John Wiley and Sons, New York, 1966.
4. H.B. Keller, Accurate difference methods for linear ordinary differential systems subject to linear constraints, SIAM J. Num. Anal. 6 (1969), 8-30.

5. H.B. Keller, A new difference scheme for parabolic problems. Numerical Solution of Partial Differential Equations II (SYNSPADE 1970), B. Hubbard ed., Academic Press, New York, 1971.
6. R.D. Richtmyer and K.W. Morton, Difference Methods for Initial-Value Problems. Interscience, New York, 1967.
7. B. Wendroff, Theoretical Numerical Analysis. Academic Press, New York, 1966.
8. J.H. Wilkinson, Error analysis of direct methods of matrix inversion. JACM 8 (1961), 281-330.

THE LIBRARY
THE COMPUTING CENTRE
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER 8, B. C.