MATHEMATICS OF COMPUTATION
VOLUME 54, NUMBER 190
APRIL 1990, PAGES 671-700

ITERATIVE METHODS FOR CYCLICALLY REDUCED NON-SELF-ADJOINT LINEAR SYSTEMS

HOWARD C. ELMAN AND GENE H. GOLUB

ABSTRACT. We study iterative methods for solving linear systems of the type arising from two-cyclic discretizations of non-self-adjoint two-dimensional elliptic partial differential equations. A prototype is the convection-diffusion equation. The methods consist of applying one step of cyclic reduction, resulting in a "reduced system" of half the order of the original discrete problem, combined with a reordering and a block iterative technique for solving the reduced system. For constant-coefficient problems, we present analytic bounds on the spectral radii of the iteration matrices in terms of cell Reynolds numbers that show the methods to be rapidly convergent. In addition, we describe numerical experiments that supplement the analysis and that indicate that the methods compare favorably with methods for solving the "unreduced" system.

1. INTRODUCTION

We consider iterative methods for solving the linear systems that arise from finite difference discretizations of non-self-adjoint elliptic problems of the form

$$-\nabla \cdot p\nabla u + q \cdot \nabla u = f \quad \text{on } \Omega, \tag{1.1a}$$

$$ru + su_n = g \quad \text{on } \partial\Omega, \tag{1.1b}$$

where Ω is a smooth domain in $\mathbf{R}^2$. Discretization of (1.1) by finite differences results in a linear system of equations

$$Au = f, \tag{1.2}$$

where u and f now denote vectors in a finite-dimensional space. A is typically nonsymmetric and it is often not diagonally dominant.

Received December 2, 1988; revised May 19, 1989.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 65F10, 65N20; Secondary 15A06.

Key words and phrases. Linear systems, reduced system, iterative methods, convection-diffusion, non-self-adjoint.

The work of the first author was supported by the National Science Foundation under grant DMS-8607478.

The work of the second author was supported by the National Science Foundation under grant DCR-8412314, the Simon Guggenheim Memorial Foundation, and the University of Maryland's Institute for Advanced Computer Studies, whose support is gratefully acknowledged.

For five-point finite difference discretizations, A has Property A [21], i.e., its rows and columns can be symmetrically permuted so that (after appropriate permutation of the entries of u and f) (1.2) has the form

$$\begin{pmatrix} D & C \\ E & F \end{pmatrix} \begin{pmatrix} u^{(r)} \\ u^{(b)} \end{pmatrix} = \begin{pmatrix} f^{(r)} \\ f^{(b)} \end{pmatrix}, \tag{1.3}$$

where D and F are nonsingular diagonal matrices. The system (1.3) corresponds to a *red-black* ordering of the underlying grid. With one step of cyclic reduction, the "red" points $u^{(r)}$ can be decoupled from the "black" points $u^{(b)}$, producing a *reduced system*

$$[F - ED^{-1}C]u^{(b)} = f^{(b)} - ED^{-1}f^{(r)}. \tag{1.4}$$

The coefficient matrix

$$S = F - ED^{-1}C \tag{1.5}$$

is also sparse, so that (1.4) can be solved by some sparse iterative method. For symmetric positive definite systems arising from self-adjoint problems, it is known that iterative schemes such as the Chebyshev and conjugate gradient methods converge more rapidly when applied to (1.4) than when applied to (1.2), see [3, 11, 12]. It has also been observed empirically for a large collection of nonsymmetric problems that preconditioned iterative methods are more effective for solving (1.4) than for solving (1.2) [7, 8].

In this paper, we present a convergence analysis of some block iterative methods for solving (1.4) based on a 1-*line* ordering of the reduced grid. For the full system (1.2), line methods of this type are known to be effective in the self-adjoint case, see e.g. [14, 20, 21], and they have also been applied successfully to non-self-adjoint problems [4, 5]. Our analysis applies to finite difference discretizations of a constant-coefficient version of (1.1) with Dirichlet boundary conditions. We show that the coefficient matrix S is symmetrizable under a wide variety of circumstances, and we use symmetrizability to derive bounds on the convergence rates in terms of cell Reynolds numbers. In addition, we present the results of numerical experiments on nonsymmetrizable and variable-coefficient problems that supplement the analysis. The results suggest that the methods considered are highly effective for computing the numerical solution to (1.1).

We remark that the choice of finite difference discretization affects the accuracy and quality of the discrete solution to (1.1), see [18] and references therein. In this paper, we are concerned with properties of the matrices arising after this choice is made. We consider two difference schemes as examples, based on either centered differences or upwind differences for the first-order terms of (1.1). The analysis of the paper can also be applied to other schemes.

An outline of the paper is as follows. In §2, we illustrate our methodology on a simple one-dimensional example, where the algebra is more transparent than for two-dimensional problems. In §3, we describe the discrete constant-coefficient two-dimensional convection-diffusion equation, and we present a convergence analysis of a block Jacobi method for solving the full system (1.2). This analysis is closely related to that of [4], which applies in a somewhat more general setting. In §4, we present the convergence analysis for the reduced system. We present conditions under which S is symmetrizable, and we derive bounds on the spectral radii of iteration matrices arising from a block Jacobi splitting, where the underlying grid is ordered by diagonals. In §5, we present some numerical experiments that confirm the analysis of the symmetrizable case and demonstrate the effectiveness of the reduced system in other cases. Finally, in §6, we draw conclusions.

2. A one-dimensional example

In this section we demonstrate the use of cyclic reduction for one-dimensional problems. Such problems are not difficult from a computational point of view; we consider them because the algebra is more transparent than for higher dimensions. Consider the constant-coefficient problem

$$-u'' + \sigma u' = f \quad \text{on } (0, 1), \ u(0), u(1) \text{ given}. \tag{2.1}$$

Let (2.1) be discretized by centered finite differences with n interior mesh points:

$$u'' \approx \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2}, \qquad u' \approx \frac{u_{i+1} - u_{i-1}}{2h},$$

where $h = 1/(n+1)$. The result is a linear system of equations $Au = f$ where A is a tridiagonal matrix,

$$A = \text{tri}[-(1+\gamma), 2, -(1-\gamma)] \tag{2.2}$$

with $\gamma = \sigma h/2$. We refer to this quantity as the *cell Reynolds number.* Here $\text{tri}[b_i, a_i, c_i]$ denotes the tridiagonal matrix whose ith row contains the values b_i, a_i, and c_i on its subdiagonal, diagonal, and superdiagonal, respectively. The subdiagonal of the first row and the superdiagonal of the last row are not defined. We omit the subscripts in the case of constant coefficients.

Consider symmetrizing A by a diagonal similarity transformation.

Lemma 1. *For $A = \text{tri}[b_i, a_i, c_i]$, where b_i, a_i, and c_i are real, there exists a real (nonsingular) diagonal matrix Q with $Q^{-1}AQ$ a real symmetric matrix if and only if for each i $(1 \le i \le n)$, one of $b_{i+1}c_i > 0$ or $b_{i+1} = c_i = 0$ holds. The symmetrized matrix is* $\text{tri}[(b_i c_{i-1})^{1/2}, a_i, (b_{i+1}c_i)^{1/2}]$.

Proof. If all $\{b_i\}$, $\{c_i\}$ are nonzero, then the entries of Q are determined by the specification $q_{i+1}^{-1} b_{i+i} q_i = q_i^{-1} c_i q_{i+1}$, i.e., $q_1 \ne 0$ is arbitrary and

$$q_{i+1} = (b_{i+1}/c_i)^{1/2} q_i = [(b_{i+1}c_i)^{1/2}/c_i] q_i, \qquad i \ge 1.$$

If $b_{i+1} = c_i = 0$, then $q_{i+1} \ne 0$ may be arbitrary. □

This result implies that the tridiagonal matrix of (2.2) can be symmetrized when $\gamma^2 < 1$. The symmetrized matrix is

$$\hat{A} = Q^{-1}AQ = \text{tri}[(1-\gamma^2)^{1/2}, 2, (1-\gamma^2)^{1/2}]. \tag{2.3}$$

$\hat{A}$ is diagonally dominant whenever Q is defined.

If the unknowns $\{u_i\}_{i=1}^n$ are ordered with the odd-numbered indices first, then (2.2) has the form (1.3) where D and F have diagonal entries equal to two, and C and E are bidiagonal. After one step of cyclic reduction (and scaling by two), the reduced matrix has the form

$$S = \text{tri}[-(1+\gamma)^2, 2(1+\gamma^2), -(1-\gamma)^2] \tag{2.4}$$

when n is odd. S also has this form when n is even, except that the last diagonal entry is $3+\gamma^2$. Lemma 1 implies that S is symmetrizable *for all* $\gamma \neq 1$, with symmetrized matrix

$$\hat{S} = \text{tri}[\pm(1-\gamma)^2, 2(1+\gamma^2), \pm(1-\gamma^2)]. \tag{2.5}$$

It is straightforward to show that S is diagonally dominant for all γ when n is odd and for $|\gamma| \leq 1$ when n is even. $\hat{S}$ is diagonally dominant and positive definite for all γ.

Consider an analysis of the point Jacobi method for solving linear systems with the coefficient matrices of (2.2)–(2.5). We use the following result, which applies even if $bc \leq 0$.

Lemma 2. *The eigenvalues of the tridiagonal matrix* tri$[b, a, c]$ *of order* m *are* $\{\lambda_j = a + \text{sign}(c)2\sqrt{bc}\cos(j\pi/(m+1)),\ j = 1, \dots, m\}$.

Proof. This can be verified directly. The eigenvector corresponding to λ_j is $v^{(j)}$ where $v_k^{(j)} = (b/c)^{k/2}\sin(jk\pi/(m+1))$, $k = 1, \dots, m$.

Corollary 1. *The spectral radius of the point Jacobi iteration matrix for* (2.2), *and for* (2.3) *when* $|\gamma| \leq 1$, *is* $|(1-\gamma^2)^{1/2}|\cos(\pi h)$. *For odd* n *and* $\gamma \neq 1$, *the spectral radius of the point Jacobi iteration matrix for both* (2.4) *and* (2.5) *is* $|(1-\gamma^2)/(1+\gamma^2)|\cos(2\pi h)$.

Proof. The Jacobi matrix for (2.2) is tri$[-(1+\gamma)/2, 0, -(1-\gamma)/2]$, and its eigenvalues are determined as in Lemma 2. The analysis for the other three matrices is identical. □

Thus, one step of cyclic reduction produces a matrix that has good numerical properties. In contrast to the full system, the reduced system is symmetrizable for all $\gamma \neq 1$. For the full system, the point Jacobi iteration is convergent only for $\gamma \leq 2$, whereas for the reduced system it is convergent for $\gamma \neq 1$; in addition, the spectral radius is always smaller for the reduced system. We remark that by expanding the reduced operator

$$-(1+\gamma)^2u_{i-2} + 2(1+\gamma^2)u_i - (1-\gamma)^2u_{i+2}$$

in a Taylor series centered at u_i [19], we find that this difference scheme can be viewed as second-order approximation to the differential operator

$$-\left(1+\frac{\sigma^2h^2}{4}\right)u''+\sigma u'.$$

A heuristic explanation for the good algebraic properties of the reduced matrix is that it corresponds to a perturbation of (2.1) in which "artificial viscosity" is added (see [18]).

