JOURNAL OF COMPUTATIONAL PHYSICS 67, 91–123 (1986)

Solution of Elliptic PDEs by Fast Poisson Solvers Using a Local Relaxation Factor*

SIN-CHUNG CHANG

National Aeronautics and Space Administration; Lewis Research Center; Cleveland, Ohio 44135

Received May 10, 1985; revised January 15, 1986

Using a local discrete Fourier analysis, one- and two-step iterative procedures are developed to solve a large class of 2-D and 3-D nonseparable elliptic partial differential equations (PDEs). The one-step procedure is a modified D'Yakanov Gunn iterative procedure in which the relaxation factor is grid point dependent. The two-step procedure is designed to accelerate the one-step procedure. Both are easy to implement, and applicable to a variety of boundary conditions. They are also computationally efficient as indicated by the results of numerical comparisons with other established methods. Furthermore, these new algorithms possess two important properties which the traditional iterative methods lack, i.e., (i) the convergence rate is relatively insensitive to grid cell size and aspect ratio, and (ii) the convergence rate can be easily estimated using the coefficient of the PDE being solved. For a set of constant coefficient model problems, it is shown theoretically that the two-step vs. one-step computational efficiency ratio ranges between 1 and 2. The higher ratio is realized whenever the need to accelerate the one-step procedure is greater, i.e., whenever its convergence rate is lower. It is also shown numerically that the two-step procedure can substantially outperform the one-step procedure in the numerical solution of many PDEs with variable coefficients.

1. INTRODUCTION

Since the middle sixties, fast direct solvers (FDS) have been developed for the numerical solution of separable elliptic partial differential equations (PDE) [1–5]. Based on Fourier analysis and cyclic reduction, FDS algorithms are most effective on a uniform rectangular grid. They can obtain the solution with efficiency far beyond the reach of the traditional iterative procedures such as Successive Over-Relaxation (SOR) methods.

Generally, FDS algorithms are not directly applicable to an elliptic problem with either (i) a computation domain of irregular shape, or (ii) a nonseparable PDE. Limitation (i) may be circumvented either by mapping the original domain onto a rectangular domain or by using the capacity matrix method [5]. For limitation (ii),

* Partial results of this paper were presented at the Ninth International Conference on Numerical Methods in Fluid Dynamics, Saclay, France, June 25–29, 1984.

0021-9991/86 $3.00

it can be circumvented by a semidirect procedure, i.e., an iterative procedure driven by an FDS. In this study, new semidirect procedures are developed and tested on both two-dimensional (2-D) and three-dimensional (3-D) problems. These new procedures are easy to implement, computationally efficient, and applicable to a variety of boundary conditions (BCS). Furthermore, they possess two important properties which the traditional iterative methods lack, i.e., (i) the convergence rate is insensitive to grid cell size and aspect ratio, and (ii) the convergence rate can be easily estimated using the coefficients of the PDE being solved.

Many elliptic PDEs can be expressed as

$$Qu = h, \tag{1.1}$$

where Q is a nonseparable second-order linear elliptic operator, u the dependent variable, and h a given source term. Equation (1.1) may be solved with the one-step procedure

$$P(u^{n+1} - u^n) = -\tau(Qu^n - h). \tag{1.2}$$

Here n is the iteration number, τ a nonzero relaxation factor and P a separable elliptic operator which can be directly inverted by an FDS. This procedure is a continuous analogue of the D'Yakanov–Gunn iterations [6]. In sharp contrast to the traditional iterative procedures, the advancement of u^n to u^{n+1} at any grid point during each D'Yakanov–Gunn iteration is influenced by the boundary conditions and the values of u^n at all grid points. As a result, the D'Yakanov–Gunn iterative procedure is far more effective than the traditional iterative procedures in spreading long-wavelength information. Therefore one may expect that it is also more effective in reducing long-wavelength errors.

The iterative procedure (1.2) was utilized by Concus and Golub [7] and Bank [8] in their works on the numerical solution of nonseparable elliptic equations. In these works the relaxation factor τ is treated as a constant and the iteration is accelerated by an optimal choice of τ. In the current paper, it will be shown that a more efficient algorithm may be obtained by using a spatially varying relaxation factor.

The use of a local (spatially varying) relaxation factor in the current study is motivated by an earlier study of a semidirect procedure [9, 10]. In the previous study, it was found that both the global and local convergence rates can be predicted using a von Neumann (Fourier) analysis. Since this analysis handles variable coefficient problems by freezing the coefficients at their local values, heuristically one may say that it is valid only for the errors whose wavelengths are much shorter than those of the coefficients. The success of this analysis in predicting the global and local convergence rates indicates that long-wavelength errors are indeed of secondary importance in a semidirect procedure. Furthermore, since the theoretical local convergence rate evaluated using the von Neumann analysis is a function of the local value of τ, this success suggests that iteration (1.2) can be accelerated if the

values of τ at all grid points are chosen to maximize the theoretical local convergence rate. Recently, SOR-related local relaxation methods were also developed by Ehrlich [11] and Botta and Veldman [12]. Although these methods generally are more efficient than the classical SOR methods, their convergence rates are sensitive to grid cell size and aspect ratio. Moreover, for variable coefficient problems, the convergence rates of these methods are difficult to predict.

As shown by the work of Bank [8], iteration (1.2) can also be accelerated by choosing an operator P which closely resembles the operator Q. Application of this technique, however, could be limited by the following considerations: (a) this technique may require the use of a general separable operator P. This, however, is computationally inefficient since an FDS code for a general separable operator is about five times slower than one for the Laplacian ∇^2 [5]; (b) to apply this technique, the FDS code for the operator P generally must be custom-made. This may require a considerable effort. The above considerations lead us to choose $P = \nabla^2$ or its equivalent in the current study.

In the current paper, we also investigate the two-step iterative procedure

$$Pw^n = -\tau(Qu^n - h) \qquad (\tau \neq 0) \tag{1.3}$$

$$P(u^{n+1} - u^n) = Rw^n, \tag{1.4}$$

where w^n is the intermediate iterative variable and R, an elliptic linear operator whose exact form will be chosen to accelerate the convergence. Assuming that (a) $u^n \to u$ and $w^n \to w$ as $n \to +\infty$, and (b) the inverse of R exists, one may conclude that $w = 0$ and u is a solution of Eq. (1.1).

If one assumes that the computational effort required to advance u^n to u^{n+1} using the two-step procedure is twice that required using the one-step procedure, the former will have no merit unless it can achieve a convergence rate at least twice that of the latter. Fortunately, for a large class of the operator Q, an operator R can be found such that the two-step vs one-step convergence rate ratio varies between 2 and 4. Moreover, as will be shown, the higher ratio is realized whenever the need to accelerate the one-step procedure is greater, i.e., whenever its convergence rate is lower.

In Section 2, Eqs. (1.1) to (1.4) are discretized and the convergence rates of the one- and two-step procedures are studied for a constant coefficient operator Q assuming the iterative errors satisfy the periodic boundary conditions. The analysis is a rigorous version of the von Neumann analysis. Its results are used to determine the optimal value of τ and, in the case of two-step procedure, the optimal form of R. In Section 3, the results obtained in Section 2 are extended to solve PDEs with variable coefficients. In Section 4, the current methods are numerically evaluated using a variety of 2-D nonseparable elliptic PDE's. In addition to the advantages noted previously, the results of this numerical evaluation indicate that the current procedures can be used to solve PDEs with a cross derivative term and works very well for many PDEs with rapidly varying coefficients.

Finally, in Section 5, the current procedures are incorporated into an Euler

Solver [9, 10] to obtain solutions for 3-D incompressible flows in a 180-degree turning channel. It is shown that the current procedures converge with rates much higher than those reported in [10].

2. Analysis

As an initial step, iteration (1.2) is studied assuming that τ is a constant and

$$Q \stackrel{\text{def}}{=} a\frac{\partial^2}{\partial x^2} + 2b\frac{\partial^2}{\partial x\,\partial y} + c\frac{\partial^2}{\partial y^2}, \tag{2.1}$$

where a, b, and c are arbitrary constants subjected to the elliptic conditions

$$a > 0, \qquad c > 0, \qquad \text{and} \qquad ac - b^2 > 0. \tag{2.2}$$

Furthermore, it is assumed that

$$P \stackrel{\text{def}}{=} a_0\frac{\partial^2}{\partial x^2} + c_0\frac{\partial^2}{\partial y^2} \qquad (a_0 > 0,\ c_0 > 0), \tag{2.3}$$

where a_0 and c_0 are two arbitrary positive constants. Using a uniform grid with grid intervals Δx and Δy in the x and the y directions, the central difference forms of (1.1) and (1.2) at a grid point (i, j) are

$$\tilde{Q}(u_{i,j}) = h_{i,j} \tag{2.4}$$

and

$$\tilde{P}(u_{i,j}^{n+1} - u_{i,j}^{n}) = -\tau[\tilde{Q}(u_{i,j}^{n}) - h_{i,j}], \tag{2.5}$$

respectively. Here $h_{i,j}$ is the source term and the finite difference operators $\tilde{Q}$ and $\tilde{P}$, respectively, are defined by

$$\begin{aligned}\tilde{Q}(v_{i,j}) \stackrel{\text{def}}{=}\ & a(\Delta x)^{-2}(v_{i+1,j} + v_{i-1,j} - 2v_{i,j}) + c(\Delta y)^{-2}(v_{i,j+1} + v_{i,j-1} - 2v_{i,j}) \\ & + b(2\Delta x\Delta y)^{-1}(v_{i+1,j+1} + v_{i-1,j-1} - v_{i+1,j-1} - v_{i-1,j+1})\end{aligned} \tag{2.6}$$

and

$$\tilde{P}(v_{i,j}) \stackrel{\text{def}}{=} a_0(\Delta x)^{-2}(v_{i+1,j} + v_{i-1,j} - 2v_{i,j}) + c_0(\Delta y)^{-2}(v_{i,j+1} + v_{i,j-1} - 2v_{i,j}), \tag{2.7}$$

where $v_{i,j}$ is any function of the grid point (i, j). One notes that the operator $\tilde{P}$ is a central difference Poisson operator for a uniform grid with grid intervals $\Delta x/\sqrt{a_0}$ and $\Delta y/\sqrt{c_0}$. Thus it may be inverted using a fast Poisson solver.

