Numerical Solution of Symmetric Positive Differential Equations

By Theodore Katsanis

Abstract. A finite-difference method for the solution of symmetric positive linear differential equations is developed. The method is applicable to any region with piecewise smooth boundaries. Methods for solution of the finite-difference equations are discussed. The finite-difference solutions are shown to converge at essentially the rate $O(h^{1/2})$ as $h \to 0$, h being the maximum distance between adjacent mesh-points.

An alternate finite-difference method is given with the advantage that the finite-difference equations can be solved iteratively. However, there are strong limitations on the mesh arrangements which can be used with this method.

Introduction. In the theory of partial differential equations there is a fundamental distinction between those of elliptic, hyperbolic and parabolic type. Generally each type of equation has different requirements as to the boundary or initial data which must be specified to assure existence and uniqueness of solutions, and to be well posed. These requirements are usually well known for an equation of any particular type. Further, many analytical and numerical techniques have been developed for solving the various types of partial differential equations, subject to the proper boundary conditions, including even many nonlinear cases. However, for equations of mixed type much less is known, and it is usually difficult to know even what the proper boundary conditions are.

As a step toward overcoming this problem Friedrichs [1] has developed a theory of symmetric positive linear differential equations independent of type. Chu [2] has shown that this theory can be used to derive finite-difference solutions in two-dimensions for rectangular regions, or more generally, by means of a transformation, for regions with four corners joined by smooth curves. In this paper a more general finite-difference method for the solution of symmetric positive equations is presented (based on [3]). The only restriction on the shape of the region is that the boundary be piecewise smooth. It is proven that the finite-difference solution converges to the solution of the differential equation at essentially the rate $O(h^{1/2})$ as $h \to 0$, h being the maximum distance between adjacent mesh-points for a two-dimensional region. Also weak convergence to weak solutions is shown.

An alternate finite-difference method is given for the two-dimensional case with the advantage that the finite-difference equation can be solved iteratively. However, there are strong limitations on the mesh arrangements which can be used with this method.

1. Symmetric Positive Linear Differential Equations. Let Ω be a bounded open set in the m-dimensional space of real numbers, R^m. The boundary of Ω will be

Received May 18, 1967. Revised May 8, 1968.

denoted by $\partial\Omega$, and its closure by $\bar{\Omega}$. It is assumed that $\partial\Omega$ is piecewise smooth. A point in R^m is denoted by $x = (x_1, x_2, \cdots, x_m)$ and an r-dimensional vector-valued function defined on Ω is given by $u = (u_1, u_2, \cdots, u_r)$. Also let $\alpha^1, \alpha^2, \cdots, \alpha^m$ and G be given $r \times r$ matrix-valued functions and $f = (f_1, f_2, \cdots, f_r)$ a given r-dimensional vector-valued function, all defined on Ω (at least). It is assumed that the α^i are piecewise differentiable. For convenience, let $\alpha = (\alpha^1, \alpha^2, \cdots, \alpha^m)$, so that we can use expressions such as

$$\nabla\cdot(\alpha u) = \sum_{i=1}^{m} \frac{\partial}{\partial x_i}(\alpha^i u)\,. \tag{1.1}$$

With this notation we can write the identity

$$\sum_{i=1}^{m} \frac{\partial}{\partial x_i}(\alpha^i u) = \sum_{i=1}^{m} \frac{\partial \alpha^i}{\partial x_i} u + \sum_{i=1}^{m} \alpha^i \frac{\partial u}{\partial x_i}$$

simply as

$$\nabla\cdot(\alpha u) = (\nabla\cdot\alpha)u + \alpha\cdot\nabla u\,. \tag{1.2}$$

The definitions for symmetric positive operators and admissible or semiadmissible boundary conditions were introduced by Friedrichs [1].

Let K be the first-order linear partial differential operator defined by

$$Ku = \alpha\cdot\nabla u + \nabla\cdot(\alpha u) + Gu\,. \tag{1.3}$$

K is *symmetric positive* if each component, α^i, of α is symmetric and the symmetric part, $(G + G^*)/2$, of G is positive definite on $\bar{\Omega}$.