Many of the observations made above carry over to the variable-coefficient case, e.g. where the differential operator is $-u''+\sigma(x)u'$. The coefficient matrix is then

$$A=\mathrm{tri}[-(1+\gamma_i),\, 2,\, -(1-\gamma_i)],$$

where $\gamma_i=\sigma(x_i)h/2$. The symmetrized form

$$\hat{A}=\mathrm{tri}[((1+\gamma_i)(1-\gamma_{i-1}))^{1/2},\, 2,\, ((1+\gamma_{i+1})(1-\gamma_i))^{1/2}]$$

is well defined if $|\gamma_i|<1$ for all i. The reduced matrix is (for odd n)

$$S=\mathrm{tri}[-(1+\gamma_{2i})(1+\gamma_{2i-1}),\, 4-(1+\gamma_{2i})(1-\gamma_{2i-1})-(1-\gamma_{2i})(1+\gamma_{2i+1}),\\ -(1-\gamma_{2i})(1-\gamma_{2i+1})],$$

where $1\le i\le \lfloor n/2\rfloor$. If the conditions

$$s_{ii}>0,\quad (1+\gamma_{2i})(1+\gamma_{2i-1})>0,\quad (1-\gamma_{2i})(1-\gamma_{2i+1})>0$$

hold, then S is diagonally dominant. If, in addition,

$$(1-\gamma_{2i-2})(1+\gamma_{2i})(1-\gamma_{2i-1}^2)>0,$$

then S is symmetrizable by a real diagonal similarity transformation. These conditions all hold if $|\gamma_i|<1$ for all i, and they also hold for large $\{\gamma_i\}$ whenever $\sigma(x)$ does not have large derivatives in regions where it changes sign. Both sets of conditions are trivially true for constant σ.

Finally, returning to the constant-coefficient case, note that the equation of (2.1) can be written in self-adjoint form

$$-(e^{\sigma x}u')'=e^{\sigma x}f.$$

Consider the symmetrizing matrices Q discussed above. If the first entry satisfies $q_1=1$, then $q_i=[(1+\gamma)/(1-\gamma)]^{i-1}$ for the full system, and $q_i=[(1+\gamma)/(1-\gamma)]^{2(i-1)}$ for the reduced system. In either case, the entries of Q are very large for many values of γ. (For example, for $0<\gamma<1$, the limiting value as $n\to\infty$ of the last entry of Q is e^{σ}, for both systems.) In this sense, the symmetrizing operator behaves like the integrating factor $e^{\sigma x}$. For large σ, both are difficult to implement in floating-point arithmetic.

3. The two-dimensional convection-diffusion equation

Consider the constant-coefficient convection-diffusion equation

$$-\Delta u+\sigma u_x+\tau u_y=f \tag{3.1}$$

on the unit square $\Omega = (0, 1) \times (0, 1)$, with Dirichlet boundary conditions $u = g$ on $\partial\Omega$. We discretize (3.1) on a uniform $n \times n$ grid, using standard second-order differences [20, 21]

$$\Delta u \approx \frac{u_{i+1,j} - 2u_{ij} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2u_{ij} + u_{i,j-1}}{h^2}$$

for the Laplacian, where $h = 1/(n+1)$. We examine two choices of finite difference schemes for the first derivative terms:

1. centered differences: $u_x \approx \dfrac{u_{i+1,j} - u_{i-1,j}}{2h}$, $u_y \approx \dfrac{u_{i,j+1} - u_{i,j-1}}{2h}$,
2. upwind differences: $u_x \approx \dfrac{u_{ij} - u_{i-1,j}}{h}$ $u_y \approx \dfrac{u_{ij} - u_{i,j-1}}{h}$,

where the latter is applicable when $\sigma, \tau \geq 0$.

Suppose the grid points are ordered using the rowwise natural ordering, i.e., the vector u is ordered lexicographically as $(u_{1,1}, u_{2,1}, \dots, u_{n,n})^T$. Then, for both discretizations of the first derivative terms, the coefficient matrix has the form

$$A = \mathrm{tri}[A_{j,j-1}, A_{jj}, A_{j,j+1}], \tag{3.2}$$

where

$$A_{j,j-1} = bI, \qquad A_{jj} = \mathrm{tri}[c, a, d], \qquad A_{j,j+1} = eI, \tag{3.3}$$

I is the identity matrix, and all blocks are of order n. After scaling by h^2, the matrix entries are given by

$$\begin{aligned} a = 4, \qquad b = -(1+\delta), \qquad c = -(1+\gamma), \\ d = -(1-\gamma), \qquad e = -(1-\delta), \end{aligned} \tag{3.4}$$

for the centered difference scheme, where the cell Reynolds numbers are $\gamma = \sigma h/2$ and $\delta = \tau h/2$; and

$$\begin{aligned} a = 4 + 2(\gamma+\delta), \qquad b = -(1+2\delta), \qquad c = -(1+2\gamma), \\ d = -1, \qquad e = -1, \end{aligned} \tag{3.5}$$

for the upwind scheme.

First, assume that $cd > 0$. This is true for the centered difference scheme ($cd = 1 - \gamma^2$) when $|\gamma| < 1$, and it always holds for the upwind scheme ($cd = 1 + 2\gamma$). Consider the block Jacobi splitting

$$A = D - C, \tag{3.6}$$

where D is the block diagonal matrix $\mathrm{diag}(A_{11}, A_{22}, \dots, A_{nn})$. For our analysis, it will be useful to define several auxiliary matrices. In particular, by Lemma 1, each A_{jj} can be symmetrized by a real diagonal similarity transformation Q_j. The choice of the first entry $q_1^{(j)}$ of each Q_j is arbitrary; for the moment we fix this choice to be one. The symmetrized matrix is

$$\hat{A}_{jj} = Q_j^{-1} A_{jj} Q_j = \mathrm{tri}[\sqrt{cd}, a, \sqrt{cd}].$$

By Lemma 2, the eigenvalues of $\widehat{A}_{jj}$ are $\{\lambda_k \equiv a + 2\sqrt{cd}\cos(k\pi h) | 1 \leq k \leq n\}$. Let V_j be an orthonormal matrix whose columns are the corresponding eigenvectors of $\widehat{A}_{jj}$. Let

$$Q = \operatorname{diag}(Q_1, Q_2, \dots, Q_n), \qquad V = \operatorname{diag}(V_1, V_2, \dots, V_n).$$

Finally, let P denote the permutation matrix that transforms the rowwise natural ordering into the columnwise natural ordering, i.e., $P^T AP$ has the form of (3.2)–(3.3), except that the roles of b and c are interchanged and the roles of d and e are interchanged.

Consider the similarity transformation $\widetilde{A} = (QVP)^{-1}A(QVP)$. This transformation first symmetrizes the block diagonal of A, then diagonalizes the resulting interior tridiagonal matrix, and then reorders to produce a block diagonal matrix with tridiagonal blocks. The splitting of $\widetilde{A}$ analogous to (3.6) is $\widetilde{D} - \widetilde{C}$ where $\widetilde{D} = (QVP)^{-1}D(QVP)$ and $\widetilde{C} = (QVP)^{-1}C(QVP)$. But $\widetilde{D}^{-1}\widetilde{C} = (QVP)^{-1}D^{-1}C(QVP)$, so that the eigenvalues of $D^{-1}C$ are the same as those of $\widetilde{D}^{-1}\widetilde{C}$. Moreover, $\widetilde{D}_j$ is a diagonal matrix, all of whose nonzero entries are λ_j, and the choice $q_1^{(j)} = 1$ implies that $\widetilde{C}$ is the block diagonal matrix whose jth diagonal block is $\operatorname{tri}[b, 0, e]$ for all j. Hence, $\widetilde{D}^{-1}\widetilde{C}$ is a block diagonal matrix whose jth diagonal block is

$$[\widetilde{D}^{-1}\widetilde{C}]_j = \operatorname{tri}[b/\lambda_j, 0, e/\lambda_j].$$

Applying Lemma 2 again, we have that the eigenvalues of $\widetilde{D}^{-1}\widetilde{C}$ are

$$\frac{2\sqrt{be}\cos(k\pi h)}{a + 2\sqrt{cd}\cos(j\pi h)}.$$

The maximum such value occurs when $j = n$, $k = 1$.

Note that this analysis imposes no condition on b and e. If $be > 0$ ($|\delta| < 1$ for centered differences and always for upwind differences), then an identical analysis could be applied to the columnwise ordered version of A. Hence, we have the following result.

Theorem 1. *If $cd > 0$, then the spectral radius of the block Jacobi iteration matrix for the rowwise ordered full system is*

$$\frac{2\sqrt{be}\cos(\pi h)}{a - 2\sqrt{cd}\cos(\pi h)}.$$

If $be > 0$, then the spectral radius of the block Jacobi iteration matrix for the columnwise ordered full system is

$$\frac{2\sqrt{cd}\cos(\pi h)}{a - 2\sqrt{be}\cos(\pi h)}.$$

The bounds for the difference schemes under consideration are derived by substituting the values of $a - e$ from (3.4) and (3.5) into the results of Theorem 1. These bounds also follow from the more general analysis given in [4].

Corollary 2. *For centered differences, if* $|\gamma| < 1$, *then the spectral radius of the block Jacobi iteration matrix for the rowwise ordered full system is*

$$\frac{\sqrt{|1-\delta^2|}\cos(\pi h)}{2-\sqrt{1-\gamma^2}\cos(\pi h)}.$$

If $|\delta| < 1$, *then the spectral radius of the block Jacobi iteration matrix for the columnwise ordered full system is*

$$\frac{\sqrt{|1-\gamma^2|}\cos(\pi h)}{2-\sqrt{1-\delta^2}\cos(\pi h)}.$$

For upwind differences, the spectral radii are

$$\frac{\sqrt{1+2\delta}\cos(\pi h)}{2+(\gamma+\delta)-\sqrt{1+2\gamma}\cos(\pi h)} \quad \textit{and} \quad \frac{\sqrt{1+2\gamma}\cos(\pi h)}{2+(\gamma+\delta)-\sqrt{1+2\delta}\cos(\pi h)}$$

for the rowwise and columnwise orderings, respectively.

Comparison of these asymptotic bounds as $h \to 0$ shows that for centered differences, when both $|\gamma| < 1$ and $|\delta| < 1$, the rowwise bound is smaller if $|\gamma| < |\delta|$ and the columnwise bound is smaller if $|\delta| < |\gamma|$. For upwind differences, the rowwise bound is smaller if $\delta < \gamma$ and the columnwise bound is smaller if $\gamma < \delta$.

Finally, recall that the choice $q_1^{(j)} = 1$ in the discussion above was arbitrary. In particular, if both $cd > 0$ and $be > 0$, then for $q_1^{(j+1)} = (b/e)^{1/2} q_1^{(j)}$, $Q^{-1}AQ$ is symmetric. The analysis of the Jacobi splittings is unaffected. Hence, we have the following result.

Theorem 2. *If both* $|\gamma| < 1$ *and* $|\delta| < 1$, *then the coefficient matrix for the centered difference scheme is symmetrizable by a real diagonal similarity transformation. For all* $\gamma > 0$ *and* $\delta > 0$, *the coefficient matrix for the upwind scheme is symmetrizable.*

We remark that these symmetrizing operations have been discussed in [6].