Let

$$e_{i,j}^{n} \stackrel{\text{def}}{=} u_{i,j}^{n} - u_{i,j}. \tag{2.8}$$

Then (2.4) and (2.5) imply that

$$\tilde{P}(e_{i,j}^{n+1} - e_{i,j}^{n}) = -\tau \tilde{Q}(e_{i,j}^{n}). \tag{2.9}$$

In Appendix A, the convergence rate of (2.9) is studied using a discrete Fourier analysis. In this analysis, one assumes that

$$e_{i,j}^{n} = e_{i+K,j}^{n} = e_{i,j+L}^{n} \qquad (n = 0, 1, 2,...;\ i, j = 0, \pm 1, \pm 2,...) \tag{2.10}$$

where K ($\geqslant 2$) and L ($\geqslant 2$) are two arbitrary integers. Let

$$\|e^{n}\| \stackrel{\text{def}}{=} \left[\sum_{i=0}^{(K-1)} \sum_{j=0}^{(L-1)} (e_{i,j}^{n})^2 \right]^{1/2} \tag{2.11}$$

$$M^{\infty} \stackrel{\text{def}}{=} \lim_{n \to +\infty} \frac{\|e^{n+1}\|}{\|e^{n}\|} \tag{2.12}$$

$$\hat{a} \stackrel{\text{def}}{=} a/a_{\text{o}} > 0, \qquad \hat{c} \stackrel{\text{def}}{=} c/c_{\text{o}} > 0, \qquad \hat{b} \stackrel{\text{def}}{=} b/\sqrt{a_{\text{o}} c_{\text{o}}} \tag{2.13}$$

$$\lambda_{\max} \stackrel{\text{def}}{=} \tfrac{1}{2}(\hat{a} + \hat{c} + \sqrt{(\hat{a} - \hat{c})^2 + 4(\hat{b})^2}) \tag{2.14}$$

$$\lambda_{\min} \stackrel{\text{def}}{=} \tfrac{1}{2}(\hat{a} + \hat{c} - \sqrt{(\hat{a} - \hat{c})^2 + 4(\hat{b})^2}) \tag{2.15}$$

and

$$\Sigma^{*} \stackrel{\text{def}}{=} \lambda_{\max}/\lambda_{\min}. \tag{2.16}$$

It is shown in Appendix A that

(a) $$M^{\infty} = \lim_{n \to +\infty} \frac{|e_{i,j}^{n+1}|}{|e_{i,j}^{n}|} \qquad (i, j = 0, \pm 1, \pm 2,...), \tag{2.17}$$

(b) the one-step procedure is optimized if

$$\tau = \tau^{*} \stackrel{\text{def}}{=} 2/(\lambda_{\max} + \lambda_{\min}), \tag{2.18}$$

(c) assuming $\tau = \tau^{*}$, then in the limit of $K, L \to +\infty$,

$$M^{\infty} \geqslant G^{*} \stackrel{\text{def}}{=} (\Sigma^{*} - 1)/(\Sigma^{*} + 1) < 1. \tag{2.19}$$

At this juncture, several comments can be made relating to (2.17)–(2.19):

(a) The parameter G^{*} and τ^{*} are independent of the grid intervals Δx and Δy. They are functions of $\hat{a}$, $\hat{b}$, and $\hat{c}$ only.

(b) Since $\lambda_{\max} + \lambda_{\min} = \hat{a} + \hat{c}$

$$\tau^{*} = 2/(\hat{a} + \hat{c}). \tag{2.20}$$

(c) Using (2.13)–(2.16), the parameter G^* can be expressed as a function of a, b, c, and (c_o/a_o). If the coefficients a, b, and c are known, G^* becomes a function of the single variable (c_o/a_o). As shown in [13], this function reaches its minimum

$$G^*_{\min} \stackrel{\text{def}}{=} |b|/\sqrt{ac} \tag{2.21}$$

when

$$c_o/a_o = c/a. \tag{2.22}$$

Next, the two-step procedure defined in (1.3) and (1.4) is studied by assuming that τ is a constant and

$$R \stackrel{\text{def}}{=} a' \frac{\partial^2}{\partial x^2} + 2b' \frac{\partial^2}{\partial x \partial y} + c' \frac{\partial^2}{\partial y^2} \tag{2.23}$$

where a', b', and c' are constant coefficients to be determined later. Furthermore, we assume that the operators Q and P are those defined in (2.1)–(2.3). The central difference forms of (1.3) and (1.4) can be expressed as

$$\tilde{P}(w^n_{i,j}) = -\tau[\tilde{Q}(u^n_{i,j}) - h_{i,j}] \tag{2.24}$$

and

$$\tilde{P}(u^{n+1}_{i,j} - u^n_{i,j}) = \tilde{R}(w^n_{i,j}) \tag{2.25}$$

respectively. Here $\tilde{Q}$ and $\tilde{P}$, respectively, are defined in (2.6) and (2.7). $\tilde{R}(v_{i,j})$, by definition, is equal to the right side of (2.6) except that the coefficients a, b, and c, respectively, are replaced by a', b', and c'.

In Appendix B, the convergence rate of the two-step procedure is also studied using a discrete Fourier analysis. Let the operator R be subjected to the elliptic conditions, i.e.,

$$a' > 0, \qquad c' > 0 \qquad \text{and} \qquad a'c' - (b')^2 > 0. \tag{2.26}$$

Then it is shown that (2.17) is also valid for the two-step procedure. Furthermore, the two-step procedure is optimized (in a sense defined in Appendix B) if one chooses

$$a' = \frac{a_o \hat{c}}{\hat{a}\hat{c} - \hat{b}^2}, \qquad c' = \frac{c_o \hat{a}}{\hat{a}\hat{c} - \hat{b}^2}, \qquad \text{and} \qquad b' = \frac{-\sqrt{a_o c_o}\,\hat{b}}{\hat{a}\hat{c} - \hat{b}^2} \tag{2.27}$$

and

$$\tau = \underline{\tau}^* \stackrel{\text{def}}{=} 2/(\underline{\underline{\Sigma}}^* + 1) \tag{2.28}$$

where

$$\underline{\underline{\Sigma}}^* \stackrel{\text{def}}{=} \frac{(\lambda_{\max} + \lambda_{\min})^2}{4\lambda_{\max}\lambda_{\min}} = \frac{(\hat{a} + \hat{c})^2}{4(\hat{a}\hat{c} - \hat{b}^2)}. \tag{2.29}$$

Assuming (2.27) and (2.28), it is also shown that, in the limit of K, $L \to +\infty$,

$$M^{\infty} = \underline{\underline{G}}^* \stackrel{\text{def}}{=} (\underline{\underline{\Sigma}}^* - 1)/(\underline{\underline{\Sigma}}^* + 1) < 1. \tag{2.30}$$

Note that here, as in the following, the symbol "$\underline{\underline{\ }}$" is used to designate a parameter for the two-step procedure in case this parameter is different from its counterpart in the one-step procedure.

By using (2.16), (2.19), (2.29), and (2.30), it can be shown that

$$\underline{\underline{G}}^* = \rho(G^*) \stackrel{\text{def}}{=} \frac{(G^*)^2}{2 - (G^*)^2}. \tag{2.31}$$

Since

$$\frac{d\rho}{dG^*} = \frac{4G^*}{[2 - (G^*)^2]^2} > 0 \tag{2.32}$$

for $1 > G^* > 0$, one concludes that ρ is a simple monotonically increasing function of G^*. Thus the same value of (c_o/a_o) which optimizes G^* (see (2.21) and (2.22)) will also optimize $\underline{\underline{G}}^*$. In other words, the value of $\underline{\underline{G}}^*$ reaches its minimum

$$\underline{\underline{G}}^*_{\min} \stackrel{\text{def}}{=} \frac{(G^*_{\min})^2}{2 - (G^*_{\min})^2} = \frac{b^2}{2ac - b^2} \tag{2.33}$$

when $c_o/a_o = c/a$.

In the limit of K and $L \to +\infty$, the parameters G^* and $\underline{\underline{G}}^*$ are the asymptotic error multiplication factors for the one- and two-step methods, respectively. Assuming that the execution of one iteration in the two-step procedure requires twice as much computational time as that in the one-step procedure, one may conclude that, asymptotically, the two-step method is faster than the one-step method by a factor of ξ if

$$\underline{\underline{G}}^* = (G^*)^{2\xi} \qquad (1 > G^* > 0).$$

With the aid of (2.31), one obtains

$$\xi = \xi(G^*) \stackrel{\text{def}}{=} 1 - \ln[2 - (G^*)^2]/\ln[(G^*)^2] \qquad (1 > G^* > 0). \tag{2.34}$$

Since (see also Fig. 1)

$$\frac{d\xi(x)}{dx} > 0 \qquad (1 > x > 0) \tag{2.35}$$

and

$$\lim_{x \to 0^+} \xi(x) = 1, \qquad \lim_{x \to 1^-} \xi(x) = 2, \tag{2.36}$$

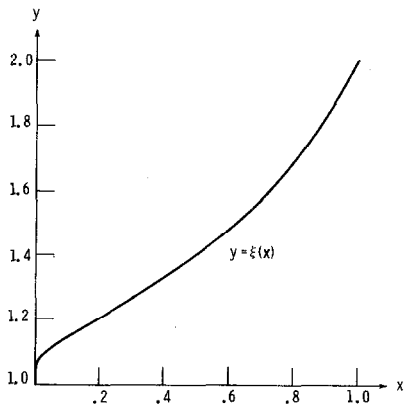

FIG. 1. The function $\xi(x) = 1 - [\ln(2-x^2)]/[\ln(x^2)]$.

the acceleration factor ξ increases from 1 to 2 monotonically as the value of G^* increases from 0 to 1, i.e., the higher value of ξ is realized whenever the need to accelerate the one-step method is greater.

This section is concluded with a discussion on the possible generalization of the 2-D results to a space of higher dimension. In an N-dimensional space ($N \geqslant 2$), (2.1) may be replaced by

$$Q = \sum_{\mu,\nu=1}^{N} \alpha_{\mu\nu} \frac{\partial^2}{\partial x_\mu \partial x_\nu} \tag{2.37}$$

where $\alpha_{\mu\nu}$ are real constants and x_μ the independent variables. Furthermore, the elliptic condition (2.2) is replaced by the requirement that the matrix

$$A \overset{\text{def}}{=} (\alpha_{\mu\nu}) \tag{2.38}$$

be symmetric and positive definite (SPD). Also the operators P and R, respectively, will assume the new forms

$$P \overset{\text{def}}{=} \sum_{\mu=1}^{N} p_\mu \frac{\partial^2}{\partial x_\mu \partial x_\mu} \qquad (p_\mu > 0,\ \mu = 1, 2, 3, \ldots, N) \tag{2.39}$$

and

$$R \overset{\text{def}}{=} \sum_{\mu,\nu=1}^{N} \alpha'_{\mu\nu} \frac{\partial^2}{\partial x_\mu \partial x_\nu} \tag{2.40}$$

where $\alpha'_{\mu\nu}$ are constant coefficients to be determined (see below). With the aid of (2.37)–(2.40) and several theorems given in [14, 15], most of the results given in this section may be generalized for $N \geqslant 2$ using arguments similar to those presented in Appendixes A and B. The exceptions are:

(a) For $N \geqslant 2$, the parameters $\lambda_{\max}$ and $\lambda_{\min}$ are defined, respectively, as the greatest and the smallest eigenvalues of the SPD matrix

$$\hat{A} \stackrel{\text{def}}{=} (\hat{\alpha}_{\mu\nu}), \tag{2.41}$$

where

$$\hat{\alpha}_{\mu\nu} \stackrel{\text{def}}{=} \alpha_{\mu\nu}/\sqrt{p_\mu p_\nu}, \qquad \mu, \nu = 1, 2, \ldots, N. \tag{2.42}$$

(b) For $N \geqslant 2$, Eq. (2.27) should be replaced by

$$\hat{A}' = (\hat{A})^{-1}, \tag{2.43}$$

where

$$\hat{A}' \stackrel{\text{def}}{=} (\hat{\alpha}'_{\mu\nu}) \tag{2.44}$$

with

$$\hat{\alpha}'_{\mu\nu} \stackrel{\text{def}}{=} \alpha'_{\mu\nu}/\sqrt{p_\mu p_\nu}, \qquad \mu, \nu = 1, 2, \ldots, N. \tag{2.45}$$

(c) Equations (2.20)–(2.22) and the second equality sign in (2.29) are valid only for $N = 2$.

3. Local Relaxation

First, the one-step procedure developed in Section 2 is extended to solve PDEs with variable coefficients. To proceed, the operator Q is initially assumed to have the form defined in (2.1) with the understanding that the coefficients a, b, and c are functions of x and y subjected to the condition (2.2).

In the variable coefficient (VC) version of the iterative procedure (2.5), the operator $\tilde{Q}$ will be defined using (2.6) with the understanding that the coefficients a, b, and c, respectively, are replaced by a_{ij}, b_{ij}, and c_{ij}, i.e., the discretized values of a, b, and c at the grid point (i, j). On the other hand, the coefficients a_0 and c_0 associated with the operator $\tilde{P}$ (see (2.7)) are again assumed to be positive constants.

The above definitions of $\tilde{P}$ and $\tilde{Q}$ are directly applicable to any internal grid point. On a periodic boundary, they are also applicable if the periodic conditions are invoked. Similarly, by using an extrapolation technique [16], the operators $\tilde{P}$ and $\tilde{Q}$ can be defined on a Neumann boundary.