For the purpose of giving suitable boundary conditions, a matrix, β, is defined (a.e.) on $\partial\Omega$ by

$$\beta = n\cdot\alpha\,, \tag{1.4}$$

where $n = (n_1, n_2, \cdots, n_m)$ is defined to be the outer normal on $\partial\Omega$.

The boundary condition $Mu = 0$ on $\partial\Omega$ is *semiadmissible* if $M = \mu - \beta$, where μ is any matrix with nonnegative definite symmetric part, $(\mu + \mu^*)/2$. If in addition, $\mathfrak{N}(\mu - \beta) \oplus \mathfrak{N}(\mu + \beta) = R^r$ on the boundary, $\partial\Omega$, the boundary condition is termed *admissible*. ($\mathfrak{N}(\mu - \beta)$ is the null space of the matrix $(\mu - \beta)$.)

The problem is to find a function u which satisfies

$$\begin{aligned} Ku &= f \quad \text{on } \Omega\,, \\ Mu &= 0 \quad \text{on } \partial\Omega\,, \end{aligned} \tag{1.5}$$

where K is symmetric positive.

Many of the usual partial differential equations may be expressed in this symmetric positive form, with the standard boundary conditions also expressed as an admissible boundary condition. This includes equations of both hyperbolic and elliptic type. However, the greatest interest lies in the fact that the definitions are completely independent of type. An example of potentially great practical importance is the Tricomi equation which arises from the equations for transonic fluid flow. The Tricomi equation is of mixed type, i.e., it is hyperbolic in part of the region, elliptic in part, and is parabolic along the line between the two parts.

The significance of the semiadmissible boundary condition is that this insures

the uniqueness of a classical solution to a symmetric positive equation. On the other hand, the stronger, admissible boundary condition is required for existence. The existence of a classical solution is generally difficult to prove for any particular case, and depends on properties at corners of the region.

Let $\mathfrak{H}$ be the Hilbert space of all square integrable r-dimensional vector-valued functions defined on Ω. The inner product is given by

$$(u, v) = \int_\Omega u \cdot v \,, \tag{1.6}$$

where

$$u \cdot v = \sum_{i=1}^{r} u_i v_i$$

and

$$\|u\|^2 = (u, u) \,. \tag{1.7}$$

A boundary inner product is defined by

$$(u, v)_B = \int_{\partial\Omega} u \cdot v \tag{1.8}$$

with the corresponding norm

$$\|u\|_B^2 = (u, u)_B \,. \tag{1.9}$$

The adjoint operators K^* and M^* are defined by

$$K^*u = -\alpha \cdot \nabla u - \nabla \cdot (\alpha u) + G^*u \,, \tag{1.10}$$

$$M^*u = (\mu^* + \beta)u \,. \tag{1.11}$$

We will make use of the following lemmas by Friedrichs.

LEMMA 1.1 (FIRST IDENTITY). *If K is symmetric positive, then*

$$(v, Ku) + (v, Mu)_B = (K^*v, u) + (M^*v, u)_B \,. \tag{1.12}$$

LEMMA 1.2 (SECOND IDENTITY). *If K is symmetric positive, then*

$$(u, Ku) + (u, Mu)_B = (u, Gu) + (u, \mu u)_B \,. \tag{1.13}$$

LEMMA 1.3. *Suppose u is a solution to* (1.5) *where M is semiadmissible. Let λ_G be the smallest eigenvalue of $(G + G^*)/2$ in $\bar{\Omega}$. Then*

$$\|u\| \leqq (1/\lambda_G)\|f\| \,. \tag{1.14}$$

LEMMA 1.4. *Let u satisfy Eq.* (1.5) *where M is semiadmissible. Further, assume that $(\mu + \mu^*)/2$ is positive definite on $\partial\Omega$ with smallest eigenvalue λ_μ. Then*

$$\|u\|_B \leqq (1/(\lambda_G \lambda_\mu)^{1/2})\|f\| \,. \tag{1.15}$$

Lemma 1.3 insures the uniqueness of a classical solution, and also that it is well posed in L^2 for homogeneous boundary conditions.