4. The two-dimensional reduced system

In this section, we discuss the construction of the two-dimensional reduced matrix S of (1.5), and we present an analysis, based on symmetrizing the reduced matrix, of a block Jacobi iteration for solving the reduced system. We remark that only the analysis depends on symmetrizability; the solution methods considered here do not require the construction of a symmetrizing operator.

4.1. Construction of the reduced matrix. The nonzero structure of S can be determined from the connection between the graph of a matrix and Gaussian elimination. It is well known that applying one step of Gaussian elimination to a linear system introduces edges into the corresponding graph of the matrix

[15]. If u is a node corresponding to an eliminated unknown, and v and w are nodes such that (u, v) and (u, w) are edges in the graph prior to eliminating u, then the edge (v, w) is introduced after elimination. In the present setting, the original matrix A is a five-point operator whose graph is a rectangular grid. The left side of Figure 4.1 shows the computational molecule for A, and the center of the figure shows a portion of the graph of A relevant to the construction of the reduced system. For the reduction, the points numbered 3, 6, 8, and 11 (the "red points") are eliminated, producing the computational molecule on the right. Thus, the reduced matrix is a skewed nine-point operator.

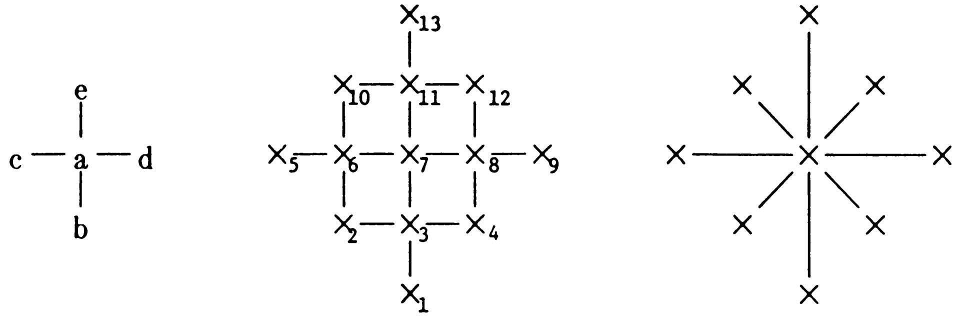

FIGURE 4.1. The computational molecule of the full system, and construction of the computational molecule of the reduced system.

To see the entries of the reduced matrix, consider the submatrix of A consisting of the rows for points 3, 6, 7, 8, and 11, and the columns for all the points of the graph in the center of Figure 4.1.

$$
\begin{array}{ll|ccccccccccccc}
\text{Column Index:} & & 3 & 6 & 8 & 11 & 7 & 1 & 2 & 4 & 5 & 9 & 10 & 12 & 13 \\
\hline
 & 3 & a & & & & e & b & c & d & & & & & \\
\text{Row} & 6 & & a & & & d & & b & & c & & e & & \\
\text{Index:} & 8 & & & a & & c & & & b & & d & & e & \\
 & 11 & & & & a & b & & & & & & c & d & e \\
 & 7 & b & c & d & e & a & & & & & & & &
\end{array}
$$

Eliminating points 3, 6, 8, and 11 is equivalent to decoupling the first four rows of this matrix by Gaussian elimination. This modifies and produces fill-in in the last row. The computations performed for the elimination are shown in the following table. The new entries of the last row are obtained by summing the columns.

Column:	7	1	2	4	5	9	10	12	13
	a								
Eliminate 3:	$-ba^{-1}e$	$-ba^{-1}b$	$-ba^{-1}c$	$-ba^{-1}d$					
Eliminate 6:	$-ca^{-1}d$		$-ca^{-1}b$		$ca^{-1}c$		$-ca^{-1}e$		
Eliminate 8:	$-da^{-1}c$			$-da^{-1}b$		$-da^{-1}d$		$-da^{-1}e$	
Eliminate 11:	$-ea^{-1}b$						$-ea^{-1}c$	$-ea^{-1}d$	$-ea^{-1}e$

Thus, the typical diagonal value in the reduced matrix is

$$a - 2ba^{-1}e - 2ca^{-1}d, \tag{4.1}$$

which occurs at all interior grid points. For grid points next to the boundary (see Figure 4.3 below), some elimination steps are not required. For example, for a point next to the right boundary, it is not necessary to eliminate d. The diagonal values for mesh points next to the boundary are

$$\begin{aligned} &a - 2ba^{-1}e - ca^{-1}d && \text{for points with one horizontal} \\ & && \text{and two vertical neighbors,} \\ &a - ba^{-1}e - 2ca^{-1}d && \text{for points with one vertical} \\ & && \text{and two horizontal neighbors,} \\ &a - ba^{-1}e - ca^{-1}d && \text{for points with just two neighbors.} \end{aligned} \tag{4.2}$$

After scaling by a, the computational molecule at an interior point for the reduced system is shown in Figure 4.2.

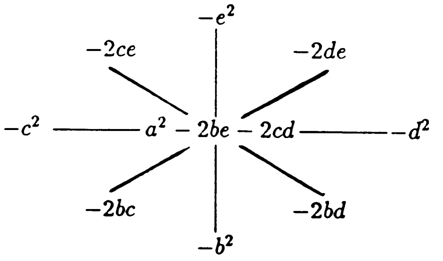

FIGURE 4.2. The computational molecule for the reduced system.

Further steps of cyclic reduction do not lead to sparse reduced matrices, so we restrict our attention to one reduction step.

4.2. Symmetrizing the reduced matrix and the block Jacobi splitting. Suppose the reduced grid is ordered by diagonal lines oriented in the NW–SE direction. An example of such an ordering derived from a 6×6 grid is shown in Figure 4.3.

$\times_7$	·	$\times_3$	·	$\times_1$	·
·	$\times_8$	·	$\times_4$	·	$\times_2$
$\times_{13}$	·	$\times_9$	·	$\times_5$	·
·	$\times_{14}$	·	$\times_{10}$	·	$\times_6$
$\times_{17}$	·	$\times_{15}$	·	$\times_{11}$	·
·	$\times_{18}$	·	$\times_{16}$	·	$\times_{12}$

FIGURE 4.3. The reduced grid derived from a 6×6 grid, with ordering by diagonals.

The reduced matrix S then has block tridiagonal form

$$\begin{pmatrix} S_{11} & S_{12} & & & \\ S_{21} & S_{22} & S_{23} & & \\ & & \ddots & & \\ & & & & S_{l-1,l} \\ & & & S_{l,l-1} & S_{ll} \end{pmatrix},$$

where $l = n - 1$ is the number of diagonal lines. The diagonal matrices $\{S_{jj}\}$ are tridiagonal,

$$S_{jj} = \text{tri}[-2ce\,,\, *\,,\, -2bd]\,, \tag{4.3}$$

where "$*$" is defined as in (4.1) and (4.2) (scaled by a). The subdiagonal blocks $\{S_{j,j-1}\}$ have nonzero structure

$$-\begin{pmatrix} d^2 & & & \\ 2de & d^2 & & \\ e^2 & 2de & \cdot & \\ & e^2 & \cdot & d^2 \\ & & \cdot & 2de \\ & & & e^2 \end{pmatrix}, \quad -\begin{pmatrix} 2de & d^2 & & \\ e^2 & 2de & d^2 & \\ & & \ddots & \\ & & & d^2 \\ & & e^2 & 2de \end{pmatrix}, \tag{4.4}$$

$$-\begin{pmatrix} e^2 & 2de & d^2 & & \\ & e^2 & 2de & d^2 & \\ & & \cdot & \cdot & \cdot \\ & & e^2 & 2de & d^2 \end{pmatrix},$$

for $2 \le j < l/2 + 1$, $j = l/2 + 1$ (l even), and $l/2 + 1 < j$, respectively. The corresponding superdiagonals $\{S_{j-1,j}\}$ are

$$-\begin{pmatrix} c^2 & 2bc & b^2 & & \\ & c^2 & 2bc & b^2 & \\ & & \cdot & \cdot & \cdot \\ & & c^2 & 2bc & b^2 \end{pmatrix}, \quad -\begin{pmatrix} 2bc & b^2 & & \\ c^2 & 2bc & b^2 & \\ & & \ddots & \\ & & & b^2 \\ & & c^2 & 2bc \end{pmatrix}, \tag{4.5}$$

$$-\begin{pmatrix} b^2 & & & \\ 2bc & b^2 & & \\ c^2 & 2bc & \cdot & \\ & c^2 & \cdot & b^2 \\ & & \cdot & 2bc \\ & & & c^2 \end{pmatrix}.$$

The following result gives circumstances under which S is symmetrizable. This result also follows from the analysis of [16].

Theorem 3. *The reduced matrix S can be symmetrized with a real diagonal similarity transformation if and only if the product $bcde$ is positive.*

Proof. We seek a matrix $Q = \operatorname{diag}(Q_1, \dots, Q_l)$, where Q_j is a real diagonal matrix of the same order as S_{jj}, such that $Q^{-1}SQ$ is symmetric. Let $Q_j = \operatorname{diag}(q_1^{(j)}, \dots, q_{r_j}^{(j)})$. First consider the diagonal block (4.3): $Q_j^{-1}S_{jj}Q_j$ is symmetric if and only if

$$-\frac{q_i^{(j)}}{q_{i-1}^{(j)}}bd = -\frac{q_{i-1}^{(j)}}{q_i^{(j)}}ce, \qquad 2 \le j \le r_j, \tag{4.6}$$

where $q_1^{(j)}$ may be arbitrary. Thus, the diagonal blocks can be symmetrized provided

$$q_i^{(j)} = \left(\frac{ce}{bd}\right)^{1/2} q_{i-1}^{(j)}, \tag{4.7}$$

and this recurrence is well defined if and only if $ce/(bd) = bcde/(bd)^2$ is positive. The (equal) quantities (4.6) are the $(i, i-1)$ and $(i-1, i)$ entries of the jth diagonal block of the symmetrized matrix.

For the off-diagonal blocks, we require

$$Q_j^{-1}S_{j,j-1}Q_{j-1} = (Q_{j-1}^{-1}S_{j-1,j}Q_j)^T. \tag{4.8}$$

There are three cases, corresponding to the three sets of indices $2 \le j < l/2+1$, $j = l/2+1$ (l even), and $l/2+1 < j$ (see (4.4)–(4.5)). When $2 \le j < l/2+1$, relation (4.8) holds if and only if the following scalar relations hold:

$$\frac{q_i^{(j)}}{q_i^{(j-1)}}c^2 = \frac{q_i^{(j-1)}}{q_i^{(j)}}d^2, \quad \text{or} \quad q_i^{(j)} = \left(\frac{d^2}{c^2}\right)^{1/2} q_i^{(j-1)}; \tag{4.9}$$

$$\frac{q_{i+1}^{(j)}}{q_i^{(j-1)}}bc = \frac{q_i^{(j-1)}}{q_{i+1}^{(j)}}de, \quad \text{or} \quad q_{i+1}^{(j)} = \left(\frac{de}{bc}\right)^{1/2} q_i^{(j-1)}; \tag{4.10}$$

$$\frac{q_{i+2}^{(j)}}{q_i^{(j-1)}}b^2 = \frac{q_i^{(j-1)}}{q_{i+2}^{(j)}}e^2, \quad \text{or} \quad q_{i+2}^{(j)} = \left(\frac{e^2}{b^2}\right)^{1/2} q_i^{(j-1)}. \tag{4.11}$$

Since the $\{q_1^{(j)}\}$ are arbitrary, (4.9) can be used to define $\{q_1^{(j)}\}$, $2 \le j < l/2+1$ (where $q_1^{(1)}$ is arbitrary). Once this choice is made, however, (4.7) completely determines $\{Q_j\}$. Thus, it is necessary to show that (4.9)–(4.11) are consistent with (4.7). But $\{q_i^{(j)}\}$ and $\{q_i^{(j-1)}\}$ both satisfy (4.7), so that (4.9) is consistent. Moreover, applying (4.7) and (4.9) gives

$$\left(q_{i+1}^{(j)}\right)^2 = \frac{ce}{bd}\left(q_i^{(j)}\right)^2 = \frac{ce}{bd}\frac{d^2}{c^2}\left(q_i^{(j-1)}\right)^2 = \frac{de}{bc}\left(q_i^{(j-1)}\right)^2,$$

and applying (4.7) twice, followed by (4.9), gives

$$\left(q_{i+2}^{(j)}\right)^2 = \left(\frac{ce}{bd}\right)^2\left(q_i^{(j)}\right)^2 = \left(\frac{ce}{bd}\right)^2\frac{d^2}{c^2}\left(q_i^{(j-1)}\right)^2 = \frac{e^2}{b^2}\left(q_i^{(j-1)}\right)^2.$$

That is, (4.10) and (4.11) follow directly from (4.7) and (4.9). The square root in (4.10) is well defined provided $de/(bc) = bcde/(bc)^2$ is positive.