The relaxation factor τ, in the VC version, is replaced by its grid point dependent version τ_{ij}. Ideally, the values of τ_{ij}'s may be chosen such that the parameter $M^{(\infty)}$ (see (2.12)) is minimized. Unfortunately, this approach is impractical due to the complexity arising from the variable nature of the coefficients of $\tilde{Q}$ and the necessity to consider the boundary conditions. The alternative adopted in the current study is based on the following heuristic arguments: Recall that the analysis described in

Section 2 and Appendix A is a rigorous von Neumann analysis for (2.9). The results of this analysis are fully justified only under very restricted conditions. However, it is well known that the von Neumann analysis often gives useful results even when its application cannot be fully justified. Particularly, by freezing the variable coefficients at their values at the grid point under consideration, this analysis has been routinely used in the stability study of the numerical procedure solving PDEs with variable coefficients. Due to the above considerations, the VC version of (2.20) is assumed to be

$$\tau_{ij} = 2/(\hat{a}_{ij} + \hat{c}_{ij}), \tag{3.1}$$

where $\hat{a}_{ij} \overset{\text{def}}{=} a_{ij}/a_o > 0$ and $\hat{c}_{ij} \overset{\text{def}}{=} c_{ij}/c_o > 0$.

In view of (2.17) and (2.19), we also assume that

$$\lim_{n \to +\infty} \frac{|e_{i,j}^{n+1}|}{|e_{i,j}^{n}|} \approx G_{ij}, \tag{3.2}$$

where G_{ij} is the local error multiplication factor defined by

$$G_{ij} \overset{\text{def}}{=} (\Sigma_{ij}^* - 1)/(\Sigma_{ij}^* + 1). \tag{3.3}$$

The parameter Σ_{ij}^* is the grid point dependent version of Σ^*. It will be evaluated using (2.13)–(2.16) with the understanding that the coefficients a, b, and c, respectively, are replaced by a_{ij}, b_{ij}, and c_{ij}. Let

$$\|e^n\| \overset{\text{def}}{=} \left[\sum_{(i,j) \in \Phi} (e_{i,j}^n)^2 \right]^{1/2} \tag{3.4}$$

and

$$G^\infty \overset{\text{def}}{=} \sup_{(i,j) \in \Phi} \{G_{ij}\}, \tag{3.5}$$

where Φ denotes the set of (i,j)'s where $u_{i,j}$'s are to be solved. Then, with the aid of (2.12), the assumption (3.2) implies that

$$M^\infty \approx G^\infty. \tag{3.6}$$

Several comments can be made relating to (3.6):

(a) Since G^∞ can be evaluated using the known coefficients a_o, c_o, a_{ij}, b_{ij}, and c_{ij}, the value of M^∞ and thus the convergence rate of the current iterative procedure can be predicted using (3.6).

(b) As long as the coefficients a_{ij}, b_{ij}, and c_{ij} do not vary greatly from one grid point to its neighbors, the value of G^∞ is not sensitive to the grid cell size and aspect ratio. This observation coupled with (3.6) implies that the convergence

behavior of the current numerical procedure, generally, may not be sensitive to the grid cell size and aspect ratio.

(c) The VC version of (2.5) can be expressed in a form in which the coefficients a_0 and c_0 appear only in the ratio (c_0/a_0). As a result, the convergence behavior of the current iterative procedure is dependent on the ratio (c_0/a_0) but not on the individual values of a_0 and c_0. Similarly, one can also show that the parameter G^{α} is dependent on the ratio (c_0/a_0) but not on the individual values of a_0 and c_0. Equation (3.6) suggests that, in order to maximize the convergence rate, the ratio (c_0/a_0) should be chosen such that G^{α} is at its minimum.

The optimal value of (c_0/a_0), generally, can only be evaluated numerically. However, in case that $b_{ij}=0$ for all $(i, j)\in\Phi$, it is shown in Appendix C that G^{α} reaches its minimum

$$G^{\alpha}_{\min} \stackrel{\text{def}}{=} \frac{\sqrt{\beta_{\max}/\beta_{\min}}-1}{\sqrt{\beta_{\max}/\beta_{\min}}+1} \tag{3.7}$$

if and only if

$$c_0/a_0=\sqrt{\beta_{\max}\cdot\beta_{\min}}, \tag{3.8}$$

where

$$\beta_{\max} \stackrel{\text{def}}{=} \sup_{(i,j)\in\Phi}\left\{\frac{c_{ij}}{a_{ij}}\right\} \quad \text{and} \quad \beta_{\min} \stackrel{\text{def}}{=} \inf_{(i,j)\in\Phi}\left\{\frac{c_{ij}}{a_{ij}}\right\} \tag{3.9}$$

may be evaluated either analytically or numerically.

The one-step procedure can be modified to solve a class of self-adjoint PDEs, i.e.,

$$Q=Q^{+}(x, y) \stackrel{\text{def}}{=} \frac{\partial}{\partial x}\left(p(x, y)\frac{\partial}{\partial x}\right)+\frac{\partial}{\partial y}\left(q(x, y)\frac{\partial}{\partial y}\right), \tag{3.10}$$

where p and q are arbitrary positive functions of x and y. A central difference operator $\tilde{Q}^{+}$ corresponding to the differential operator Q^{+} is defined by [17]

$$\begin{aligned}\tilde{Q}^{+}(v_{i,j}) \stackrel{\text{def}}{=} {} & (\Delta x)^{-2}[p_{(i-1/2)j}v_{i-1,j}+p_{(i+1/2)j}v_{i+1,j}-(p_{(i-1/2)j}+p_{(i+1/2)j})v_{i,j}] \\ & +(\Delta y)^{-2}[q_{i(j-1/2)}v_{i,j-1}+q_{i(j+1/2)}v_{i,j+1}-(q_{i(j-1/2)}+q_{i(j+1/2)})v_{i,j}],\end{aligned} \tag{3.11}$$

where

$$p_{(i+1/2)j} \stackrel{\text{def}}{=} p(x_i\pm\Delta x/2, y_j), \qquad q_{i(j+1/2)} \stackrel{\text{def}}{=} q(x_i, y_j\pm\Delta y/2).$$

Assuming the coefficients p and q do not vary greatly from one grid point to its neighbors, then

$$\tilde{Q}^{+}(v_{i,j}) \doteq p_{ij}(\Delta x)^{-2}(v_{i-1,j}+v_{i+1,j}-2v_{i,j})+q_{ij}(\Delta y)^{-2}(v_{i,j-1}+v_{i,j+1}-2v_{i,j}).$$

Thus $\tilde{Q}^+(v_{i,j}) \doteq \tilde{Q}(v_{i,j})$ (see (2.6)) if $a_{ij} = p_{ij}$, $c_{ij} = q_{ij}$ and $b_{ij} = 0$ for all $(i, j) \in \Phi$. This observation coupled with (3.1) leads to the assumption

$$\tau_{ij} = 2/(\hat{p}_{ij} + \hat{q}_{ij}), \tag{3.12}$$

where $\hat{p}_{ij} \stackrel{\text{def}}{=} p_{ij}/a_o$ and $\hat{q}_{ij} \stackrel{\text{def}}{=} q_{ij}/c_o$. Similarly, in case that $Q = Q^+(x, y)$, the parameter G^∞ will be evaluated assuming $a_{ij} = p_{ij}$, $c_{ij} = q_{ij}$, and $b_{ij} = 0$. Also the right sides of (3.7) and (3.8) will be evaluated with

$$\beta_{\max} \stackrel{\text{def}}{=} \sup_{(i,j)\in\Phi} \left\{\frac{q_{ij}}{p_{ij}}\right\} \quad \text{and} \quad \beta_{\min} \stackrel{\text{def}}{=} \inf_{(i,j)\in\Phi} \left\{\frac{q_{ij}}{p_{ij}}\right\}. \tag{3.13}$$

The one-step procedure has been extended to solve PDEs with variable coefficients. Similarly, one can obtain the VC version of the two-step procedure. Thus, with the aid of (2.28) and (2.29), one concludes that (3.1) and (3.12), respectively, may be replaced by

$$\underline{\tau}_{ij} = 2 \cdot \left[1 + \frac{(\hat{a}_{ij} + \hat{c}_{ij})^2}{4\{\hat{a}_{ij}\hat{c}_{ij} - (\hat{b}_{ij})^2\}}\right]^{-1} \tag{3.14}$$

and

$$\underline{\tau}_{ij} = 2 \cdot \left[1 + \frac{(\hat{p}_{ij} + \hat{q}_{ij})^2}{4\hat{p}_{ij}\hat{q}_{ij}}\right]^{-1} \tag{3.15}$$

in the two-step procedure.

Moreover, the VC versions of $\tilde{R}(v_{i,j})$ and (2.27) are obtained by substituting the coefficients a', b', c', $\hat{a}$, $\hat{b}$, and $\hat{c}$, respectively, with a'_{ij}, b'_{ij}, c'_{ij}, $\hat{a}_{ij}$, $\hat{b}_{ij}$, and $\hat{c}_{ij}$. For the case when the operator Q is a self-adjoint operator defined in (3.10), these VC versions are modified by replacing $\hat{a}_{ij}$, $\hat{c}_{ij}$, and $\hat{b}_{ij}$, respectively, with $\hat{p}_{ij}$, $\hat{q}_{ij}$, and 0. It should be cautioned that, for the special case where $\hat{p}_{ij} = \hat{q}_{ij}$,

$$\tilde{R}(w^n_{i,j}) = \tilde{P}(w^n_{i,j})/\hat{p}_{ij}, \qquad \underline{\tau}_{ij} = 1, \qquad \tau_{ij} = 1/\hat{p}_{ij}.$$

Thus the VC versions of (2.24) and (2.25) imply that

$$\tilde{P}(u^{n+1}_{i,j} - u^n_{i,j}) = -\tau_{ij}[\tilde{Q}(u^n_{i,j}) - h_{ij}].$$

In other words, the extra computational effort required for the two-step method is completely wasted since the convergence achieved in one iteration is identical for both the one- and two-step methods. Note that a similar situation also arises in the solution of the PDEs with constant coefficients. Let $\hat{a} = \hat{c}$ and $\hat{b} = 0$. Then it is easy to see that $G^* = \underline{G}^* = 0$. In other words, the machine accuracy solution is obtained in one iteration for both the one- and two-step methods. Since the parameter ξ is ill-defined at $G^* = 0$ (see Eq. (2.34)), the assertion made in Section 2 concerning the advantage of using the two-step method apparently is not valid for this special case.

Let $\underline{\underline{G}}^{\infty}$ be defined using (3.3) and (3.5) with the understanding that G^{∞}, G_{ij}, and Σ^{*}_{ij}, respectively, are replaced by their two-step method counterparts $\underline{\underline{G}}^{\infty}$, $\underline{\underline{G}}_{ij}$, and $\underline{\underline{\Sigma}}^{*}_{ij}$. Then (2.31), (2.32), and (3.6)–(3.9) can be used to show that

(a) $$M^{\infty} \approx \underline{\underline{G}}^{\infty} = \rho(G^{\infty}) \tag{3.16}$$

(b) $\underline{\underline{G}}^{\infty}$ reaches its minimum when G^{∞} reaches its minimum and vice versa.

Thus, for the case in which $b_{ij} = 0$ for all $(i, j) \in \Phi$, $\underline{\underline{G}}^{\infty}$ reaches its minimum

$$\underline{\underline{G}}^{\infty}_{\min} \stackrel{\text{def}}{=} \rho(G^{\infty}_{\min}) \tag{3.17}$$

if and only if $c_o/a_o = \sqrt{\beta_{\max} \cdot \beta_{\min}}$.

The technique of local relaxation has been described for 2-D problems. In a similar fashion, it can be applied to 3-D problems. The value of this technique as a tool to solve PDEs with variable coefficients will be demonstrated in the subsequent sections.