By widening the class of solutions to (1.5) to include weak solutions it is quite easy to prove existence of a solution to a symmetric positive equation under only semiadmissible boundary conditions. We will use Friedrichs' definition of weak

solution. Let $V = C_1(\Omega) \cap \{v | M^*v = 0 \text{ on } \partial\Omega\}$. A function $u \in \mathfrak{H}$ (defined above) is a *weak solution* of (1.5) if $f \in \mathfrak{H}$ and for all $v \in V$

$$(v, f) = (K^*v, u) . \tag{1.16}$$

It follows from the "first identity" (1.12) that a classical solution is also a weak solution.

Friedrichs [1] proved the existence of weak solutions if M is semiadmissible. He also showed that, if, in addition, M is admissible and the weak solution is continuously differentiable, then the weak solution must also be a classical solution.

2. Finite-Difference Solution of Symmetric Positive Differential Equations. First we will express K in a form slightly different from (1.3), by the use of (1.2). We have

$$Ku = 2\nabla\cdot(\alpha u) - (\nabla\cdot\alpha)u + Gu . \tag{2.1}$$

Using the concept of vectors whose components are themselves matrices or vectors leads to somewhat simpler notation for the application of Green's theorem.

LEMMA 2.1 (GREEN'S THEOREM). *Let g be a continuously differentiable m-dimensional vector-valued function defined on $\Omega \subset R^m$, with vector components in either R, R^r or $R^r \times R^r$. Then*

$$\int_\Omega \nabla\cdot g = \int_{\partial\Omega} g\cdot n . \tag{2.2}$$

This result follows directly from the definitions, using Green's theorem.

We now integrate the equation $Ku = f$ over any region $P \subset \Omega$ using (2.1) and Green's theorem to obtain

$$\int_P Ku = 2\int_{\partial P} \beta u - \int_P (\nabla\cdot\alpha)u + \int_P Gu = \int_P f . \tag{2.3}$$

By a suitable approximation to (2.3) the desired finite-difference equations will be obtained.

Let H be a set of N mesh-points for Ω. It is not required for the theory that the mesh-points all lie in Ω. With each mesh-point $x_j \in H$ we identify a mesh-region $P_j \subset \Omega$ by

$$P_j = \{x | |x - x_j| < |x - x_k|, \forall x_k \in H, k \neq j; x \in \Omega\} .$$

If P_j is adjacent to P_k we say that x_j is connected to x_k (corresponding to the fact that the directed graph of the resulting matrix will have a directed path in both directions between j and k, see [4, p. 16]). Let $l_{j,k} = |x_j - x_k|$, where x_j is connected to x_k, and let $h = \max l_{j,k}$. Now define A_j to be the "volume" of P_j and $L_{j,k}$ to be the "area" of the $(r-1)$-dimensional "surface" between P_j and P_k. We put $\Gamma_{j,k} = \overline{P}_j \cap \overline{P}_k$. Fig. 1 illustrates mesh-points and corresponding mesh-regions for two dimensions. This concept of mesh-regions is based on the suggestions of MacNeal [5]. We will always use the notation $\sum_j$ to indicate a sum over all points, x_j, in H, and $\sum_k$ to indicate a sum over points, x_k, which are connected to some one point, x_j.

The desired finite-difference equation can now be obtained by a suitable approximation to Eq. (2.3). We use the symbol $\doteq$ to indicate the discrete approximation that will be used for each expression. First

$$\int_{\Gamma_{j,k}} \beta u \doteq L_{j,k}\beta_{j,k}\frac{u_j + u_k}{2} \tag{2.4}$$

where $u_j = u(x_j)$ and $\beta_{j,k}$ is the value of β for P_j at the center of $\Gamma_{j,k}$. (Note that $\beta_{j,k} = -\beta_{k,j}$.) The approximation to the next term of Eq. (2.3) requires approximating u with u_j first, and then applying Green's theorem before approximating α. With this we obtain

$$\int_{Pj} (\nabla\cdot\alpha)u \doteq \int_{Pj} (\nabla\cdot\alpha)u_j = \int_{\partial Pj} \beta u_j \,. \tag{2.5}$$