For the other two cases, the analogues of (4.9)–(4.11) are

$$q_i^{(j)} = \left(\frac{de}{bc}\right)^{1/2} q_i^{(j-1)}, \quad q_{i+1}^{(j)} = \left(\frac{e^2}{b^2}\right)^{1/2} q_i^{(j-1)}, \quad q_i^{(j)} = \left(\frac{d^2}{c^2}\right)^{1/2} q_{i+1}^{(j-1)},$$

for l even and $j = l/2 + 1$, and

$$q_i^{(j)} = \left(\frac{e^2}{b^2}\right)^{1/2} q_i^{(j-1)}, \quad q_i^{(j)} = \left(\frac{de}{bc}\right)^{1/2} q_{i+1}^{(j-1)}, \quad q_i^{(j)} = \left(\frac{d^2}{c^2}\right)^{1/2} q_{i+2}^{(j-1)},$$

for $l/2 + 1 < j$. Here $q_1^{(j)}$ is defined using the first expression of these relations. Proofs that these are consistent with (4.7) are essentially identical to the argument above. □

Let $D = \operatorname{diag}(S_{11}, \dots, S_{ll})$ denote the block diagonal of S, and let $S = D - C$ denote the block Jacobi splitting of S. Let $\widehat{S}$ denote the symmetrized matrix $Q^{-1}SQ$ (when it exists), and let $\widehat{S} = \widehat{D} - \widehat{C}$ denote the block Jacobi splitting. Here $\widehat{D} = Q^{-1}DQ$ and $\widehat{C} = Q^{-1}CQ$. Also, note that S and $\widehat{S}$ are irreducible.

Corollary 3. *If both $be > 0$ and $cd > 0$, then S is symmetrizable by a real diagonal similarity transformation Q. If the diagonal entries of S are positive, then S is an irreducibly diagonally dominant M-matrix provided*

$$a^2 \geq (|b| + |c| + |d| + |e|)^2, \tag{4.12}$$

and $\widehat{S}$ is an irreducibly diagonally dominant M-matrix provided

$$a^2 \geq 4(\sqrt{be} + \sqrt{cd})^2.$$

Proof. The existence of the symmetrizing matrix Q follows immediately from Theorem 3. The off-diagonal entries of S are negative, and the corresponding off-diagonal entries of $\widehat{S}$ are negative if the positive square root is used in the recurrences defining Q. Diagonal dominance is established by direct computation. For rows of S corresponding to interior mesh points, diagonal dominance holds if and only if

$$a^2 - 2be - 2cd \geq b^2 + c^2 + d^2 + e^2 + 2(|bc| + |bd| + |ce| + |de|),$$

which is equivalent to (4.12). For mesh points next to the boundary, the diagonal values are greater than those for the interior points (see (4.2)), so that strict diagonal dominance applies. The argument for diagonal dominance in $\widehat{S}$ is the same. That S and $\widehat{S}$ are M-matrices follows from Corollary 1, p. 85, of Varga [20]. □

Let the nonzero off-diagonal entries of S and $\widehat{S}$ be identified as the north, south, east, and west, and northeast, northwest, southeast, and southwest values, according to their location in the computational molecule. (See Figure 4.3.) D and $\widehat{D}$ consist of the center, northwest, and southeast entries.

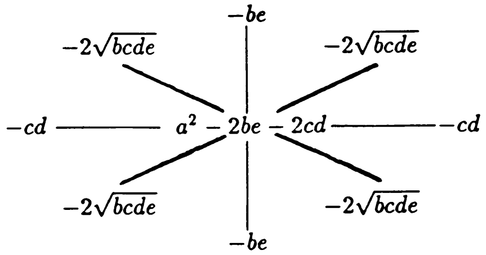

FIGURE 4.4. A computational molecule for the symmetrized reduced system.

Corollary 4. *If both* $be < 0$ *and* $cd < 0$, *then* S *is symmetrizable by a real diagonal similarity transformation* Q. *Depending on the choice of* Q, *any of the following distributions of signs can occur in* $\widehat{S}$:

(a) *the northwest and southeast values are positive and all other off-diagonal values are negative*;

(b) *all nonzero off-diagonal entries of* $\widehat{S}$ *are positive*;

(c) *the northwest, southeast, northeast, and southwest values are negative, and the north, south, east, and west values are positive*;

(d) *the northwest, southeast, north, south, east, and west values are negative, and the northeast and southwest values are positive.*

In cases (c) *and* (d), $\widehat{D}$ *is a diagonally dominant* M*-matrix.*

Proof. The existence of the symmetrizing matrix Q follows immediately from Theorem 3. The signs of the square roots defining the entries of Q, which are arbitrary, determine the distributions of signs in $\widehat{S}$. Without loss of generality, assume that $b < 0$ and $c < 0$, so that the hypotheses imply $d > 0$ and $e > 0$. The arguments for other combinations of signs are identical. If all entries of Q have the same sign (i.e., the positive square root is used throughout), then $\widehat{S}$ is as in (a). If all entries of Q within any of its blocks have the same sign (i.e., the positive square root is used in (4.7)), but neighboring blocks have the opposite sign (e.g., the negative square root is used in (4.9) when this expression defines $q_1^{(j)}$), then the distribution of signs in $\widehat{S}$ is as in (b). On the other hand, if the signs within the blocks of Q alternate, then the northwest and southeast entries (the off-diagonal entries of $\widehat{D}$) are negative, and the signs in the off-diagonal blocks of $\widehat{S}$ vary by diagonal. Distributions (c) and (d) are specified by choosing appropriate signs for the first entries of each block; the choice of sign does not obey a simple fixed rule, but instead varies with the index j of the block Q_j. In the latter two cases, each block of $\widehat{D}$ is irreducibly diagonally dominant with nonpositive off-diagonal entries (see Figure 4.4), so that $\widehat{D}$ is an M-matrix. □

Figure 4.4 shows a computational molecule for the symmetrized reduced system. The figure is valid for both Corollaries 3 and 4. When b, c, d, and e are as in Corollary 3, all off-diagonal entries are negative. When b, c, d, and e are as in Corollary 4, the signs correspond to case (c). For the choices of difference schemes that we are considering, Corollary 3 applies to the centered difference scheme for small $|\gamma|$ and $|\delta|$ (less than one), and to the upwind scheme. Corollary 4 applies to the centered difference scheme for large $|\gamma|$ and $|\delta|$. The following result summarizes the analysis above for the two difference schemes.

Corollary 5. *If A is constructed using centered differences, then S is symmetrizable via a real diagonal matrix Q if and only if either $|\gamma| < 1$ and $|\delta| < 1$ both hold, or $|\gamma| > 1$ and $|\delta| > 1$ both hold. If $|\gamma| < 1$ and $|\delta| < 1$, then S is an irreducibly diagonally dominant M-matrix and Q can be chosen so that $\widehat{S}$ is an irreducibly diagonally dominant M-matrix. If $|\gamma| > 1$ and $|\delta| > 1$, then Q can be chosen so that $\widehat{D}$ is a diagonally dominant M-matrix. If A is constructed using upwind differences, then S is symmetrizable for all $\gamma \geq 0$ and $\delta \geq 0$, and S and (for appropriately chosen Q) $\widehat{S}$ are irreducibly diagonally dominant M-matrices.*

Proof. For centered differences with $|\gamma| < 1$ and $|\delta| < 1$, and for upwind differences, the assertions follow from Corollary 3. Diagonal dominance follows from direct computation. For centered differences with $|\gamma| > 1$ and $|\delta| > 1$, the result corresponds to case (c) or (d) of Corollary 4. □

4.3. Bounds for solving the convection-diffusion equation. We now derive bounds for the spectral radius of the iteration matrix $B = D^{-1}C$ based on the block Jacobi splitting of S, in the case where S is symmetrizable. Note that

$$B = Q\widehat{D}^{-1}\widehat{C}Q^{-1}, \tag{4.13}$$

i.e., B is similar to $\widehat{B} = \widehat{D}^{-1}\widehat{C}$. Hence, we can restrict our attention to $\widehat{B}$. The analysis is essentially based on the result

$$\rho(\widehat{D}^{-1}\widehat{C}) \leq \|\widehat{D}^{-1}\|_2\|\widehat{C}\|_2 = \frac{\rho(\widehat{C})}{\lambda_{\min}(\widehat{D})}, \tag{4.14}$$

where the equality follows from the symmetry of $\widehat{D}$ and $\widehat{C}$. Explicit bounds are obtained by replacing $\widehat{D}$ by a matrix with constant diagonal values (cf. (4.2)). There are two cases, corresponding to Corollary 3 ($be > 0$ and $cd > 0$) and Corollary 4 ($be < 0$ and $cd < 0$). We assume for the second case that Q is chosen so that $\widehat{D}$ is an M-matrix (e.g. with the sign distribution (c)). Hence, in both cases, $\widehat{D}$ is symmetric positive definite and can be factored symmetrically as $\widehat{D} = LL^T$. Consequently,

$$L^T\widehat{D}^{-1}\widehat{C}L^{-T} = L^{-1}\widehat{C}L^{-T}. \tag{4.15}$$

That is, $\widehat{B}$, and therefore B, are similar to a symmetric matrix, and their eigenvalues are real.