4. Numerical Evaluation

The one- and two-step method were evaluated using PDEs with constant coefficients [18]. The test problems were designed such that the key results of the theoretical development, i.e., (2.17)–(2.22), (2.27)–(2.34), (A.9), (A.19)–(A.26), (B.4), (B.13), and (B.14), can be tested numerically under the most ideal conditions. For several test problems, the numerical results differ from the theoretical predictions only by roundoff errors. In the current evaluation, we concentrate on the PDEs with variable coefficients. To proceed, the following preliminaries are given:

(a) In this section the domain for all numerical problems, except specified otherwise, is assumed to be $1 \geqslant x \geqslant 0$ and $1 \geqslant y \geqslant 0$. Moreover, the operator $\tilde{P}$ is inverted using a Fast Poisson Solver [19].

(b) The convergence rate is evaluated using

$$O_e(n) \stackrel{\text{def}}{=} -\log_{10}\left(\frac{\|e^n\|}{\|e^0\|}\right) \tag{4.1}$$

or

$$O_r(n) \stackrel{\text{def}}{=} -\log_{10}\left(\frac{\|r^n\|}{\|r^0\|}\right), \tag{4.2}$$

where $\|e^n\|$ is the error norm defined in (3.4) while $\|r^n\|$ is the residual norm defined by

$$\|r^n\| \stackrel{\text{def}}{=} \left[\sum_{(i,j) \in \Phi} \{\tilde{Q}(u^n_{i,j}) - h_{i,j}\}^2\right]^{1/2}. \tag{4.3}$$

One notes that the solution $u_{i,j}$ obtained to machine accuracy is used to evaluate $O_e(n)$. Furthermore, since $u^0_{i,j}=0$ for all $(i,j)\in\Phi$ in the current numerical study, $O_e(n)$ can be interpreted as the number of correct digits in $u^n_{i,j}$.

(c) In view of (3.6), the values of $O_e(n)$ and $O_r(n)$ obtained using the one-step method will be predicted using

$$O_t(n)=-n\cdot\log_{10}(G^\infty) \tag{4.4}$$

or its continuous version, i.e.,

$$O_t^*(n)\overset{\text{def}}{=}\lim_{\Delta x,\Delta y\to 0} O_t(n). \tag{4.5}$$

(d) Equation (3.16) combined with (2.31) and (2.34) suggests that the parameter

$$\xi(G^\infty)\overset{\text{def}}{=}1-\ln[2-(G^\infty)^2]/\ln[(G^\infty)^2] \qquad (1>G^\infty>0) \tag{4.6}$$

may be used to predict the numerical acceleration factor

$$\xi_r(n)\overset{\text{def}}{=}\frac{[O_r(n)]_{\text{two-step procedure}}}{[O_r(2n)]_{\text{one-step procedure}}}. \tag{4.7}$$

The evaluation will begin with the one-step method, which is then followed by the two-step method. The first group of PDEs to be studied includes

$$(1+2x^2+2y^2)\frac{\partial^2 u}{\partial x^2}+(1+x^2+y^2)\frac{\partial^2 u}{\partial y^2}=1 \tag{4.8}$$

$$(1+2x^2+2y^2)\frac{\partial^2 u}{\partial x^2}+(1+x^2+y^2)\frac{\partial^2 u}{\partial x\,\partial y}+(1+x^2+y^2)\frac{\partial^2 u}{\partial y^2}=1 \tag{4.9}$$

$$(1+2x^2+2y^2)\frac{\partial^2 u}{\partial x^2}+(1+x^2+y^2)\frac{\partial^2 u}{\partial x\,\partial y}+(1+x^2+y^2)\frac{\partial^2 u}{\partial y^2}$$
$$+[1+3e^{(x^2+y^2)}]\frac{\partial u}{\partial x}+[1+3e^{(x^2+y^2)}]\frac{\partial u}{\partial y}-[1+3e^{(x^2+y^2)}]u=1 \tag{4.10}$$

$$\frac{\partial}{\partial x}\left[\{1+(x^2+y^2)^l\}\frac{\partial u}{\partial x}\right]+\frac{\partial}{\partial y}\left[\{2+(x^2+y^2)^l\}\frac{\partial u}{\partial y}\right]=1 \qquad l=2,4,6,8. \tag{4.11}$$

Fifteen numerical problems associated with the above PDEs are defined in Table I. The parameters MX and MY, respectively, are the numbers of grid intervals in the x and y directions. The other parameter IB specifies the particular set of boundary conditions (see Fig. 2). All these problems are solved assuming $a_o=c_o=1$.

Problems 1–5 are associated with the same PDE (4.8). They differ on the grid cell size, aspect ratio, and boundary conditions. For the one-step method, as shown in

TABLE I

Definitions of Problems 1–15 and Comparisons of Numerical Results and Theoretical Predictions

Problem number	Equation	l	IB	MX	MY	$O_1(n)$	$O_2(n)$	$O_1^*(n)$	$\xi_1(n/2)$	$\xi(G^*)$
1	(4.8)	NA	1	16	16	14.50	12.34	12.04	1.077	1.234
2	(4.8)	NA	1	64	64	13.78	12.11	12.04	1.133	1.237
3	(4.8)	NA	1	64	4	13.96	12.68	12.04	1.144	1.228
4	(4.8)	NA	2	16	16	13.96	12.34	12.04	0.904	1.234
5	(4.8)	NA	3	16	16	13.62	12.18	12.04	1.024	1.236
6	(4.9)	NA	1	16	16	14.43	9.69	9.63	0.962	1.402
7	(4.9)	NA	1	64	64	12.58	9.64	9.63	1.068	1.403
8	(4.9)	NA	1	64	4	19.04	10.01	9.63	0.730	1.394
9	(4.9)	NA	2	16	16	14.50	9.66	9.63	0.977	1.403
10	(4.9)	NA	3	16	16	15.44	9.66	9.63	0.897	1.403
11	(4.10)	NA	1	16	16	13.05	9.69	9.63	0.804	1.402
12	(4.11)	2	1	16	16	17.24	15.27	15.27	0.809	1.289
13	(4.11)	4	1	16	16	16.83	15.27	15.27	0.521	1.289
14	(4.11)	6	1	16	16	13.44	15.27	15.27	0.347	1.289
15	(4.11)	8	1	16	16	7.55	15.27	15.27	0.169	1.289

Note. $O_1(n)$ is obtained using the one-step method. $n = 20$ for Problems 1–5 and $n = 32$ for Problems 6–15.

Table I, the values of either $O_1(20)$ or $O_1^*(20)$ are fairly accurate estimate of $O_2(20)$. Also, as expected from the current theoretical development and the experiences of other researchers [7, 8], the effects of grid cell size and aspect ratio on the convergence rate are minimal. Even the very large aspect ratio (16:1) does not cause any significant reduction in the convergence rate. Furthermore, it is seen that the convergence rate is insensitive to the particular set of boundary conditions used.

Problems 6–10 are associated with (4.9) which differs from (4.8) only in the appearance of a cross derivative term. The numerical results indicate that the convergence rate of the one-step method may be substantially underestimated by the

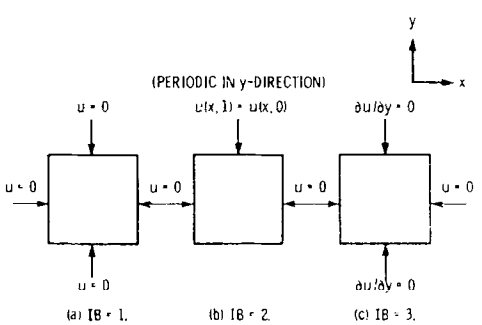

FIG. 2. Three sets of boundary conditions on a unit square.

parameter $O_t(32)$ or $O_t^*(32)$. Furthermore, it is more sensitive to the change of grid cell size and aspect ratio. An explanation for these peculiar behaviors associated with a PDE with a cross derivative term is given in [13].

The success of the one-step method in solving a PDE with a cross derivative term is rather significant. This author is unaware of any earlier work which solves PDEs of this type using a semidirect procedure. The lack of progress in this area may be due to the fact that it is very difficult to choose a separable operator P (by definition, it cannot have a cross derivative term) which closely resembles a non-separable operator Q containing a cross derivative term.

Equation (4.10) contains first-order and zero-order derivative terms. To solve PDEs of this type, we simply add the central difference form of those terms to the term $\tilde{Q}(u_{i,j}^n)$ in (2.5). The value of $O_r(32)$ for Problem 11 indicates that the one-step procedure works very well even though the coefficients of first-order and zero-order derivative terms in (4.10) are several times larger than the second-order terms. This is rather unexpected because the coefficients of lower order terms are completely neglected in the evaluation of the local relaxation factor.

Equation (4.11) belongs to the class of self-adjoint PDEs defined in (3.10). It is seen that the variation of the values of the coefficients p and q increases progressively as one goes from $l=2$ to $l=4$ and so on. For $l=8$, the increase in the values of p and q from one corner $(x=y=0)$ to another corner $(x=y=1)$ on the unit square is of the order of 100 times. It might appear that the technique of local relaxation is no longer valid. The results shown in Table I indicate the one-step method is still useful in this extreme case.

As predicted by (4.4), the $O_r(n)$ vs n curves evaluated using the one-step method are closely approximated by straight lines for the above problems with the exception of perhaps Problems 13–15 (see Fig. 3). For these problems, the linear relation between $O_r(n)$ and n gradually deteriorates as the variation of the coefficients p and q increases. One also notes that the robustness of the one-step method is most

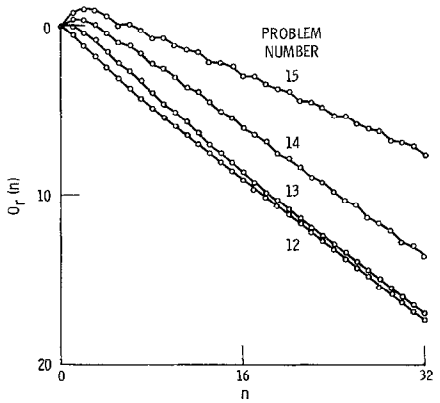

FIG. 3. One-step method convergence histories for Problems 12–15.

TABLE II

Definitions of Problems 16–21 and Comparisons of Numerical Results and Theoretical Predictions

Problem number	Equation	Computational domain	MX	MY	c_o/a_o [a]	$O_e(10)$	$O_r(10)$	$O_t(10)$	$\xi_r(5)$	$\xi(G^\infty)$
16	(4.12)	$1 \geqslant x \geqslant 0,\ 1 \geqslant y \geqslant 0$	16	16	0.423	5.96	5.33	3.92	0.793	1.336
17	(4.12)		64	64	0.379	5.27	4.22	3.47	1.216	1.367
18	(4.13)		16	16	1.0	8.59	8.09	$+\infty$	[b]	[c]
19	(4.13)		64	64	1.0	8.47	7.95	$+\infty$	[b]	[c]
20	(4.12)	$1.5 \geqslant x \geqslant 0,\ 1.5 \geqslant y \geqslant 0$	16	16	0.1239	2.44	1.526	1.083	2.009	1.654
21	(4.12)	$1.5 \geqslant x \geqslant 0,\ 1.5 \geqslant y \geqslant 0$	64	64	0.1065	2.24	1.217	0.928	1.741	1.698

Note. $O_e(10)$ and $O_r(10)$ are obtained using the one-step method.

[a] Evaluated from Eq. (3.8).

[b] Do not compute numerically.

[c] Undefined since $G^\infty = 0$.

evident in its ability to reverse the trend toward divergence during the first few iterations.