The final approximation is then

$$\int_{\Gamma_{j,k}} \beta u_j \doteq L_{j,k}\beta_{j,k}u_j \,. \tag{2.6}$$

Equations (2.4) and (2.6) take care of the integration over the interface between any P_j and P_k. Now we need to make an approximation for the boundary sides. It will be convenient to be able to subdivide $\overline{P}_j \cap \partial\Omega$ into more than one piece. We will label each piece $\Gamma_{j,B}$ and we will use the convention that $\sum_B$ will mean a summation over the B for just one j. We use $l_{j,B}$ to denote the distance from x_j to x_B, where x_B is located at the "center" of $\Gamma_{j,B}$ and $L_{j,B}$ is used for the "area" of $\Gamma_{j,B}$. Also $\beta_{j,B} = \beta(x_B)$. This notation is indicated for the two-dimensional case in Fig. 1. The desired approximations are now given by

$$\int_{\Gamma_{j,B}} \beta u \doteq L_{j,B}\beta_{j,B}u_B \,, \tag{2.7}$$

$$\int_{\Gamma_{j,B}} \beta u_j \doteq L_{j,B}\beta_{j,B}u_j \,. \tag{2.8}$$

Finally the remaining terms in equation (2.3) are approximated by

$$\int_{Pj} Gu \doteq A_jG_ju_j \,, \tag{2.9}$$

$$\int_{Pj} f \doteq A_jf_j \,, \tag{2.10}$$

where $G_j = G(x_j)$ and $f_j = f(x_j)$. Also we can approximate $\int Ku$ by

$$\int_{Pj} Ku \doteq A_j(K_hu)_j \,, \tag{2.11}$$

where K_h is the finite-difference operator to be defined and which will approximate K. Using approximations (2.4) to (2.11) in Eq. (2.3) we arrive at the following definition of K_h,

$$A_j(K_hu)_j = \sum_k L_{j,k}\beta_{j,k}u_k + \sum_B L_{j,B}\beta_{j,B}(2u_B - u_j) + A_jG_ju_j \,, \tag{2.12}$$

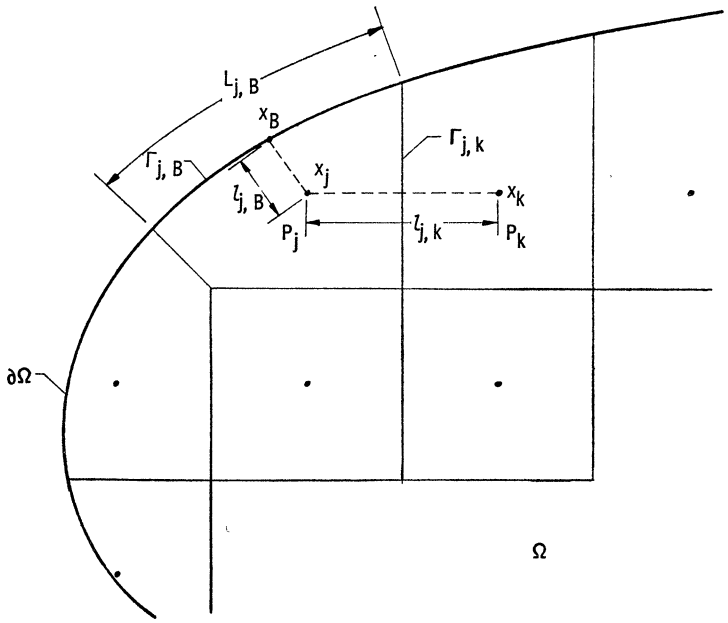

FIGURE 1. Typical mesh-regions in the two-dimensional case.

where u here denotes a discrete function defined on $\overline{H} = H \cup \{x_B\}$, and $u_j = u(x_j)$. We will seek to find a function defined on $\overline{H}$ and satisfying $(K_h u)_j = f_j$ for every $x_j \in H$. Of course the solution is not yet uniquely determined since there are more unknowns than equations. The boundary condition $Mu = 0$ will furnish us with the necessary information to determine u uniquely on H (but not necessarily on all of $\overline{H}$).

Using M_h to denote the boundary operator used to approximate M, we make the following definition

$$(M_h u)_{j,B} = \mu_{j,B} u_j - \beta_{j,B}(2u_B - u_j) \tag{2.13}$$

for all j where P_j is a boundary polygon, and for all boundary surfaces of P_j (each of which is associated with a point x_B). It is easily seen that M_h is consistent with M (i.e., $(M_h u)_{j,B} \to Mu(x_{j,B})$ as $h \to 0$ if u is continuous). The reason for this choice of M_h is that the condition $M_h u = 0$ can be used to eliminate u_B in $K_h u$ in a simple manner, and also we will be able to prove basic identities for the finite-difference operators analogous to those for the continuous operators (Eqs. (1.12) and (1.13)).