Suppose that $be > 0$ and $cd > 0$. The M-matrix $\widehat{D}$ has block diagonal form $\operatorname{diag}(T_1, \ldots, T_l)$, where each block T_j is a tridiagonal matrix. Therefore,

$\sigma(\widehat{D}) = \bigcup_j \sigma(T_j)$, and

$$\lambda_{\min}(\widehat{D}) = \min_j \lambda_{\min}(T_j). \tag{4.16}$$

Let T denote one of the tridiagonal blocks T_j of $\widehat{D}$, of order r. T has the form $\widehat{T}+P$, where $\widehat{T} = \operatorname{tri}[-\hat{b}, \hat{a}, -\hat{b}]$ and $P = \operatorname{diag}(p_1, 0, \dots, 0, p_r)$. Here, $\hat{a} = a^2 - 2be - 2cd$, $\hat{b} = 2\sqrt{bcde}$. The perturbations p_1 and p_r each have the form $\alpha be + \beta cd$, where α and β are zero or one, and at least one of α, β is nonzero, see (4.1), (4.2). Hence, the perturbations are positive. Consequently,

$$\begin{aligned} \lambda_{\min}(T) &= \min_{v\neq 0}(v, Tv) = \min_{v\neq 0}[(v, \widehat{T}v) + (v, Pv)] \\ &\geq \min_{v\neq 0}(v, \widehat{T}v) = \lambda_{\min}(\widehat{T}). \end{aligned} \tag{4.17}$$

By Lemma 2, $\sigma(\widehat{T}) = \{\lambda_k \equiv \hat{a} - 2\hat{b}\cos(k\pi/(r+1))\}$. The minimum value of λ_k is

$$\lambda_{\min}(\widehat{T}) = \hat{a} - 2\hat{b}\cos\theta_1. \tag{4.18}$$

The smallest such minimum (over all choices of $T = T_j$ from $\widehat{D}$) occurs with $r = n$. Thus, (4.16), (4.17), and (4.18) imply

$$\lambda_{\min}(\widehat{D}) \geq \hat{a} - 2\hat{b}\cos(\pi h). \tag{4.19}$$

In terms of the entries of A, this expression is

$$\hat{a} - 2\hat{b} + 2\hat{b}(1 - \cos(\pi h)) = a^2 - 2(\sqrt{be} + \sqrt{cd})^2 + 4\sqrt{bcde}(1 - \cos(\pi h)).$$

The spectral radius of $\widehat{C}$ is bounded by Gerschgorin's theorem [20]:

$$\rho(\widehat{C}) \leq 4\sqrt{bcde} + 2be + 2cd = 2(\sqrt{be} + \sqrt{cd})^2.$$

Hence, from (4.14) we have the following result:

Theorem 4. *If* $be > 0$ *and* $cd > 0$, *then the spectral radius of the block Jacobi iteration matrix for the reduced system satisfies*

$$\rho(B) \leq \frac{2(\sqrt{be} + \sqrt{cd})^2}{a^2 - 2(\sqrt{be} + \sqrt{cd})^2 + 4\sqrt{bcde}(1 - \cos(\pi h))}.$$

When $be < 0$ and $cd < 0$, the perturbations p_1 and p_r are negative. Consequently, the inequality of (4.17) is not valid and (4.14) cannot be used directly. For an alternative approach, let $\widehat{C} = \widehat{C}^{(a)} + \widehat{C}^{(b)}$, where $\widehat{C}^{(a)}$ is the part of $\widehat{C}$ corresponding to the northeast and southwest neighbors in the computational molecule, and $\widehat{C}^{(b)}$ is the part of $\widehat{C}$ corresponding to the north, south, east, and west neighbors. Assume that Q is chosen so that case (c) of Corollary 4 holds. Then $\widehat{C}^{(a)} \geq 0$ and $\widehat{C}^{(b)} \leq 0$. (See Figure 4.4.) This implies that the splittings

$$\widehat{S}^{(a)} = \widehat{D} - \widehat{C}^{(a)}, \qquad \widehat{S}^{(b)} = \widehat{D} - [-\widehat{C}^{(b)}] \tag{4.20}$$

are regular splittings [20]. Moreover, $\widehat{S}^{(b)}$ is an irreducibly diagonally dominant M-matrix, and $\widehat{S}^{(a)}$ is an irreducibly diagonally dominant M-matrix provided

$$a^2/2 + (\sqrt{-cd} - \sqrt{-be})^2 - 2\sqrt{bcde} \geq 0 . \tag{4.21}$$

In the following discussion, we assume this inequality holds, so that $\widehat{S}^{(a)-1} > 0$ and $\widehat{S}^{(b)-1} > 0$.

The similarity transformations (4.13) and (4.15) imply that

$$\begin{aligned} \rho(B) &= \rho(L^{-1}\widehat{C}L^{-T}) = \|L^{-1}\widehat{C}L^{-T}\|_2 \\ &\leq \|L^{-1}\widehat{C}^{(a)}L^{-T}\|_2 + \|L^{-1}\widehat{C}^{(b)}L^{-T}\|_2 \\ &= \rho(L^{-1}\widehat{C}^{(a)}L^{-T}) + \rho(L^{-1}\widehat{C}^{(b)}L^{-T}) \\ &= \rho(\widehat{D}^{-1}\widehat{C}^{(a)}) + \rho(\widehat{D}^{-1}\widehat{C}^{(b)}). \end{aligned} \tag{4.22}$$

We bound the last two quantities of this expression using (4.20). Let $\widehat{D} = \widetilde{D} + \widetilde{P}$, where $\widetilde{D}$ is a block diagonal matrix, each of whose blocks is a constant-coefficient tridiagonal matrix of the form of $\widehat{T}$ above, and $\widetilde{P}$ is block diagonal with blocks of the form P above. The nonzero entries of $\widetilde{P}$ are now negative. Let $\widetilde{C}^{(a)} = \widehat{C}^{(a)} - \widetilde{P}$ and $\widetilde{C}^{(b)} = \widehat{C}^{(b)} + \widetilde{P}$. Then the alternative splittings

$$\widehat{S}^{(a)} = \widetilde{D} - \widetilde{C}^{(a)}, \qquad \widehat{S}^{(b)} = \widetilde{D} - [-\widetilde{C}^{(b)}]$$

are also regular splittings, and the inequalities $\widetilde{C}^{(a)} \geq \widehat{C}^{(a)}$, $-\widetilde{C}^{(b)} \geq -\widehat{C}^{(b)}$ hold. Varga's theorem on regular splittings [20, Theorem 3.15] implies that

$$\rho(\widehat{D}^{-1}\widehat{C}^{(a)}) \leq \rho(\widetilde{D}^{-1}\widetilde{C}^{(a)}), \qquad \rho(\widehat{D}^{-1}\widehat{C}^{(b)}) \leq \rho(\widetilde{D}^{-1}\widetilde{C}^{(b)}).$$

Therefore, from (4.22), we have

$$\begin{aligned} \rho(B) &\leq \rho(\widetilde{D}^{-1}\widetilde{C}^{(a)}) + \rho(\widetilde{D}^{-1}\widetilde{C}^{(b)}) \leq \|\widetilde{D}^{-1}\|_2(\|\widetilde{C}^{(a)}\|_2 + \|\widetilde{C}^{(a)}\|_2) \\ &\leq \rho(\widetilde{D}^{-1})(\rho(\widetilde{C}^{(a)}) + \rho(\widetilde{C}^{(b)})). \end{aligned}$$

The spectral radius of $\widetilde{D}^{-1}$ is bounded as in (4.19) above; the denominator is now

$$a^2 + 2(\sqrt{|be|} - \sqrt{|cd|})^2 + 4\sqrt{bcde}(1 - \cos(\pi h)) .$$

The spectral radii of $\widetilde{C}^{(a)}$ and $\widetilde{C}^{(b)}$ are bounded by Gerschgorin's theorem:

$$\begin{gathered} \rho(\widetilde{C}^{(a)}) \leq \max(4\sqrt{bcde}, 2\sqrt{bcde} + |be|, 2\sqrt{bcde} + |cd|, |be| + |cd|), \\ \rho(\widetilde{C}^{(b)}) \leq 2(|be| + |cd|). \end{gathered}$$

We summarize this discussion as follows:

Theorem 5. *If* $be < 0$, $cd < 0$, *and inequality* (4.21) *holds, then the spectral radius of the block Jacobi iteration matrix for the reduced system satisfies*

$$\rho(B) \leq \frac{\max(4\sqrt{bcde}\,, 2\sqrt{bcde}+|be|\,, 2\sqrt{bcde}+|cd|\,, |be|+|cd|)+2(|be|+|cd|)}{a^2+2(\sqrt{|be|}-\sqrt{|cd|})^2+4\sqrt{bcde}(1-\cos(\pi h))}\,.$$

Substitution of the expressions of (3.4) and (3.5) into the results of Theorems 4 and 5 gives bounds for the specific difference schemes. Inequality (4.21) permits us to replace $\widehat{D}$ with (the constant diagonal) $\widetilde{D}$, but the inequality is not valid for arbitrary matrices. In terms of the expressions of (3.4) for large cell Reynolds numbers, a sufficient condition for (4.21) to hold is that $\sqrt{(\gamma^2-1)(\delta^2-1)} \leq 4$. We simplify the expression derived from Theorem 5 using the notation

$$\mu(\gamma\,, \delta) \equiv \max(4\sqrt{(\gamma^2-1)(\delta^2-1)}\,, 2\sqrt{(\gamma^2-1)(\delta^2-1)}+\gamma^2-1\,,$$
$$2\sqrt{(\gamma^2-1)(\delta^2-1)}+\delta^2-1\,, \gamma^2-1+\delta^2-1).$$

Depending on the values of γ and δ, any of the four quantities defining μ can determine its value.

Corollary 6. *For the centered difference scheme, if* $|\gamma| < 1$ *and* $|\delta| < 1$, *then the spectral radius of the block Jacobi iteration matrix for the reduced system is bounded by*

$$\frac{(\sqrt{1-\gamma^2}+\sqrt{1-\delta^2})^2}{8-(\sqrt{1-\gamma^2}+\sqrt{1-\delta^2})^2+2\sqrt{(1-\gamma^2)(1-\delta^2)}(1-\cos(\pi h))}\,.$$

If $|\gamma| > 1$, $|\delta| > 1$, *and* $\sqrt{(\gamma^2-1)(\delta^2-1)} \leq 4$, *then the spectral radius is bounded by*

$$\frac{\frac{1}{2}\mu(\gamma\,, \delta)+\gamma^2-1+\delta^2-1}{8+(\sqrt{\gamma^2-1}-\sqrt{\delta^2-1})^2+2\sqrt{(\gamma^2-1)(\delta^2-1)}(1-\cos(\pi h))}\,.$$

For the upwind difference scheme, the spectral radius is bounded by

$$\frac{(\sqrt{1+2\gamma}+\sqrt{1+2\delta})^2}{2(2+\gamma+\delta)^2-(\sqrt{1+2\gamma}+\sqrt{1+2\delta})^2+2\sqrt{(1+2\gamma)(1+2\delta)}(1-\cos(\pi h))}\,.$$

See §4.5 for comparisons of these asymptotic bounds with those for the full system (Corollary 2).

Although these results are stated only for the unit square, the arguments are trivially adapted for general rectangular domains with uniform meshes. They also provide upper bounds for irregular domains. For example, the number of points in the longest diagonal line determines the inequality of (4.19).

4.4. Fourier analysis. As we will show in §5, the bounds of §4.3 agree with the results of numerical computations when $be > 0$ and $cd > 0$, but they are pessimistic when $be < 0$ and $cd < 0$. We now present a Fourier analysis of a variant of the symmetrized reduced operator using the methodology of [2]. Consider the discrete nine-point operator of Figure 4.5. This operator is based on the version of $\widehat{S}$ of Figure 4.4, except that it is defined on a rectilinear grid with periodic boundary conditions. The horizontal lines of the rectilinear grid correspond to the lines oriented in the NW–SE direction of the skewed grid. (In the figure, the orientation of the skewed grid is indicated in parentheses.) We refer to this operator as the rectilinear periodic reduced operator.