The second group of PDEs to be studied includes

$$\frac{\partial}{\partial x}\left[\{1+(x+y)^2\}^2\frac{\partial u}{\partial x}\right]+\frac{\partial}{\partial y}\left[\{1+\sin^2(x+y)\}^2\frac{\partial u}{\partial y}\right]=h_1(x,\,y) \tag{4.12}$$

$$\frac{\partial}{\partial x}\left[\{1+\tfrac{1}{2}(x^4+y^4)\}^2\frac{\partial u}{\partial x}\right]+\frac{\partial}{\partial y}\left[\{1+\tfrac{1}{2}(x^4+y^4)\}^2\frac{\partial u}{\partial y}\right]=h_2(x,\,y), \tag{4.13}$$

where $h_1(x, y)$ and $h_2(x, y)$ are source terms chosen such that

$$u=u_1(x,\,y)\overset{\text{def}}{=}\sin x\sin y \tag{4.14}$$

and

$$u=u_2(x,\,y)\overset{\text{def}}{=}[x(1-x)\,y(1-y)]^2, \tag{4.15}$$

respectively, are the exact solutions of (4.12) and (4.13). Definitions of the four associated numerical problems along with the values of $O_e(10)$, $O_r(10)$, and $O_t(10)$ are given in Table II. The boundary values of u in these problems are specified using (4.14) or (4.15).

Problem 16 is one of the test problems used by Bank [8]. Compared with the

Since the operator $\tilde{P}$ used in Bank's method is a general separable operator, the corresponding FDS code usually must be custom-made and is about five times slower than that for the Laplacian ∇^2 [5]. Thus the current one-step method is easier to use and, for Problem 16, at least a factor of 4 more efficient.

Problem 18 is another test problem used by Bank. Compared with the current value of $O_e(10)=8.59$, the value obtained by Bank is 5.88 without using his scaling technique and 14.79 if a scaling function is used. This problem along with problem 19 was also solved by Concus and Golub [7]. The method of Concus and Golub is also driven by a fast Poisson solver and the results obtained are comparable with ours. However, their method is applicable only when $p=q$ as in the case of (4.13).

The last PDEs to be studied in this section are

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial}{\partial y}\left[\{1+\tfrac{1}{2}(x-y)\}\frac{\partial u}{\partial y}\right]=0 \tag{4.16}$$

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial}{\partial y}\left[\{10+\tfrac{1}{2}(x-y)\}\frac{\partial u}{\partial y}\right]=0. \tag{4.17}$$

An exact solution for both (4.16) and (4.17) is

$$u=y+\tfrac{1}{4}x^2. \tag{4.18}$$

TABLE III

CPU Time Comparisons of Current Methods and SLOR Method

Equation	Solution method	c_o/a_o	ω_o	N_c	T_t (sec)	T_f (sec)
(4.16)	SLOR	NA	1.752	83	3.790	NA
(4.16)	One-step	[a]0.8839	NA	13	1.871	1.345
(4.16)	Two-step	[a]0.8839	NA	6	1.531	1.210
(4.17)	SLOR	NA	1.510	44	2.039	NA
(4.17)	One-step	[a]9.989	NA	6	0.888	0.614
(4.17)	Two-step	[a]9.989	NA	4	1.050	0.815
(4.17)	One-step	1.0	NA	104	14.04	10.32
(4.17)	Two-step	1.0	NA	31	7.61	6.18

[a] Evaluated from Eq. (3.8).

Equations (4.16) and (4.17) were numerically solved using a grid with $MX = MY = 31$. It is assumed that the values of u at all boundary grid points are specified using (4.18). To compare the efficiency of the current one-step method with traditional iterative methods, these numerical problems are solved using the one-step method along with SLOR (successive line over-relaxation). The results of CPU time comparisons are summarized in Table III. Those parameters used in this table which were not defined previously are

(a) $\omega_o \stackrel{\text{def}}{=}$, the optimal value of the relaxation factor used in the SLOR method (determined by repeated numerical experiments);

(b) $N_c \stackrel{\text{def}}{=}$, the smallest value of n which satisfies the convergence criterion

$$(\Delta x)^2 \cdot \|r^n\| < 10^{-8}, \tag{4.19}$$

where $\|r^n\|$ is defined in (4.3) and $\Delta x = 1/31$;

(c) $T_t \stackrel{\text{def}}{=}$, the total CPU time (IBM 370/3033AP) used to satisfy the convergence criterion (4.19);

(d) $T_f \stackrel{\text{def}}{=}$, the CPU time used in the execution of the FDS code.

According to Table III, the total CPU time required for the solution of either (4.16) or (4.17) using SLOR is about twice that using the one-step method. This comparison becomes even more favorable toward the current method if one recalls that the prediction of ω_o is elusive. A small error in this prediction may result in a large increase in the value of T_t. For example, in the solution of (4.16) using SLOR, a change of the value of the relaxation factor from 1.752 to 1.680 resuls in an increase in the value of T_t from 3.790 to 6.565 seconds. On the other hand, as shown in Table IV, the optimal value of (c_o/a_o) evaluated using (3.8) usually is very accurate. Moreover, since the fast Poisson solver [19] currently used is a general purpose code, the value of T_f can be reduced further if the fast Poisson solver is optimized.

TABLE IV

N_c and $O_r(13)$ as Functions of (c_o/a_o) in the Numerical Solution of Eq. (4.16) (One-Step Method)

(c_o/a_o)	N_c	$O_r(13)$
0.8	14	8.3343
[a]0.8839	13	9.4040
0.895	13	9.4891
0.897	13	9.4955
0.899	13	9.4987
[b]0.900	13	9.4990
0.901	13	9.4985
0.903	13	9.4949
0.905	13	9.4882
1.0	14	8.4831

[a] Evaluated from Eq. (3.8).
[b] Actual optimal value.

We complete the evaluation of the one-step procedure by comparing it with a procedure which differs from it only in the use of a constant relaxation factor τ_c. Assuming $a_o = c_o = 1$, problem 16 was solved using different values of τ_c. As shown in Table V, $O_e(10)$ reaches its best value ($\doteq 2.278$) at $\tau_c \doteq 0.103$. Even this best value is substantially below that ($\doteq 3.47$) obtained using the current method ($a_o = c_o = 1$). If we further consider the fact that the accurate prediction of optimal τ_c is by no means easy (see, for example, pp. 964–966 of Ref. [8]), one may con-

TABLE V

$O_e(10)$ as a Function τ_c for Problem 16

τ_c	$O_e(10)$
0.02	0.6
0.06	1.514
0.10	2.259
0.102	2.276
0.103	2.278
0.104	2.272
0.105	2.255
0.11	1.989
0.12	1.2
0.13	[a]

[a] Do not converge if $\tau_c \geqslant 0.13$.

clude that the current procedure has an edge over a procedure using a constant relaxation factor.

The two-step method will be evaluated by comparing it with the one-step method. The initial comparisons involve problems 1-17. The relative efficiencies of these two methods, measured in terms of the parameter $\xi_r(n)$ (see (4.7)), are given in Tables I and II. It is seen that (a) the convergence is accelerated, i.e., $\xi_r(n) > 1$, by the two-step procedure in only six test cases, and (b) without any exception, $\xi_r(n)$ is smaller than the corresponding theoretical parameter $\xi(G^\alpha)$ and the discrepancy is rather large for the test problems with rapidly varying coefficients, i.e., problems 13-15. These disappointing results, however, are not surprising due to the following considerations: (a) the technique of local relaxation is less viable for the two-step procedure and this is particularly true for the test problems with rapidly varying coefficients, and (b) the one-step method convergence rates associated with problems 1-17 are all relatively high and thus (see Fig. 1) the advantage of using the two-step procedure is greatly reduced.

To demonstrate that the two-step method could be substantially faster than the one-step method if the one-step method convergence rate is low, we introduce problems 20 and 21 (see Table II) which, respectively, are the modified versions of problems 16 and 17. The modification involves only the enlargement of the computational domain. As shown in Table II, this results in a large reduction in the values of $O_t(10)$ and thus the one-step method convergence rates. Furthermore, it is seen that the two-step method, as indicated by the values of $\xi_r(5)$, is indeed substantially faster than the one-step method.

Problems 18 and 19 are associated with a self-adjoint PDE (4.13) in which the coefficients p and q are identical. As noted in Section 3, for these test problems, the two-step method is inferior to the one-step method.

For the test problems associated with (4.16) and (4.17), it is shown in Table III that the efficiencies of the one- and two-step methods are about equal if the values of (c_o/a_o) are chosen according to (3.8). If instead, one chooses $c_o/a_o = 1$ for the numerical solution of (4.17), then the convergence rates are sharply reduced and the two-step method is about twice faster than the one-step method.

As a final comment, it is noted that the numerical results and theoretical predictions shown in Tables I and II generally are in better agreement if the values of MX and MY are larger. This is consistent with the fact that all the prediction methods were developed assuming the integers K and L (see (2.10)) are infinitely large.

5. Application to a 3-D Flow Problem

In this section, the current semidirect procedures are incorporated into an Euler solver [9, 10] to obtain the inviscid solution for 3-D steady incompressible rotational flow in a 180-degree turning channel (Fig. 4).

The Euler solver is formed by the inner and the outer loops. The inner loop solves the elliptic equations while the outer loop solves the hyperbolic equations. In

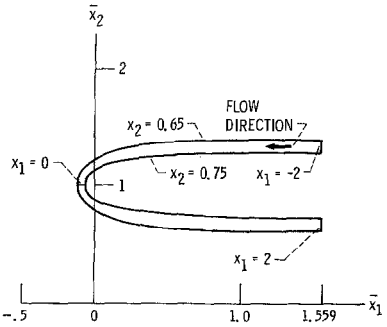

FIG. 4. A converging–diverging turning channel ($\bar{x}_3$ is suppressed).

each pass through the inner loop, the velocity field $\mathbf{V}$ is updated to satisfy the continuity equation

$$\nabla \cdot \mathbf{V} = 0 \tag{5.1}$$

and the velocity–vorticitiy relation

$$\nabla \times \mathbf{V} = \boldsymbol{\Omega}, \tag{5.2}$$

where $\boldsymbol{\Omega}$ is a known divergence free (i.e., $\nabla \cdot \boldsymbol{\Omega} = 0$) vorticity field. In the current study, a solution procedure different from that described in [9, 10] is used to solve (5.1) and (5.2). It is noted that the general solution of (5.1) and (5.2) can be expressed as

$$\mathbf{V} = \mathbf{V}_s + \nabla u \tag{5.3}$$

where $\mathbf{V}_s$ is any special solution of (5.2) and u is a solution of

$$\nabla^2 u = -\nabla \cdot \mathbf{V}_s. \tag{5.4}$$

As a result, once a special solution $\mathbf{V}_s$ is obtained [9], the solution of (5.4) becomes the focal point of the inner loop calculations.

Let the coordinates $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ refer to physical space and (x_1, x_2, x_3) to computational space. It is shown in [10] that the turning channel in Fig. 4 is a mapping of a parallelepiped $(2 \geqslant x_1 \geqslant -2,\ 0.75 \geqslant x_2 \geqslant 0.65,\ 0.1 \geqslant x_3 \geqslant 0)$ in computational space. In physical space, (5.4) is a Poisson equation. However, it cannot be solved using the current procedure since the physical domain is not a parallelepiped. On the other hand, in computational space, the domain is a parallelepiped and (5.4) becomes

$$\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \eta \frac{\partial^2 u}{\partial x_3^2} = -\left(\frac{\partial V_{s,1}}{\partial x_1} + \frac{\partial V_{s,2}}{\partial x_2} + \eta \frac{\partial V_{s,3}}{\partial x_3}\right) \tag{5.5}$$

where $V_{s,1}$, $V_{s,2}$ and $V_{s,3}$ are the covariant components of the known vector field $\mathbf{V}_s$ and

$$\eta = \eta(x_1, x_2) \stackrel{\text{def}}{=} \frac{\cosh(\pi x_1) + \cos(\pi x_2)}{\cosh(\pi x_1) - \cos(\pi x_2)}. \tag{5.6}$$

The 3-D operator Q associated with (5.5) is a special case of that defined in (2.37). Let the parameter G^{∞} and the set Φ for 3-D problems be defined similar to their counterparts for 2-D problems. Using a line of arguments similar to that presented in Appendix C, it can be shown that the parameter G^{∞} reaches its minimum

$$G^{\infty}_{\min} \stackrel{\text{def}}{=} \frac{\sqrt{\eta_{\max}/\eta_{\min}} - 1}{\sqrt{\eta_{\max}/\eta_{\min}} + 1} \tag{5.7}$$

if and only if the coefficients of the operator P (see (2.39)) are chosen such that

$$\frac{p_3}{p_1} = \frac{p_3}{p_2} = \sqrt{\eta_{\max} \cdot \eta_{\min}}. \tag{5.8}$$

Here $\eta_{\max}$ and $\eta_{\min}$, respectively, are the maximum and minimum of the function η in Φ.