The existence and uniqueness of a solution to the finite-difference equation and the convergence to a continuous solution as $h \to 0$ depends on proving the basic identities for the discrete operators. Let $\mathfrak{H}_h$ be the finite-dimensional Hilbert space of discrete functions defined on H. The inner product is given by

$$(u, v)_h = \sum_j A_j u_j \cdot v_j \tag{2.14}$$

and

$$\|u\|_h^2 = (u, u)_h \,. \tag{2.15}$$

Also a "boundary" inner product is given by

$$(u, v)_{B_h} = \sum_j \sum_B L_{j,B} u_{j,B} \cdot v_{j,B} \tag{2.16}$$

for P_j a boundary mesh region, and

$$\|u\|^2_{B_h} = (u, u)_{B_h} . \tag{2.17}$$

The discrete adjoint operators K_h^* and M_h^* are defined in the obvious way,

$$A_j(K_h^* u)_j = -\sum_k L_{j,k}\beta_{j,k}u_k - \sum_B L_{j,B}\beta_{j,B}(2u_B - u_j) + A_j G_j^* u_j , \tag{2.18}$$

$$(M_h^* u)_{j,B} = u^*_{j,B} u_j + \beta_{j,B}(2u_B - u_j) . \tag{2.19}$$

We can now give the "first identity" for the discrete operators.

LEMMA 2.2. *If K is symmetric positive, then*

$$(v, K_h u)_h + (v, M_h u)_{B_h} = (K_h^* v, u)_h + (M_h^* v, u)_{B_h} \tag{2.20}$$

for any functions u, v defined on $\overline{H}$.

Proof. Using the definitions, Eqs. (2.12) and (2.18), we have

$$\begin{aligned}(v, K_h u)_h - (K_h^* v, u)_h = \sum_j \Big[& \sum_k L_{j,k} v_j \cdot \beta_{j,k} u_k \\ & + \sum_B L_{j,B} v_j \cdot \beta_{j,B}(2u_B - u_j) + A_j v_j \cdot G_j u_j \\ & + \sum_k L_{j,k}\beta_{j,k} v_k \cdot u_j \\ & + \sum_B L_{j,B}\beta_{j,B}(2v_B - v_j)\cdot u_j - A_j G_j^* v_j \cdot u_j \Big] .\end{aligned}$$

By rearrangement, since $\beta_{j,k} = -\beta_{k,j}$, and since $\beta_{j,k}$ is symmetric we have

$$\sum_j \sum_k L_{j,k}\beta_{j,k} v_k \cdot u_j = -\sum_j \sum_k L_{j,k} v_j \cdot \beta_{j,k} u_k$$

and we see that all terms cancel with the exception of the boundary terms, so that

$$\begin{aligned}(v, K_h u)_h - (K_h^* v, u)_h \\ = \sum_j \sum_B L_{j,B}(v_j \cdot \beta_{j,B}(2u_B - u_j) + \beta_{j,B}(2v_B - v_j)\cdot u_j) .\end{aligned} \tag{2.21}$$

On the other hand, using Eqs. (2.13) and (2.19),

$$\begin{aligned}(M_h^* v, u)_{B_h} - (v, M_h u)_{B_h} = & \sum_j \sum_B L_{j,B}(\mu^*_{j,B} v_j \cdot u_j + \beta_{j,B}(2v_B - v_j)\cdot u_j) \\ & - \sum_j \sum_B L_{j,B}(v_j \cdot \mu_{j,B} u_j - v_j \cdot \beta_{j,B}(2u_B - u_j))\end{aligned}$$

which is the same as the right side of (2.21). Hence the "first identity" for the difference operators is proved.

The discrete operators have been defined so that $K_h + K_h^* = G + G^*$ and $M_h + M_h^* = \mu + \mu^*$. By letting $v = u$ in (2.20) we can prove the discrete "second identity" as for the continuous case.