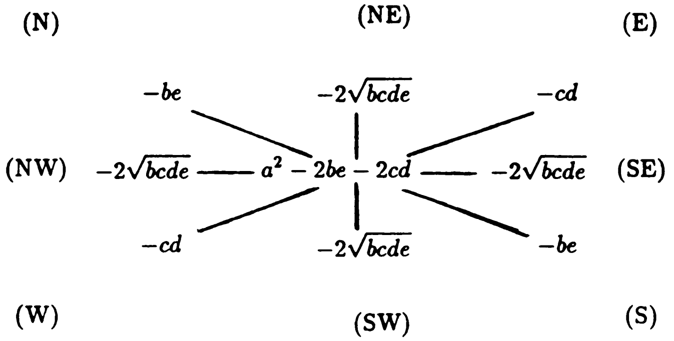

FIGURE 4.5. A computational molecule for the rectilinear periodic reduced operator.

The Fourier analysis is defined as follows; see [2] for a more detailed description. Suppose the rectilinear grid is contained in a square domain with n interior points in each direction and periodic boundary conditions. Let $\widehat{S}_P$ denote the operator defined by the computational molecule of Figure 4.5. That is, if v is a mesh function with value v_{jk} at the (j,k) mesh point, $1 \le j,\ k \le n$, then

$$\begin{aligned}(\widehat{S}_P v)_{jk} \equiv (a^2 - 2be - 2cd)v_{jk} &- 2\sqrt{bcde}v_{j-1,k} - 2\sqrt{bcde}v_{j+1,k}\\ &- bev_{j-1,k+1} - 2\sqrt{bcde}v_{j,k+1} - cdv_{j+1,k+1}\\ &- cdv_{j-1,k-1} - 2\sqrt{bcde}v_{j,k-1} - bev_{j+1,k-1}.\end{aligned}$$

The analogue of the line Jacobi splitting considered above is

$$\widehat{S}_P = \widehat{D}_P - \widehat{C}_P,$$

where $\widehat{D}_P$ corresponds to the horizontal connections of Figure 4.5 (indices $(j-1,k)$, (j,k), and $(j,k+1)$), and $\widehat{C}_P$ corresponds to the other connections. Let $v = v^{(s,t)}$ have values $v_{jk}^{(s,t)} = e^{ij\theta_s}e^{ik\phi_t}$, where $\theta_s = 2\pi sh$, $\phi_t = 2\pi th$, $1 \le s,\ t \le n$, and $h = 1/(n+1)$. It is straightforward to show by direct substitution that

$$(\widehat{S}_P v)_{jk} = \lambda v_{jk}, \qquad (\widehat{D}_P v)_{jk} = \psi v_{jk}, \qquad (\widehat{C}_P v)_{jk} = \mu v_{jk},$$

where

$$\psi = \psi_{st} = a^2 - 2be - 2cd - 4\sqrt{bcde}\cos\theta_s\,,$$
$$\mu = \mu_{st} = 4\sqrt{bcde}\cos\phi_t + 2cd\cos(\theta_s + \phi_t) + 2be\cos(\theta_s - \phi_t)\,,$$

and $\lambda = \psi - \mu$. The quantities ψ and μ are the eigenvalues of $\widehat{D}_P$ and $\widehat{C}_P$, respectively, corresponding to the (shared) eigenvector v. The analogous eigenvalue of $\widehat{D}_P^{-1}\widehat{C}_P$ is

$$\frac{4\sqrt{bcde}\cos\phi_t + 2cd\cos(\theta_s + \phi_t) + 2be\cos(\theta_s - \phi_t)}{a^2 - 2be - 2cd - 4\sqrt{bcde}\cos\theta_s}\,. \tag{4.23}$$

The maximal value of this expression over all θ_s and ϕ_t, $1 \le s$, $t \le n$, is a heuristic bound for the maximal eigenvalue for the analogous Dirichlet operator. In the following, we will be concerned with asymptotic bounds as $h \to 0$ (for *fixed* γ and δ). For simplicity we examine (4.23) for all θ_s, $\phi_t \in [0, 2\pi]$. This ignores some $O(h^2)$ effects that are significant only when $\gamma = \delta = 0$.

Lemma 3. *For $be > 0$ and $cd > 0$, the maximum Fourier eigenvalue* (4.23) *is*

$$\frac{2(\sqrt{be} + \sqrt{cd})^2}{a^2 - 2(\sqrt{be} + \sqrt{cd})^2}\,. \tag{4.24}$$

For $be < 0$ and $cd < 0$, the maximum Fourier eigenvalue is

$$\frac{2(\sqrt{|be|} + \sqrt{|cd|})^2}{a^2 + 2(\sqrt{|be|} + \sqrt{|cd|})^2}\,. \tag{4.25}$$

Proof. When $be > 0$ and $cd > 0$, the numerator of (4.23) is maximized when $\cos\phi_t = 1$, $\cos(\theta_s + \phi_t) = 1$, and $\cos(\theta_s + \phi_t) = 1$. The denominator is minimized when $\cos\theta_s = 1$. These conditions are all satisfied when $\theta_s = \phi_t = 0$, which results in (4.24).

When $be < 0$ and $cd < 0$, it is not possible to maximize the numerator and minimize the denominator simultaneously. Expression (4.25) corresponds to the values $\theta = \pi$ and $\phi = 0$. To see that (4.25) is maximal, consider the change of variables $x = \sqrt{|cd|}$ and $y = \sqrt{|be|}$. Then we wish to establish the inequality

$$\frac{2xy\cos\phi - x^2\cos(\theta + \phi) - y^2\cos(\theta - \phi)}{a^2/2 + x^2 + y^2 - 2xy\cos\theta} \le \frac{(x + y)^2}{a^2/2 + (x + y)^2}\,,$$

for $x > 0$, $y > 0$, and θ, $\phi \in [0, 2\pi]$. We need only consider the case where the left-hand side is positive, so that the inequality is equivalent to

$$(x + y)^2 - 2xy\cos\phi + x^2\cos(\theta + \phi) + y^2\cos(\theta - \phi)$$
$$\ge (1 + \cos\theta)2xy\frac{(x + y)^2}{a^2/2 + (x + y)^2}\,.$$

Since $(x+y)^2/(a^2/2+(x+y)^2) < 1$, it suffices to show that

$$x^2+y^2+x^2\cos(\theta+\phi)+y^2\cos(\theta-\phi)-2xy(\cos\theta+\cos\phi)\geq 0.$$

For any fixed y, θ, and ϕ (provided $\cos(\theta+\phi)\neq -1$), it follows from elementary calculus that (considered as a function of x) the expression on the left of this inequality has minimum value

$$y^2\left[(1+\cos(\theta-\phi))-\frac{(\cos\theta+\cos\phi)^2}{(1+\cos(\theta+\phi))}\right],$$

which is identically zero. The case $\cos(\theta+\phi)=-1$ is covered by minimizing with respect to y in an analogous way. □

Substitution of the values of b, c, d, and e from the two difference schemes yields the following result.

Theorem 6. *Fourier analysis of the rectilinear periodic reduced operator yields the following asymptotic bounds on the Jacobi iteration matrix. For centered differences with* $|\gamma|<1$, $|\delta|<1$:

$$\frac{(\sqrt{1-\gamma^2}+\sqrt{1-\delta^2})^2}{8-(\sqrt{1-\gamma^2}+\sqrt{1-\delta^2})^2}.$$

For centered differences with $|\gamma|>1$, $|\delta|>1$:

$$\frac{(\sqrt{\gamma^2-1}+\sqrt{\delta^2-1})^2}{8+(\sqrt{\gamma^2-1}+\sqrt{\delta^2-1})^2}.$$

For upwind differences:

$$\frac{(\sqrt{1+2\gamma}+\sqrt{1+2\delta})^2}{2(2+\gamma+\delta)^2-(\sqrt{1+2\gamma}+\sqrt{1+2\delta})^2}.$$

Note that the first and third of these bounds agree with the analogous asymptotic results from Corollary 6. The second bound does not depend on any restrictions on γ and δ.

4.5. Comparison of bounds. We compare the asymptotic bounds (as $h\to 0$) on the spectral radii for the block Jacobi operators from the full system (Corollary 2) and the reduced system (Corollary 6 and Theorem 6). For the full system, we consider the minimum of the bounds for the rowwise and columnwise ordered matrices. We do not have analytic results that cover all cell Reynolds numbers. Instead, we make a numerical comparison. In each of three cases, we graph the bounds on four cross-sections of a square region in the (γ,δ) plane. These cross-sections correspond to the four choices $\delta\approx 0$, $\delta=\gamma$, $\delta\approx$ its maximal value on the region, and $\delta\approx$ its midpoint on the region. The bounds are symmetric with respect to γ and δ (since we are using the minimum of the two

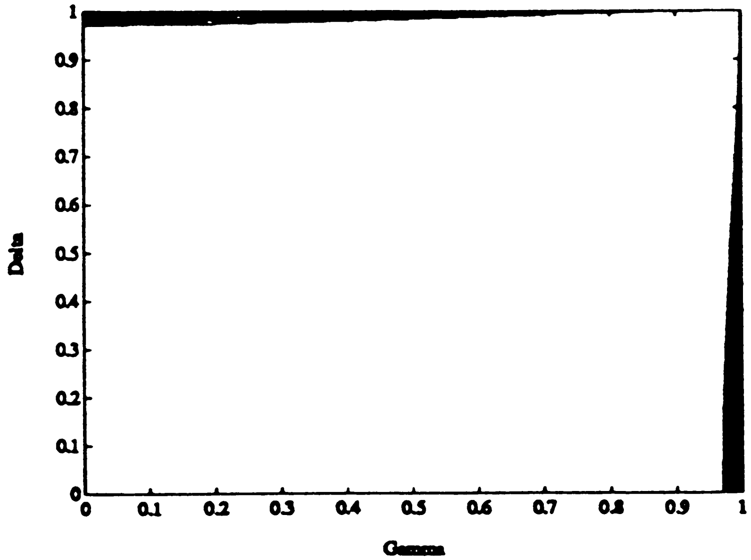

FIGURE 4.6. Values of γ, δ where the full system bounds are smaller than the reduced system bounds, for centered differences and small cell Reynolds numbers.

full system bounds), so that the graphs of Figures 4.7–4.9 below also compare the results with the roles of γ and δ interchanged.

1. *Centered differences,* $\gamma < 1$ *and* $\delta < 1$. In this case, neither bound is uniformly better for all choices of γ and δ. In Figure 4.6, the shaded region shows the values of γ and δ where the full system bounds are smaller than the reduced system bounds.[1] Figure 4.7 compares the bounds on four cross-sections corresponding to $\delta = .02$, $\delta = \gamma$, $\delta = .98$, and $\delta = .48$. A significant part of the bottom left picture corresponds to the shaded region of Figure 4.6. Taken together, these figures show that the reduced system bound is smaller for most values of γ and δ, and the two bounds are very close (and small) when the full system bound is better.