In a successful effort to obtain a secondary flow solution in the turning channel, Eq. (5.5) was solved once during each of 25 passes through the inner loop. (Note: the source terms on the right side of (5.5) vary from one pass to another.) The cen-

the x_1 direction and 12 uniform intervals in both the x_2 and x_3 directions. It is assumed that the normal derivative of u vanishes at all boundaries except at the exit plane ($x_1 = 2$), where $u = 0$. Thus $\eta_{\max} = \eta(-2, 0.65) \doteq 0.9966$ and $\eta_{\min} = \eta(0, 0.75) = 0.1716$. As suggested by (5.8), (5.5) was solved assuming $P_1 = P_2 = 1$ and $P_3 = 0.4135$. The values of $O_r(8)$ obtained using the one-step method range from 3.69 to 4.48. All are higher than the value of $O_t(8) \doteq 3.07$ as evaluated from (5.7). Furthermore, the two-step method is consistently faster than the one-step method as indicated by the values of $\xi_r(4)$ which range from 1.140 to 1.308, with the average being 1.256. The convergence rates achieved using the current procedures are dramatic improvements over those reported in [10].

6. Conclusions and Comments on Future Developments

Efficient semidirect procedures have been developed using a mix of rigorous analyses and heuristic arguments. A cornerstone of the current development is the observation that, generally, the coefficients of a PDE vary only slightly across a grid interval. With the aid of a local discrete Fourier analysis, this observation is

used to develop the current local relaxation methods and the corresponding convergence rate prediction methods. The merit of the present approach was demonstrated using numerous numerical examples.

The current methods are similar to the multigrid methods [20] in one aspect: Both take advantage of the existence of an underlying continuous problem in their developments. Particularly, both use the local Fourier analysis as a means to estimate the convergence rate of the overall iterative procedure. However, the current methods are unique in their ability to annihilate the errors of both short and long wavelengths efficiently on the same grid.

The local Fourier analysis handles variable coefficient problems by freezing the coefficients at their local values and ignoring the boundaries. Heuristically, one may argue that the local Fourier analysis is valid only for the errors whose wavelengths are much shorter than those of the coefficients and it can not be applied in the neighborhood of any boundary. However, the results of the current numerical study indicate that, as a means of analyzing the current procedures, the local Fourier analysis may have a much wider realm of applicability than that expected from the heuristic arguments.

The current algorithms are extermely easy to implement. In applying the one-step method, a user simply evaluates the terms on the right side of (2.5) and uses them as the input for a fast Poisson solver. He is not required to deal with a large sparse matrix as in the case of a traditional iterative procedure. The two-step method can also be implemented with similar ease.

For both the one- and two-step procedures, the convergence rate can be accelerated by optimizing the coefficients of the finite difference operator $\tilde{P}$. It is shown that this optimization can be carried out easily for a class of self-adjoint elliptic PDEs.

It is also shown that the convergence rate of the current procedures is relatively insensitive to the grid cell size and aspect ratio. The underlying reason for this insensitivity is the existence of the uniform bounds $\lambda_{\max}$, $\lambda_{\min}$, and $\underline{\underline{\Sigma}}^*$ (see (A.25), (A.26), (B.13), and (B.14)) which are independent of the grid intervals Δx and Δy. The existence of these bounds, as shown in Appendixes A and B, can be attributed to the fact that $\tilde{P}$, $\tilde{Q}$, and $\tilde{R}$ in (2.4), (2.5), (2.24), and (2.25) are similar in one aspect, i.e., they are all central difference elliptic operators involving second-order derivatives.

The two-step procedure is developed to accelerate the one-step procedure. A key element in this development is to choose the coefficients of $\tilde{R}$ such that the convergence rate can be maximized. This optimization problem is solved by (2.27). With the aid of (2.27), a simple relation, i.e., (2.31), is found to exist between the asymptotic error multiplification factors for the one- and two-step procedures. For a constant coefficient model problem, this relation can be used to establish the superiority of the two-step procedure over the one-step procedure. This superiority, generally, carries over to the variable coefficient problems if the one-step method convergence rate is low (i.e., $\xi(G^\infty) \gtrsim 1.5$) and the coefficients do not vary as rapidly as in the case of (4.11).

As for future developments, several comments can be made:

1. The development of the current algorithms is greatly simplified by ignoring the effects of the boundary conditions and by omitting the first- and zero-order derivative terms in the operator $\tilde{Q}$. Although the numerical results indicate that the applicability of the current algorithms is not limited by these simplifying assumptions, they should be further addressed in future development.

2. For the PDEs without a cross derivative term, the parameters $G_{\min}^{x}$ and $G_{\min}^{y}$, respectively, are given by (3.7) and (3.17). As a result, a smaller value of $(\beta_{\max}/\beta_{\min})$ leads to a larger convergence rate. Let $\beta'_{\max}$ and $\beta'_{\min}$ be defined using (3.9) (or (3.13) if $Q = Q^{+}(x, y)$) with the understanding that $\beta_{\max}$, $\beta_{\min}$, and Φ, respectively, are replaced by $\beta'_{\max}$, $\beta'_{\min}$, and Φ' where Φ' is a subdomain of Φ. Since $(\beta_{\max}/\beta_{\min}) \geq (\beta'_{\max}/\beta'_{\min})$, the convergence rate, generally, will be larger if the computational domain is Φ' instead of Φ. The demonstration of this principle is provided by Problems 16, 17, 20, and 21 of Section 4 (see Table II). Note that the value of (c_o/a_o) determined using (3.8) is a function of the computational domain under consideration.

The above observation suggests that the current algorithms may be accelerated by using the domain decomposition methods [21–23]. In these methods, the original computational domain is divided into several overlapping or nonoverlapping subdomains and the overall iterative procedure consists of the separate iterations performed over these subdomains. If the current one-step (two-step) method is used in these iterations with the understanding that different optimal values of (c_o/a_o) are used in different subdomains, the subdomain iteration convergence rates will be higher than the convergence rates associated with the original domain. Thus an iterative procedure formed using a subdomain decomposition method may have a higher convergence rate than an iterative procedure in which only one value of (c_o/a_o) is used in the entire computational domain.

Other advantages of using the domain decomposition method are: (i) the subdomain iterations may be executed simultaneously using parallel computation, (ii) the computational domain can be any shape as long as it is the union of rectangular subdomains, and (iii) the grid cell sizes and aspect ratios may vary from one subdomain to another.

3. The use of splitting matrices involving a fast Poisson solver limits the applicability of the current procedures with respect to the geometry of the domain, the type of discretization and the boundary conditions. It also imposes a constraint on the possible forms of the splitting matrices and thus the types of PDEs which can be solved efficiently using the current procedures. As a result, it is imperative to find other types of splitting matrices which can be inverted easily under less restrictive conditions and, at the same time, do not upset the basic characteristics of the current methods. The success of this effort may require the use of several concepts which were used or proposed, e.g., the local Fourier analysis and the subdomain decomposition.

4. The numerical study of a semidirect procedure [10] indicates that the current procedure may also be used to solve a second-order quasi-linear elliptic PDE. Since the coefficients of this PDE are functions of the dependent variables and their derivatives, the local relaxation factor must be updated during each iteration for this type of applicaton.

Appendix A

The mathematical foundation for the one-step method is established in this appendix. We begin with

Theorem 1. *Given a set of* $e_{i,j}^0$'s *which satisfies the periodic condition* (2.10), *one has*:

(a) $e_{i,j}^n$'s $(n=1, 2,...;\ i, j=0, \pm 1, \pm 2,...)$ *are uniquely determined by* (2.9), (2.10), *and the auxiliary conditions*

$$\sum_{i=0}^{(K-1)} \sum_{j=0}^{(L-1)} e_{i,j}^n = 0 \qquad (n=1, 2,...). \tag{A.1}$$

(b) *Let*

$$\varphi_{i,j}^{(k,l)} \stackrel{\text{def}}{=} \frac{1}{\sqrt{KL}} \exp\left[2\pi I\left(\frac{k\cdot i}{K}+\frac{l\cdot j}{L}\right)\right], \qquad I\equiv\sqrt{-1}$$

$$(i, j=0, \pm 1, \pm 2,...; k=0, 1, 2,..., (K-1); l=0, 1, 2,..., (L-1)) \tag{A.2}$$

$$E^{0,(k,l)} \stackrel{\text{def}}{=} \sum_{i=0}^{(K-1)} \sum_{j=0}^{(L-1)} e_{i,j}^0 \overline{\varphi_{i,j}^{(k,l)}}, \qquad (k, l)\in \Psi \tag{A.3}$$

$$\sigma_p^{(k,l)} \stackrel{\text{def}}{=} -4\left[a_0\left\{\frac{1}{\Delta x}\sin(\pi k/K)\right\}^2 + c_0\left\{\frac{1}{\Delta y}\sin(\pi l/L)\right\}^2\right]$$

$$(k=0, 1, 2,... (K-1); l=0, 1, 2,..., (L-1)) \tag{A.4}$$

$$\sigma_q^{(k,l)} \stackrel{\text{def}}{=} -4\left[a\left\{\frac{1}{\Delta x}\sin(\pi k/K)\right\}^2 + c\left\{\frac{1}{\Delta y}\sin(\pi l/L)\right\}^2 + 2b\left\{\frac{1}{\Delta x}\sin(\pi k/K)\right\}\left\{\frac{1}{\Delta y}\sin(\pi l/L)\right\}\cos(\pi k/K)\cos(\pi l/L)\right]$$

$$(k=0, 1, 2,..., (K-1); l=0, 1, 2,..., (L-1)) \tag{A.5}$$

$$\gamma^{(k,l)} \stackrel{\text{def}}{=} \sigma_q^{(k,l)}/\sigma_p^{(k,l)} \qquad (k, l)\in \Psi \tag{A.6}$$

and

$$G^{(k,l)}(\tau) \stackrel{\text{def}}{=} 1-\tau\gamma^{(k,l)} \qquad (k, l)\in \Psi, \tag{A.7}$$

where $\in \overset{\text{def}}{=}$ "*belongs to*" *and* Ψ *is the set of ordered pairs defined by*

$$\Psi \overset{\text{def}}{=} \{(k, l) | k = 0, 1, 2, \ldots, (K-1); l = 0, 1, 2, \ldots, (L-1); (k, l) \neq (0, 0)\}. \quad \text{(A.8)}$$

Then the unique solution referred to in (a) *is explicitly given by*

$$e_{i,j}^{n} = \sum_{(k,l) \in \Psi} [G^{(k,l)}(\tau)]^{n} \cdot E^{0,(k,l)} \cdot \varphi_{i,j}^{(k,l)}. \quad \text{(A.9)}$$

Proof. Note that $\varphi_{i,j}^{(k,l)}$'s are periodic and orthonormal, i.e.,

$$\varphi_{i,j}^{(k,l)} = \varphi_{i+K,j}^{(k,l)} = \varphi_{i,j+L}^{(k,l)} \qquad (i, j = 0, \pm 1, \pm 2, \ldots) \quad \text{(A.10)}$$

and

$$\sum_{i=0}^{(K-1)} \sum_{j=0}^{(L-1)} \varphi_{i,j}^{(k,l)} \overline{\varphi_{i,j}^{(k',l')}} = \delta_{kk'} \delta_{ll'}$$
$$(k, k' = 0, 1, 2, \ldots, (K-1); l, l' = 0, 1, 2, \ldots, (L-1)), \quad \text{(A.11)}$$

where $\delta_{kk'}$ is the Kronecker delta symbol. Also, it can be shown that

$$\tilde{P}(\varphi_{i,j}^{(k,l)}) = \sigma_{p}^{(k,l)} \varphi_{i,j}^{(k,l)} \quad \text{(A.12)}$$

and

$$\tilde{Q}(\varphi_{i,j}^{(k,l)}) = \sigma_{q}^{(k,l)} \varphi_{i,j}^{(k,l)}, \quad \text{(A.13)}$$

where $\tilde{P}$ and $\tilde{Q}$ are defined in Section 2.