LEMMA 2.3. *If K is symmetric positive, then*

$$(u, K_h u)_h + (u, M_h u)_{B_h} = (u, Gu)_h + (u, \mu u)_{B_h} . \tag{2.22}$$

Using Eq. (2.13) and $M_h u = 0$ we can eliminate u_B from Eq. (2.12) so that the equation $K_h u = f$ can be reduced to

$$\sum_k L_{j,k}\beta_{j,k}u_k + \sum_B L_{j,B}\mu_{j,B}u_j + A_j G_j u_j = A_j f_j , \quad \forall j . \tag{2.23}$$

If we consider the case when Ω is two dimensional and rectangular, and the P_j are all equal rectangles, we can compare (2.23) with the finite-difference equation obtained by Chu [2]. The equation obtained by Chu is the same as (2.23) for interior rectangles, but is different for boundary rectangles.

Let A be the $rN \times rN$ matrix of coefficients of (2.23). Letting $\langle u, v\rangle = \sum_j u_j \cdot v_j$, the ordinary vector inner product, we have

$$\langle u, Au\rangle = (u, K_h u)_h + (u, M_h u)_{B_h} . \tag{2.24}$$

Hence, by the "second identity" (2.22), A has positive definite symmetric part which shows that A is nonsingular. We can also obtain an a priori bound for $\|u\|_h$ just as in the continuous case.

LEMMA 2.4. *Suppose u is a solution to $K_h u = f$, $M_h u = 0$, where K is symmetric positive and M is semiadmissible. Then*

$$\|u\|_h \leqq (1/\lambda_G)\|f\|_h . \tag{2.25}$$

If in addition, $(\mu + \mu^*)$ is positive definite on $\partial\Omega$, then

$$\|u\|_{B_h} \leqq \frac{1}{(\lambda_G \lambda_u)^{1/2}} \|f\|_h . \tag{2.26}$$

These bounds are obtained from the "second identity."

It is possible to show that the solution of the finite-difference equation (2.23) converges strongly to a continuously differentiable solution of equation (1.5), under the proper hypotheses. For simplicity we prove convergence only for the case when Ω is two dimensional ($m = 2$). Extension to regions in higher dimensions, with the same rate of convergence, follows directly. To allow the type of comparison we wish to make we will define operators mapping $\mathfrak{C}$ into $\mathfrak{C}_h$ and vice versa. Let $r_h : \mathfrak{C} \to \mathfrak{C}_h$ be the projection defined by

$$(r_h u)_j = u(x_j) \qquad \text{for all } x_j \in H . \tag{2.27}$$

In the other direction, let $p_h : \mathfrak{C}_h \to \mathfrak{C}$ be an injection mapping defined by

$$p_h u_h(x) = (u_h)_j , \qquad \text{for all } x \in P_j . \tag{2.28}$$

We immediately have the following relations,

$$r_h p_h = I , \tag{2.29}$$

$$\|p_h u_h\| = \|u_h\|_h \qquad \text{for all } u_h \in \mathfrak{C}_h . \tag{2.30}$$

We can now state our basic convergence theorem for two-dimensional regions.

THEOREM 2.1. *Suppose that $u \in C^2(\bar{\Omega})$ satisfies*

$$Ku = f \qquad \text{on } \Omega \subset R^2 ,$$
$$Mu = 0 \qquad \text{on } \partial\Omega ,$$

where K is symmetric positive, and $\mu + \mu^*$ is positive definite on $\partial\Omega$. For any given

$h > 0$, let H_h be a set of associated mesh-points such that the maximum distance between connected nodes is less than h and also that $L_{j,k}$, $L_{j,B}$ and $|x - x_j|$ for $x \in P_j$ are all less than h. It is assumed that the mesh is sufficiently regular so that h^2/A_j for each P_j is bounded independently of h by a constant $K_1 > 0$, which is possible for sufficiently nice regions. Also it is assumed that a uniform rectangular mesh is used for all P_j any point of which is at a distance greater than $K_2 h$ from $\partial\Omega$, where K_2 is a positive constant. It is assumed that $\alpha \in C^2(\bar{\Omega})$.