2. *Centered differences,* $\gamma > 1$ and $\delta > 1$. For this case, we consider both the rigorous reduced system bound from §4.3 and the bound of the Fourier analysis from §4.4. In our examples, the restrictions on γ and δ in Corollary 6 are satisfied whenever the bounds are less than one. As we show in §5, the Fourier bounds are closer to the spectral radii observed in experiments. Figure 4.8 shows the bounds on four cross-sections in the region $1 < \gamma$, $\delta < 3$. The figure clearly shows that the bounds for the reduced system are smaller than those for the full system. For most large γ and δ, the latter bounds (which are

[1] These regions were determined by first approximately identifying the shaded region using a mesh of size .02, and then using a fine mesh in $[0, 1] \times [.97, 1]$, with horizontal length .01 and vertical length .0005. The bounding curve at the top of the figure is a plot of values (γ, δ) on the fine mesh such that δ is the largest value for each γ where the reduced system bound is lower. The curve on the right comes from symmetry of the bounds.

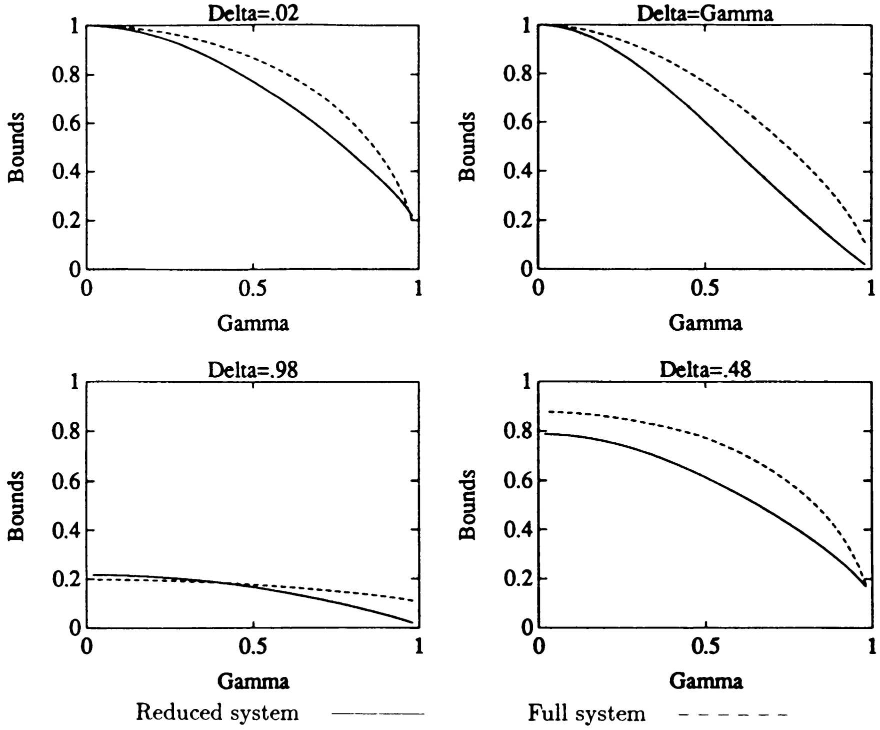

FIGURE 4.7. Asymptotic bounds for the block Jacobi iteration matrices on four cross-sections of the (γ, δ) plane, for centered differences and small cell Reynolds numbers.

tight) are greater than one, so that the block Jacobi iteration applied to the full system is divergent.

3. *Upwind differences.* In numerical comparisons for $0 < \gamma, \delta < 5$ (using ninety-nine mesh points in each direction), we observed that the bounds for the reduced system are uniformly smaller than those for the full system. Figure 4.9 graphs the bounds on four cross-sections of $0 < \gamma, \delta < 3$.

Note that although we only consider positive γ and δ here, the results for centered differences apply in general when absolute values are used. The analysis does not apply in cases where one of $|\gamma|$, $|\delta|$ is greater than one and the other is less than one; numerical experiments for cases of this type are described in §5.

We conclude with some remarks concerning acceleration, using either block SOR or the conjugate gradient method (CG) [10] with preconditioning by the block diagonal. In the symmetrizable cases, the eigenvalues of the block Jacobi matrices are real. Hence, Young's SOR analysis [20, 21] applies, i.e., the optimal SOR iteration parameter can be obtained from the spectral radius of the block Jacobi matrix. Symmetrization is needed only for the analysis; the actual computations can be performed with the nonsymmetric matrices. In contrast,

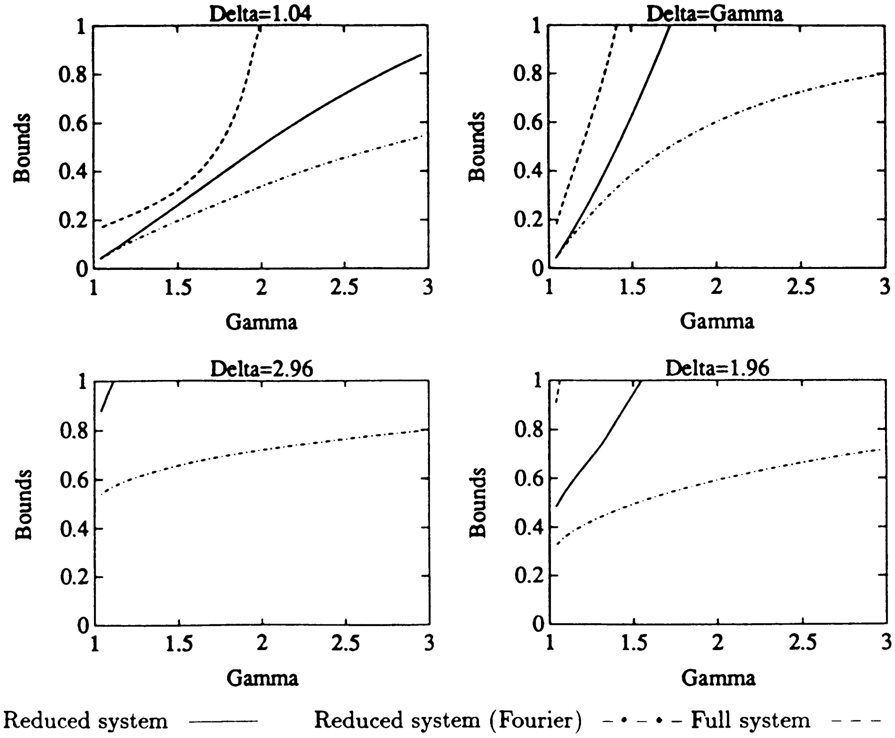

FIGURE 4.8. Asymptotic bounds for the block Jacobi iteration matrices on four cross-sections of the (γ, δ) plane, for centered differences and large cell Reynolds numbers.

CG requires either the explicit computation of the symmetrized matrices $\widehat{C}$ and $\widehat{D}$, or (what is equivalent), a change in the inner product used in the CG iteration. Both schemes require the explicit use of the symmetrizing operator Q. But, as in the one-dimensional case, the entries of Q may be very large. (For example, if $0 < \gamma < 1$ and $\delta = 0$, then (4.7) is identical to the recurrence for one dimension.) As a result, CG is not always viable in floating-point arithmetic. Finally, we note that for the full system, it is shown in [4] that SOR iteration may be convergent for values of γ and δ where the block Jacobi method is divergent, as well as for values where our analysis does not apply.

5. NUMERICAL EXPERIMENTS

In this section, we present the results of numerical experiments that confirm and supplement the analysis of §4. In all cases, we present our results for the block Gauss-Seidel splitting $S = [D - L] - U$, where L and U are the lower and upper triangles of C, respectively. (Since S has block Property A, $\rho((D-L)^{-1}U) = [\rho(D^{-1}C)]^2$.) In particular, we compare the bounds of Corollary 6 with computed values for several different mesh sizes, and we also examine the effectiveness of the block Gauss-Seidel method in cases where the analy-

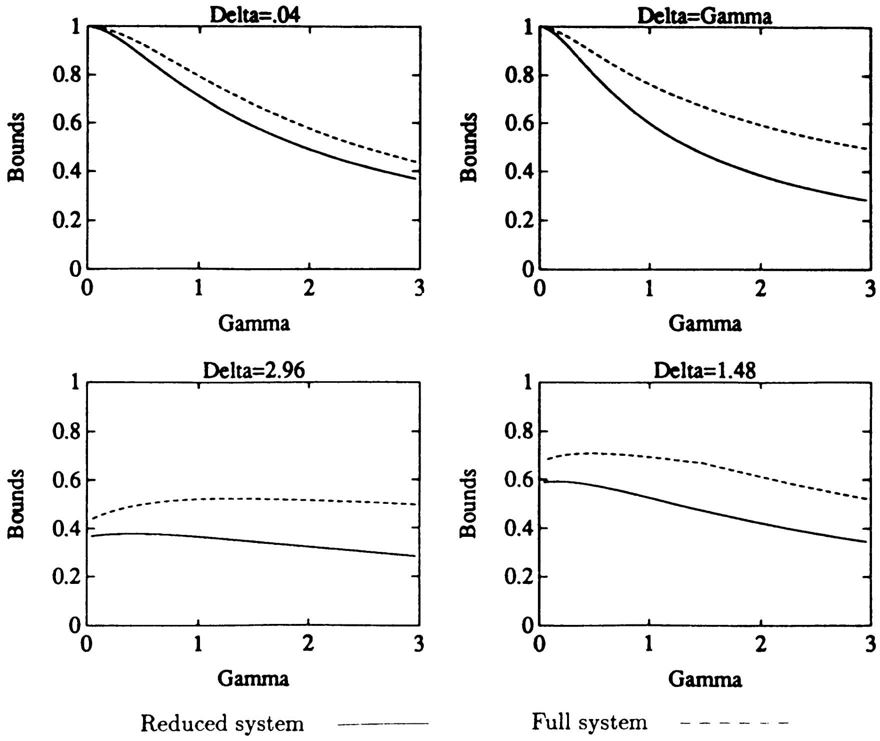

FIGURE 4.9. Asymptotic bounds for the block Jacobi iteration matrices on four cross-sections of the (γ, δ) plane, for upwind differences.

sis is not applicable (e.g. for the centered difference discretization where $|\gamma| > 1$ and $|\delta| < 1$). In addition, we present numerical results for constant-coefficient problems where Dirichlet boundary conditions are replaced by outflow boundary conditions on portions of the boundary, and for several variable-coefficient problems. All experiments were performed on a VAX-8600 in double-precision Fortran. The reduced matrices were computed using PCGPAK [17]. The spectral radii $\rho((D-L)^{-1}U)$ were determined using the QZ algorithm in EISPACK [9, 10].

Table 5.1 shows computed values of the spectral radii of the block Gauss-Seidel iteration matrices derived from the centered difference discretization of (3.1), for $\tau = 0$ (so that $\delta = 0$). For $\gamma \leq 1$, the table also shows the asymptotic bound for these quantities derived by squaring the first expression of Corollary 6, with $h = 0$. The value for $\gamma = 1$ is the limit as $\gamma \rightarrow 1$. Numerical experiments with $\gamma = 0$ and varying δ produced the same spectral radii; since the bounds of Corollary 6 are symmetric with respect to γ and δ, Table 5.1 also applies for the case $\gamma = 0$ and δ taking on the values in the first column. The results for $\gamma < 1$ show that the limiting values of the spectral radii tend to the bounding value as $h \rightarrow 0$ (for γ fixed). The analytic bounds for the values of h in the table are closer to the asymptotic bounds than to the computed spectral radii. The results also show that the method is highly effective for $\gamma > 1$, where S is not symmetrizable and our analysis does not apply. In this case, some computed eigenvalues of each Gauss-Seidel matrix are complex.