Equations (2.10), (A.10), and (A.11) imply that

$$e_{i,j}^{n} = \sum_{k=0}^{(K-1)} \sum_{l=0}^{(L-1)} E^{n,(k,l)} \varphi_{i,j}^{(k,l)}$$
$$(n = 0, 1, 2, \ldots; i, j = 0, \pm 1, \pm 2, \ldots), \quad \text{(A.14)}$$

where

$$E^{n,(k,l)} \overset{\text{def}}{=} \sum_{i=0}^{(K-1)} \sum_{j=0}^{(L-1)} e_{i,j}^{n} \overline{\varphi_{i,j}^{(k,l)}}. \quad \text{(A.15)}$$

Substituting (A.14) into (2.9) and using (A.11)–(A.13), one obtains

$$(E^{n+1,(k,l)} - E^{n,(k,l)}) \sigma_{p}^{(k,l)} = -\tau E^{n,(k,l)} \sigma_{q}^{(k,l)}$$
$$(n = 0, 1, 2, \ldots; k = 0, 1, 2, \ldots, (K-1); l = 0, 1, 2, \ldots, (L-1)). \quad \text{(A.16)}$$

For $k = l = 0$, (A.16) is an identity since $\sigma_{p}^{(0,0)} = \sigma_{q}^{(0,0)} = 0$. On the other hand, since $\sigma_{p}^{(k,l)} < 0$ if $(k, l) \in \Psi$, (A.16) implies that

$$E^{n+1,(k,l)} = [G^{(k,l)}(\tau)] \cdot E^{n,(k,l)} \qquad (k, l) \in \Psi. \quad \text{(A.17)}$$

Equation (A.9) is an immediate result of (A.14), (A.17), and the assumption

$$E^{n,(0,0)} = 0 \qquad (n = 1, 2, 3, \ldots). \tag{A.18}$$

It should be noted that (A.18) is introduced to insure the uniqueness of the solution. Using (A.15) and the fact that $\varphi_{i,j}^{(0,0)} = 1/\sqrt{KL}$ for $i, j = 0, \pm 1, \pm 2, \ldots$, it can be shown that (A.18) is equivalent to (A.1). Q.E.D.

According to (A.9), $e_{i,j}^n$ is a sum of $(K \times L - 1)$ terms and each term is multiplied by the factor $G^{(k,l)}(\tau)$ as the iteration number n increased by one. Obviously, the term with the greatest value of $|G^{(k,l)}(\tau)|$ eventually will become dominant assuming the corresponding $E^{0,(k,l)}$ does not vanish. Thus, assuming every $E^{0,(k,l)} \neq 0$, one may conclude that

$$M^{\infty} = \lim_{n \to +\infty} \frac{|e_{i,j}^{n+1}|}{|e_{i,j}^n|} = G(\tau) \qquad (i, j = 0, \pm 1, \pm 2, \ldots), \tag{A.19}$$

where M^{∞} is defined in (2.12) and

$$G(\tau) \stackrel{\text{def}}{=} \sup_{(k,l) \in \Psi} \{|G^{(k,l)}(\tau)|\}. \tag{A.20}$$

A direct implication of (A.19) is that the value of M^{∞} reaches its minimum if the parameter τ is chosen such that the function $G(\tau)$ is at its minimum. Let

$$\gamma_{\max} \stackrel{\text{def}}{=} \sup_{(k,l) \in \Psi} \{\gamma^{(k,l)}\}, \qquad \gamma_{\min} \stackrel{\text{def}}{=} \inf_{(k,l) \in \Psi} \{\gamma^{(k,l)}\} \tag{A.21}$$

and use the fact that $\gamma_{\max} \geqslant \gamma_{\min} > 0$ (see (A.25)). One concludes from (A.7) that $G(\tau)$ reaches its minimum

$$G(\tau^0) = (\Sigma - 1)/(\Sigma + 1) < 1 \tag{A.22}$$

when

$$\tau = \tau^0 \stackrel{\text{def}}{=} 2/(\gamma_{\max} + \gamma_{\min}) \tag{A.23}$$

where

$$\Sigma \stackrel{\text{def}}{=} \gamma_{\max}/\gamma_{\min}. \tag{A.24}$$

Combining (A.19) and (A.22), one concludes that (a) $M^{\infty} < 1$ if $\tau = \tau^0$, and (b) M^{∞} increases with an increase of Σ.

The values of $\gamma_{\max}$ and $\gamma_{\min}$, generally, are functions of the integers K and L. However, as will be shown,

$$\lambda_{\max} \geqslant \gamma_{\max} \geqslant \gamma_{\min} \geqslant \lambda_{\min} > 0 \tag{A.25}$$

and

$$\lim_{K,L \to +\infty} \gamma_{\max} = \lambda_{\max}, \qquad \lim_{K,L \to +\infty} \gamma_{\min} = \lambda_{\min}. \tag{A.26}$$

Expressions (A.25) and (A.26) imply that Σ, τ^0, and $G(\tau^0)$, respectively, approach Σ^*, τ^*, and G^* as K and L approach $+\infty$. Thus the statements (a) to (c) made following (2.16) are the results of (A.19) and (A.22)–(A.26).

To prove expressions (A.25) and (A.26), one notes that

$$\gamma^{(k,l)} = \hat{a}(s_x)^2 + \hat{c}(s_y)^2 + 2\hat{b}s_x s_y t_x t_y, \qquad (k, l) \in \Psi, \tag{A.27}$$

where

$$s_x \stackrel{\text{def}}{=} \frac{(\sqrt{a_0}/\Delta x)\sin(\pi k/K)}{\sqrt{[(\sqrt{a_0}/\Delta x)\sin(\pi k/K)]^2 + [(\sqrt{c_0}/\Delta y)\sin(\pi l/L)]^2}}, \qquad (k, l) \in \Psi \tag{A.28}$$

$$s_y \stackrel{\text{def}}{=} \frac{(\sqrt{c_0}/\Delta y)\sin(\pi l/L)}{\sqrt{[(\sqrt{a_0}/\Delta x)\sin(\pi k/K)]^2 + [(\sqrt{c_0}/\Delta y)\sin(\pi l/L)]^2}} \qquad (k, l) \in \Psi \tag{A.29}$$

$$t_x \stackrel{\text{def}}{=} \cos(\pi k/K) \qquad (k = 0, 1, 2, \ldots, (K-1)) \tag{A.30}$$

and

$$t_y \stackrel{\text{def}}{=} \cos(\pi l/L) \qquad (l = 0, 1, 2, \ldots, (L-1)). \tag{A.31}$$

Using (A.28)–(A.31), one concludes that (a) $(s_x)^2 + (s_y)^2 = 1$, and (b) $(t_x)^2 \leqslant 1$, $(t_y)^2 \leqslant 1$. As a result, Theorem 1 in [14] implies that

$$\lambda_{\max} \geqslant \gamma^{(k,l)} \geqslant \lambda_{\min} > 0 \qquad (k, l) \in \Psi. \tag{A.32}$$

Inequality (A.25) follows immediately from (A.32).

Equation (A.26) can also be shown using (A.27)–(A.31) and Theorem 1 in [14]. The proof is rather lengthy and the full details are given in [13].

Appendix B

The mathematical foundation for the two-step method is developed in this appendix. With the understanding that the notations defined previously are inherited here, we define

$$\gamma'^{(k,l)} \stackrel{\text{def}}{=} (\sigma_r^{(k,l)}/\sigma_p^{(k,l)}) \qquad (k, l) \in \Psi \tag{B.1}$$

$$\underline{\gamma}^{(k,l)} \stackrel{\text{def}}{=} \gamma^{(k,l)} \cdot \gamma'^{(k,l)} \qquad (k, l) \in \Psi \tag{B.2}$$

$$\underline{G}^{(k,l)}(\tau) \stackrel{\text{def}}{=} 1 - \tau\underline{\gamma}^{(k,l)} \qquad (k, l) \in \Psi, \tag{B.3}$$

where, by definition, $\sigma_r^{(k,l)}$ is equal to the right side of (A.5) except that the coefficients a, b, and c, respectively, are replaced by a', b', and c'. Let both $w_{i,j}^n$'s and

$e_{i,j}^n$'s satisfy the periodic and uniqueness conditions as given in (2.10) and (A.1). Then a line of arguments similar to that presented in Appendix A can be used to show that the two-step method counterpart of (A.9) is

$$e_{i,j}^n = \sum_{(k,l)\in\Psi} [\underline{\underline{G}}^{(k,l)}(\tau)]^n \cdot E^{0,(k,l)} \cdot \varphi_{i,j}^{(k,l)}$$
$$(n = 1, 2, 3,...; i, j = 0, \pm 1, \pm 2,...). \tag{B.4}$$

Note that (B.3) and (B.4), respectively, are identical to (A.7) and (A.9) except that $\gamma^{(k,l)}$ and $G^{(k,l)}(\tau)$, respectively, are replaced by $\underline{\gamma}^{(k,l)}$ and $\underline{\underline{G}}^{(k,l)}(\tau)$. As a result, (A.19)–(A.24) and the statements made relating to them can be repeated for the two-step procedure if (a) the parameters $G(\tau)$, $G^{(k,l)}(\tau)$, $\gamma^{(k,l)}$, $\gamma_{\max}$, $\gamma_{\min}$, τ^0, $G(\tau^0)$, and Σ are replaced by $\underline{\underline{G}}(\tau)$, $\underline{\underline{G}}^{(k,l)}(\tau)$, $\underline{\gamma}^{(k,l)}$, $\underline{\gamma}_{\max}$, $\underline{\gamma}_{\min}$, $\underline{\tau}^0$, $\underline{\underline{G}}(\underline{\tau}^0)$, and $\underline{\underline{\Sigma}}$ and (b)

$$\underline{\gamma}_{\max} \geqslant \underline{\gamma}_{\min} > 0. \tag{B.5}$$

At this juncture, it should be noted that the difference between a one-step method parameter and its two-step method counterpart can be traced back to the difference in the definition of $\gamma^{(k,l)}$ and $\underline{\gamma}^{(k,l)}$ (see (B.2)). This difference, essentially, is all that separates the analysis of the one-step method from that of the two-step method.

To prove (B.5), we define $\hat{a}'$, $\hat{b}'$, $\hat{c}'$, $\lambda'_{\max}$, and $\lambda'_{\min}$ using (2.13)–(2.15) with the understanding that the parameters a, b, c, $\hat{a}$, $\hat{b}$, $\hat{c}$, $\lambda_{\max}$, and $\lambda_{\min}$ are replaced by a', b', c', $\hat{a}'$, $\hat{b}'$, $\hat{c}'$, $\lambda'_{\max}$, and $\lambda'_{\min}$. A result similar to (A.32) is

$$\lambda'_{\max} \geqslant \gamma'^{(k,l)} \geqslant \lambda'_{\min} > 0 \qquad (k, l) \in \Psi. \tag{B.6}$$

Combining, (A.32), (B.2), and (B.6), one concludes that

$$\lambda_{\max} \cdot \lambda'_{\max} \geqslant \underline{\gamma}_{\max} \geqslant \underline{\gamma}_{\min} \geqslant \lambda_{\min} \cdot \lambda'_{\min} > 0. \tag{B.7}$$

Inequality (B.5) follows directly from (B.7).