Let $u_h \in \mathfrak{H}_h$ be the unique solution to

$$K_h u_h = r_h f \quad \text{on } H_h\,, \qquad M_h u_h = 0\,.$$

Then $\|p_h u_h - u\| = O(h^\nu)$ as $h \to 0$ for any positive $\nu < 1/2$.

Chu [2] proved convergence of his finite-difference scheme, where Ω is a rectangle or a region with four corners, but the rate of convergence was not established.

Proof. Define $w_h = u_h - r_h u$. Let λ_G be the smallest eigenvalue of $(G + G^*)/2$ in $\bar{\Omega}$. Using the "second identity" (2.22), we have

$$\|w_h\|_h^2 \leqq (1/\lambda_G)[(w_h, K_h w_h)_h + (w_h, M_h w_h)_{B_h}]\,.$$

Using the Cauchy-Schwartz inequality, we have

$$\|w_h\|_h^2 \leqq (1/\lambda_G)(\|w_h\|_h \|K_h w_h\|_h + \|w_h\|_{B_h} \|M_h w_h\|_{B_h})\,. \tag{2.31}$$

We will show that $\|K_h w_h\|_h = O(h^{1/2})$ and $\|M_h w_h\|_{B_h} = O(h)$, as $h \to 0$. We shall need the following lemma.

LEMMA 2.5. *Let g be a function defined on a finite region $P \subset R^2$, and suppose that g satisfies a Lipschitz condition, i.e., there is a constant $K_3 > 0$ such that $|g(x) - g(y)| \leqq K_3|x - y|$, for all $x, y \in P$. Then, if A_0 is the area of P and $|x - x_0| \leqq h$ in P,*

$$\left| g(x_0) - \frac{1}{A_0} \int_P g(x) \right| \leqq K_3 h\,.$$

We proceed now with the proof of the theorem. Let Ω_1 denote that portion of Ω consisting of those P_j which are rectangular, and let Ω_2 denote the rest of the P_j. From the hypothesis we see that the area of Ω_2 is less than the length of $\partial\Omega$ times $K_2 h$. We have now that

$$\|K_h w_h\|_h^2 = \sum_{j \in J_1} \int_{P_j} (Ku(x_j) - (K_h r_h u)_j)^2 + \sum_{j \in J_2} \int_{P_j} (Ku(x_j) - (K_h r_h u)_j)^2\,, \tag{2.32}$$

where

$$J_i = \{j | P_j \subset \Omega_i\}\,, \qquad i = 1, 2\,.$$

To simplify notation we will use u_j for $u(x_j)$ and u_B for $u(x_B)$. We now obtain a suitable bound for $|Ku(x_j) - (K_h r_h u)_j|$

$$\begin{aligned} |Ku(x_j) - (K_h r_h u)_j| \leqq & \left| 2\nabla\cdot(\alpha u)(x_j) - \sum_k \frac{L_{j,k}}{A_j} \beta_{j,k}(u_j + u_k) - 2\sum_B \frac{L_{j,B}}{A_j} \beta_{j,B} u_B \right| \\ & + \left| (\nabla\cdot\alpha)u(x_j) - \sum_k \frac{L_{j,k}}{A_j} \beta_{j,k} u_j - \sum_B \frac{L_{j,B}}{A_j} \beta_{j,B} u_j \right|. \end{aligned} \tag{2.33}$$

Consider the first term in the last expression above

$$\left|2\nabla\cdot(\alpha u)(x_j) - \sum_k \frac{L_{j,k}}{A_j}\beta_{j,k}(u_j + u_k) - 2\sum_B \frac{L_{j,B}}{A_j}\beta_{j,B}u_B\right|$$

$$\leqq \left|2\nabla\cdot(\alpha u)(x_j) - \frac{2}{A_j}\int_{P_j}\nabla\cdot(\alpha u)\right|$$

$$+\frac{1}{A_j}\left|\sum_k \int_{\Gamma_{j,k}} 2(\beta u - (\beta u)_{j,k}) + \sum_B \int_{\Gamma_{j,B}} 2(\beta u - (\beta u)_{j,B})\right| \tag{2.34}$$

$$+\frac{1}{A_j}\left|\sum_k \int_{\Gamma_{j,k}} \beta_{j,k}(2u_{j,k} - (u_j + u_k))\right|.$$

By Lemma 2.5, since α and $u \in C^2(\bar{\Omega})$ imply that their derivatives satisfy a Lipschitz condition,

$$\left|2\nabla\cdot(\alpha u)(x_j) - \frac{2}{A_j}\int_{P_j}\nabla\cdot(\alpha u)\right| = O(h)\,. \tag{2.35}$$

We consider now the case when $j \in J_1$, so that P_j is a rectangle with x_j at the center.