TABLE 5.1. Spectral radii and bounds for the Gauss-Seidel iteration matrices, centered differences, $\delta = 0$.

γ	$h = 1/8$	$h = 1/16$	$h = 1/32$	Asymptotic Bound
.2	.50	.79	.89	.92
.4	.40	.62	.69	.72
.6	.26	.40	.45	.46
.8	.13	.19	.21	.22
1.0	.01	.02	.05[2]	.02
1.2	.03	.03	.04	–
1.4	.04	.05	.06	–
1.6	.08	.08	.08	–
1.8	.10	.10	.11	–
2.0	.10	.10	.15	–

TABLE 5.2. Spectral radii and bounds for the Gauss-Seidel iteration matrices, centered differences, $\gamma = \delta$.

γ	$h = 1/8$	$h = 1/16$	$h = 1/32$	Asymptotic Bound	Fourier
.2	.46	.73	.82	.85	.85
.4	.30	.46	.51	.52	.52
.6	.13	.19	.21	.22	.22
.8	.03	.04	.05	.05	.05
1.0	0	0	0	0	0
1.2	.02	.03	.03	.05	.03
1.4	.07	.10	.10	.23	.11
1.6	.14	.18	.19	.61	.19
1.8	.21	.26	.27	1.25	.28
2.0	.27	.33	.35	2.25	.36

Table 5.2 shows the computed spectral radii of the block Gauss-Seidel iteration matrices for the centered difference discretization and $\gamma = \delta$. This table contains both the rigorous bounds from Corollary 6 and the bounds of Theorem 6 derived using Fourier analysis. The two bounds agree when $\gamma < 1$ and

[2] We suspect that the eigenvalue problem is ill-conditioned when $\gamma = 1$, and that this is why this computed spectral radius exceeds the asymptotic bound.

$\delta < 1$. The rigorous bounds are pessimistic when $\gamma > 1$ and $\delta > 1$, and the Fourier results agree with experimental results. We believe the source of the pessimism is the inequality of (4.22), which prevents any cancellations due to the opposite signs of $\widehat{C}^{(a)}$ and $\widehat{C}^{(b)}$ from being put to use. The Fourier analysis does not require an analogous inequality. Note that in both Tables 5.1 and 5.2, for moderate values of γ, the more highly nonsymmetric problems are easier to solve than the nearly symmetric problems.

Our analysis applies only for discretizations of problems with Dirichlet boundary conditions. However, it is known that the discrete solutions are more meaningful physically when *outflow* boundary conditions

$$u_x = 0 \quad \text{if } \sigma > 0, \qquad u_y = 0 \quad \text{if } \tau > 0$$

are used [4, 13]. (In particular, for centered difference discretizations, when $\gamma > 1$ or $\delta > 1$, Dirichlet boundary conditions result in oscillatory discrete solutions at boundary layers, whereas outflow boundary conditions result in smooth solutions.) In Table 5.3, we compare the spectral radii for the block Gauss-Seidel iteration matrices arising from Dirichlet problems with those arising from outflow boundary conditions. The left side of the table is for the case $\delta = 0$ (i.e., $\tau = 0$ in (3.1)), and the right-hand side is for $\delta = \gamma$ $(\sigma = \tau)$. The data is for mesh size $h = 1/32$. The outflow boundary conditions were discretized by first-order differences [13]

$$u_x(1, y_j) \approx \frac{u_{n,j} - u_{n-1,j}}{h}, \qquad u_y(x_i, 1) \approx \frac{u_{i,n} - u_{i,n-1}}{h}.$$

The results show that the behavior for outflow boundary conditions is nearly identical to that for Dirichlet conditions.

TABLE 5.3. Comparison of spectral radii of the Gauss-Seidel iteration matrices derived from centered differences and either Dirichlet or outflow boundary conditions, $h = 1/32$.

$\delta = 0$		
γ	Dirichlet	Outflow
.2	.888	.892
.4	.694	.695
.6	.447	.448
.8	.214	.214
1.0	.053	.053
1.2	.036	.036
1.4	.056	.056
1.6	.081	.081
1.8	.112	.109
2.0	.147	.141

$\delta = \gamma$		
γ	Dirichlet	Outflow
.2	.820	.826
.4	.506	.507
.6	.214	.215
.8	.047	.047
1.0	0	0
1.2	.032	.032
1.4	.103	.103
1.6	.188	.188
1.8	.273	.273
2.0	.353	.353

Finally, we describe some results for problems with variable coefficients, of the form $-\Delta u + f(x, y)u_x + g(x, y)u_y = 0$ on $\Omega = (0, 1) \times (0, 1)$, $u = g$ on $\partial\Omega$. We consider four problems, which are taken from [1, 19]:

1. $-\Delta u + \sigma x^2 u_x + \sigma x^2 u_y = 0$,
2. $-\Delta u + \frac{1}{2}\sigma(1 + x^2)u_x + 100u_y = 0$,
3. $-\Delta u + \sigma x^2 u_x = 0$,
4. $-\Delta u + \sigma(1 - 2x)u_x + \sigma(1 - 2y)u_y = 0$,

with $u = 0$ on $\partial\Omega$. As in [1, 19], we descretized each problem using centered differences and mesh size $h = 1/20$. The spectral radii of the Gauss-Seidel iteration matrices are shown in Table 5.4. For reference, the table also reports $\gamma = \sigma h/2$. For Problems 1 and 3, γ represents the maximum cell Reynolds number on the mesh, the minimum being 0. For Problem 2, it represents the maximum cell Reynolds number for the x-coordinate; the minimum is $\gamma/2$, and the y-coordinate has constant value $\delta = 5$. For Problem 4, the coefficients change sign in Ω; the cell Reynolds numbers in each coordinate vary between $-\gamma$ and γ.

TABLE 5.4. Spectral radii for the Gauss-Seidel iteration matrices of four problems with variable coefficients, centered differences, $h = 1/20$.

σ	$\gamma = \frac{\sigma h}{2}$	Problem 1	Problem 2	Problem 3	Problem 4
1	.03	.91	.23	.91	.90
10	.26	.91	.23	.92	.80
100	2.6	.78	.40	.83	.18
1000	26	.96	.94	.89	.95
10000	263	.998	.999	.994	.999

Although it is difficult to make definitive statements about these results, they appear to be consistent with the analysis of the constant-coefficient case. In particular, for moderate σ (the three smaller values), the spectral radii are bounded well below one for all four problems. For Problems 1, 3, and 4, performance improves as $|\gamma|$ increases from 0; and for Problem 2 (where $\delta = 5$) it is very good for moderate γ. Performance declines when γ gets very large; in these cases finer meshes will improve both performance of the iterative solver and accuracy of the discrete solution. We remark that for Problem 2 with $\sigma \leq 10$, the Gauss-Seidel matrices have complex eigenvalues close in modulus to their spectral radii. In many of the other cases (e.g., all instances of Problem 1), some computed eigenvalues contain small imaginary parts, of order at most 10^{-2}.

6. CONCLUSIONS

We have performed an analytic and experimental study of a block iterative method for solving the reduced system derived from a class of discrete non-self-adjoint elliptic problems. The analysis provides rigorous justification for the effectiveness of the reduced system methodology previously observed empirically, and it shows that the use of the reduced system often results in faster convergence than if the full system is solved by analogous iterative methods. The experimental results show that the method is also effective for problems where the analysis does not apply. We close with the observation that the computations under consideration are naturally divided into individual subtasks (corresponding to tridiagonal subblocks), so that they can be implemented very efficiently on parallel computers.

Acknowledgment

The authors wish to thank Clyde Kruskal for help with the proof of Lemma 3, and Paul Concus for several helpful discussions.

Bibliography

1. E. F. F. Botta and A. E. P. Veldman, *On local relaxation methods and their application to convection-diffusion equations*, J. Comput. Phys. **48** (1981), 127–149.
2. T. F. Chan and H. C. Elman, *Fourier analysis of iterative methods for elliptic problems*, SIAM Rev. **31** (1989), 20–49.
3. R. Chandra, *Conjugate gradient methods for partial differential equations*, Ph.D. Thesis, Department of Computer Science, Yale University, 1978.
4. R. C. Y. Chin and T. A. Manteuffel, *An analysis of block successive overrelaxation for a class of matrices with complex spectra*, SIAM J. Numer. Anal. **25** (1988), 564–585.
5. R. C. Y. Chin, T. A. Manteuffel, and J. de Pillis, *ADI as a preconditioning for solving the convection-diffusion equation*, SIAM J. Sci. Statist. Comput. **5** (1984), 281–299.
6. A. R. Curtis, *On a property of some test equations for finite difference methods*, IMA J. Numer. Anal. **1** (1981), 369–375.
7. S. C. Eisenstat, H. C. Elman, and M. H. Schultz, *Block-preconditioned conjugate-gradientlike methods for numerical reservoir simulation*, SPE Reservoir Engineering **3** (1988), 307–312.
8. H. C. Elman, *Iterative methods for large, sparse, nonsymmetric systems of linear equations*, Ph.D. Thesis, Department of Computer Science, Yale University, 1982.
9. R. S. Garbow, J. M. Boyle, J. J. Dongarra, and C. B. Moler, *Matrix eigensystem routines: EISPACK guide extension*, Springer-Verlag, New York, 1972.
10. G. H. Golub and C. F. van Loan, *Matrix computations*, The Johns Hopkins University Press, Baltimore, MD, 1983.
11. L. A. Hageman, F. T. Luk, and D. M. Young, *On the equivalence of certain acceleration methods*, SIAM J. Numer. Anal. **17** (1980), 852–873.
12. L. A. Hageman and D. M. Young, *Applied iterative methods*, Academic Press, New York, 1981.
13. G. W. Hedstrom and A. Osterheld, *The effect of cell Reynolds number on the computation of a boundary layer*, J. Comput. Phys. **37** (1980), 399–421.
14. S. V. Parter, *On estimating the "rates of convergence" of iterative methods for elliptic difference equations*, Trans. Amer. Math. Soc. **114** (1965), 320–354.
15. ——, *The use of linear graphs in Gaussian elimination*, SIAM Rev. **3** (1961), 119–130.
16. S. V. Parter and J. W. T. Youngs, *The symmetrization of matrices by diagonal matrices*, J. Math. Anal. Appl. **4** (1962), 102–110.
17. *PCGPAK User's Guide*, Version 1.04, Scientific Computing Associates, New Haven, CT, 1987.

18. A. Segal, *Aspects of numerical methods for elliptic singular perturbation problems*, SIAM J. Sci. Statist. Comput. **3** (1982), 327–349.

19. M. C. Thompson, J. H. Ferziger, and G. H. Golub, *Block SOR applied to the cyclically-reduced equations as an efficient solution technique for convection-diffusion equations*, in Computational Techniques and Applications, CTAC-87 (John Noye and Clive Fletcher, eds.), North-Holland, New York, 1988, pp. 637–646.

20. R. S. Varga, *Matrix iterative analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1962.

21. D. M. Young, *Iterative solution of large linear systems*, Academic Press, New York, 1971.

Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland 20742. *E-mail*: elman@helios.cs.umd.edu

Department of Computer Science, Stanford University, Stanford, California 94305. *E-mail*: na.golub@na-net.stanford.edu