If the coefficients a, b, c, a_o, c_o, the aspect ratio ($\Delta y/\Delta x$), and the integers K and L are known, the parameter $\underline{\underline{\Sigma}}$ is a function of the coefficients a', b', and c'. Since the two-step method convergence rate increases with a decrease of $\underline{\underline{\Sigma}}$, ideally, the coefficients a', b', and c' should be chosen such that $\underline{\underline{\Sigma}}$ is at its minimum. Unfortunately, this optimization problem is too complicated to be solved by a simple analytical technique. In the current study, as will be shown, a simplified version of this optimization problem is solved.

To proceed, note that a direct result of (B.7) is

$$(\lambda_{\max} \cdot \lambda'_{\max})/(\lambda_{\min} \cdot \lambda'_{\min}) \geqslant \underline{\underline{\Sigma}} \stackrel{\text{def}}{=} \underline{\gamma}_{\max}/\underline{\gamma}_{\min} \geqslant 1. \tag{B.8}$$

Since the upper bound of $\underline{\underline{\Sigma}}$ given in (B.8) is independent of the integers K and L, the existence of

$$\underline{\underline{\Sigma}}^{\infty} \stackrel{\text{def}}{=} \sup_{K\geqslant 2, L\geqslant 2} \{\underline{\underline{\Sigma}}\} \tag{B.9}$$

is guaranteed. With the parameter $\underline{\underline{\Sigma}}^{\infty}$ being defined, the simplified optimization problem can be stated as follows: Given a, b, c, a_o, c_o, and $(\Delta y/\Delta x)$, find the values of a', b', and c' (subjected to the condition (2.26)) which minimize $\underline{\underline{\Sigma}}^{\infty}$. The complete solution of this problem is given as a part of

THEOREM 2. *Let a_o, b_o, a, b, c, a', b', c' be subjected to the conditions specified previously. Then*

(a)
$$\underline{\underline{\Sigma}}^{\infty} \geqslant \underline{\underline{\Sigma}}^{*}, \tag{B.10}$$

where $\underline{\underline{\Sigma}}^{*}$ is the parameter defined in (2.29) *and is independent of a', b', c', Δx, and Δy. Furthermore, the equality sign in* (B.10) *is valid if and only if*

$$\hat{D}\hat{D}' = \Lambda I_2 \qquad (\Lambda > 0), \tag{B.11}$$

where Λ is an arbitrary positive scalar, I_2 the 2×2 identity matrix, and

$$\hat{D} \overset{\text{def}}{=} \begin{pmatrix} \hat{a} & \hat{b} \\ \hat{b} & \hat{c} \end{pmatrix} \qquad \text{and} \qquad \hat{D}' \overset{\text{def}}{=} \begin{pmatrix} \hat{a}' & \hat{b}' \\ \hat{b}' & \hat{c}' \end{pmatrix}. \tag{B.12}$$

(b) *Equation* (B.11) *is equivalent to* (2.27) *if $\Lambda = 1$.*

(c) *With the coefficients a', b', and c' specified according to* (2.27)

$$\underline{\underline{\Sigma}}^{*} \geqslant \underline{\gamma}_{\max} \geqslant \underline{\gamma}_{\min} \geqslant 1 \tag{B.13}$$

and

$$\lim_{K,L \to +\infty} \underline{\gamma}_{\max} = \underline{\underline{\Sigma}}^{*}, \qquad \lim_{K,L \to \infty} \underline{\gamma}_{\min} = 1. \tag{B.14}$$

The Proof of Theorem 2 is given in [18].

Using (B.14), it can be shown that, in the limit of K, $L \to +\infty$, $\underline{\underline{\Sigma}}$, $\underline{\tau}^0$, and $\underline{\underline{G}}(\underline{\tau}^0)$, respectively, approach $\underline{\underline{\Sigma}}^{*}$, $\underline{\tau}^{*}$, and $\underline{\underline{G}}^{*}$. As a result, (2.27)–(2.30) and the statements made relating to them can be justified with the aid of Theorem 2 and the two-step method version of (A.19).

APPENDIX C

To obtain the results given in (3.7) and (3.8), we introduce the parameters $\alpha_o \overset{\text{def}}{=} c_o/a_o > 0$ and $\beta_{ij} \overset{\text{def}}{=} c_{ij}/a_{ij} > 0$. Assuming $b_{ij} = 0$ for all $(i, j) \in \Phi$, (2.13)–(2.16), and (3.3) can be used to show that

$$G_{ij} = J(\alpha_o/\beta_{ij}), \tag{C.1}$$

where J is a function defined by

$$J(t) \stackrel{\text{def}}{=} \begin{cases} \dfrac{t-1}{t+1} & \text{if} \quad t \geqslant 1 \\ \dfrac{1-t}{t+1} & \text{if} \quad 1 \geqslant t > 0. \end{cases} \tag{C.2}$$

Note that

$$\frac{dJ(t)}{dt} = \begin{cases} \dfrac{2}{(t+1)^2} > 0 & \text{if} \quad t > 1 \\ \dfrac{-2}{(t+1)^2} < 0 & \text{if} \quad 1 > t > 0, \end{cases} \tag{C.3}$$

i.e., the function $J(t)$ increases monotonically if $t > 1$ and decreases monotonically if $1 > t > 0$.

Let $\beta_{\max}$ and $\beta_{\min}$ be the parameters defined in (3.9) and, for a given α_o,

$$\hat{J}(\alpha_o) \stackrel{\text{def}}{=} \sup_{(i,j) \in \Phi} \{J(\alpha_o/\beta_{ij})\}. \tag{C.4}$$

Assuming $\beta_{\max} > \beta_{\min}$, it can be shown that

$$\frac{d\hat{J}(\alpha_o)}{d\alpha_o} > 0 \quad \text{if } \alpha_o > \alpha_m \qquad \text{and} \qquad \frac{d\hat{J}(\alpha_o)}{d\alpha_o} < 0 \quad \text{if } \alpha_o < \alpha_m, \tag{C.5}$$

where $\alpha_m \stackrel{\text{def}}{=} \sqrt{\beta_{\max} \cdot \beta_{\min}}$.

Proof. As a result of (C.3), $\hat{J}(\alpha_o)$ is equal to the greater of $J(\alpha_o/\beta_{\min})$ and $J(\alpha_o/\beta_{\max})$. Since (a) $\alpha_m/\beta_{\min} > 1$, (b) $\alpha_m/\beta_{\max} < 1$, and (c) $J(\alpha_m/\beta_{\min}) = J(\alpha_m/\beta_{\max})$, one concludes that

$$\hat{J}(\alpha_o) = \begin{cases} J(\alpha_o/\beta_{\min}) & \text{if} \quad \alpha_o \geqslant \alpha_m \\ J(\alpha_o/\beta_{\max}) & \text{if} \quad \alpha_o \leqslant \alpha_m. \end{cases} \tag{C.6}$$

Inequality (C.5) is a direct result of (C.3) and (C.6). Q.E.D.

Inequality (C.5) implies that $\hat{J}(\alpha_o)$ increases monotonically if $\alpha_o > \alpha_m$ and decreases monotonically if $\alpha_o < \alpha_m$. Since $\hat{J}(\alpha_o) = G^{\infty}$ (see (3.5), (C.1) and (C.4)), (3.7) and (3.8) simply state the fact that $\hat{J}(\alpha_o)$ reaches its minimum $\hat{J}(\alpha_m) = J(\alpha_m/\beta_{\min}) = G^{\infty}_{\min}$ if and only if $\alpha_o = \alpha_m$.

In case that $\beta_{\max} = \beta_{\min}$, (C.2) and (C.4) imply that the minimum of $\hat{J}(\alpha_o)$ is zero and it is reached if and only if $\alpha_o = \beta_o$, where β_o denotes the value of either $\beta_{\max}$ or $\beta_{\min}$. Equations (3.7) and (3.8) obviously are valid for this special case.

ACKNOWLEDGMENTS

The SLOR Calculations referred to in Section 4 were carried out using a code developed by Professor Shih-Hung Chang of Cleveland State University. A special thanks is extended to Mr. Richard H. Cavicchi for his help in preparing this manuscript.

REFERENCES

1. R. W. HOCKNEY, *J. Assoc. Comput. Mach.* **12**, 95 (1965).
2. R. W. HOCKNEY, in *Methods in Computational Physics*, Vol. 9, edited by B. Adler, S. Fernbach, and M. Rotenberg (Academic Press, New York, 1970), p. 135.
3. F. W. DORR, *SIAM Rev.* **12**, 248 (1970).
4. O. BUNEMAN, "A Compact Non-Iterative Poisson Solver," Institute for Plasma Research, Stanford University, Report SU-IPR-294, May 1969 (unpublished).
5. R. W. HOCKNEY, in *Numerical Methods in Applied Fluid Dynamics*, edited by B. Hunt (Academic Press, New York, 1980), p. 1.
6. E. G. D'YAKANOV, *Dokl. Akad. Nauk. SSSR* **138**, 522 (1961).
7. P. CONCUS AND G. H. GOLUB, *SIAM J. Numer. Anal.* **10**, 1103 (1973).
8. R. E. BANK, *SIAM J. Numer. Anal.* **14**, 950 (1977).
9. S. C. CHANG AND J. J. ADAMCZYK, *J. Comput. Phys.* **60**, 23 (1985).
10. S. C. CHANG AND J. J. ADAMCZYK, *J. Comput. Phys.* **60**, 41 (1985).
11. L. W. EHRLICH, *J. Comput. Phys.* **44**, 31 (1981).
12. E. F. F. BOTTA AND A. E. P. VELDMAN, *J. Comput. Phys.* **48**, 127 (1982).
13. S. C. CHANG, "Solution of Elliptic Partial Differential Equations by Fast Poisson Solvers Using a Local Relaxation Factor. I. One-Step Method," NASA Technical Paper 2529, 1986 (unpublished).
14. S. C. CHANG, *Linear Algebra Appl.* **65**, 179 (1985).
16. G. D. SMITH, *Numerical Solution of Partial Differential Equations* (Oxford Univ. Press (Clarendon), 1978), p. 29.
17. L. A. HAGEMAN AND D. M. YOUNG, *Applied Iterative Analysis* (Academic Press, New York, 1981), p. 12.
18. S. C. CHANG, "Solution of Elliptic Partial Differential Equations by Fast Poisson Solvers Using a Local Relaxation Factor. II. Two-Step Method," NASA Technical Paper 2530, 1986 (unpublished).
19. J. ADAMS, P. SWARTZTRAUBER, AND R. SWEET, "FISHPAK: A Package of FORTRAN Subprograms for the Solution of Separable Elliptic Partial Differential Equations, Version 3," National Center for Atmospheric Research, Boulder, Co., 1978 (unpublished).
20. *Multigrid Methods* (Proceedings of a Conference Held at Koln-Porz, Nov. 23–27, 1981), edited by W. Hackbusch and U. Trottenberg (Springer–Verlag, Berlin/New York, 1982).
21. K. MILLER, *Numer. Math.* **7**, 91 (1965).
22. G. RODRIGUE AND J. SIMON, in *Proceedings of 6th Int. Conf. on Comp. Meth. in App. Sci and Eng., INRIA, Dec.* 1983.
23. Q. V. DIHN, R. GLOWINSKI, AND J. PERIAUX, "Solving Elliptic Problems by Domain Decomposition Methods with Applications." *Elliptic Problem Solvers II*, G. Birkhoff, and A. Schoenstadt (Academic Press, New York, 1984), p. 395.