Since $u \in C^2(\Omega)$, we have

$$u_j = u_{j,k} - \frac{l_{j,k}}{2}u'_{j,k} + \frac{l_{j,k}^2}{(4)2}u''(\xi_1)\,,$$

$$u_k = u_{j,k} + \frac{l_{j,k}}{2}u'_{j,k} + \frac{l_{j,k}^2}{(4)2}u''(\xi_2)\,,$$

where the derivatives are directional derivatives in the direction $x_k - x_j$. Hence, if $|u''| < K_3$ in Ω, we have

$$|2u_{j,k} - (u_j + u_k)| < (K_3/4)h^2$$

This means that

$$\left|\int_{\Gamma_{j,k}} \beta_{j,k}(2u_{j,k} - (u_j + u_k))\right| \leqq L_{j,k}\|\beta_{j,k}\|\,|2u_{j,k} - (u_j + u_k)| = O(h^3) \tag{2.36}$$

when $j \in J_1$.

We now examine a Taylor series expansion for βu about the point $x_{j,k} = (x_j + x_k)/2$.

$$\begin{aligned}\beta(x_{j,k} + tz)u(x_{j,k} + tz) &= (\beta u)_{j,k} + t\left(\frac{d}{dt}(\beta u)\right)_{j,k} + \frac{t^2}{2}g(\xi^1)\,,\\ \beta(x_{j,k} - tz)u(x_{j,k} - tz) &= (\beta u)_{j,k} - t\left(\frac{d}{dt}(\beta u)\right)_{j,k} + \frac{t^2}{2}g(\xi^2)\end{aligned} \tag{2.37}$$

where z is a unit vector orthogonal to $x_j - x_k$, t is a scalar parameter, $g(\xi) = (g_1(\xi_1), g_2(\xi_2), \cdots, g_r(\xi_r))$, g_i is the ith component of the vector $(d^2/dt^2)(\beta u)$, and ξ_i is a point on the straight line between $x_{j,k} + (L_{j,k}/2)z$ and $x_{j,k} - (L_{j,k}/2)z$. Using (2.37) we obtain the following bound,

$$\left|\int_{\Gamma_{j,k}} \beta u - (\beta u)_{j,k}\right| = O(h^3) . \tag{2.38}$$

Now, using (2.35), (2.36) and (2.38) in (2.34) we obtain

$$\left|2\nabla\cdot(\alpha u)(x_j) - \sum_k \frac{L_{j,k}}{A_j}\beta_{j,k}(u_j + u_k)\right| = O(h) \tag{2.39}$$

for all $j \in J_1$, since $h^2/A_j \leqq K_1$ and the boundary terms are not present.

Consider now the second term on the right of (2.33):

$$\begin{aligned}&\left|(\nabla\cdot\alpha)u(x_j) - \sum_k \frac{L_{j,k}}{A_j}\beta_{j,k}u_j - \sum_B \frac{L_{j,B}}{A_j}\beta_{j,B}u_j\right| \\ &\leqq \left|(\nabla\cdot\alpha)u(x_j) - \frac{1}{A_j}\int_{P_j}(\nabla\cdot\alpha)u\right| + \frac{1}{A_j}\left|\int_{P_j}(\nabla\cdot\alpha)(u-u_j)\right| \\ &\quad + \frac{1}{A_j}\left|\sum_k \int_{\Gamma_{j,k}}(\beta-\beta_{j,k})u_j + \sum_B \int_{\Gamma_{j,B}}(\beta-\beta_{j,B})u_j\right| .\end{aligned} \tag{2.40}$$

By Lemma 2.5

$$\left|(\nabla\cdot\alpha)u(x_j) - \frac{1}{A_j}\int_{P_j}(\nabla\cdot\alpha)u\right| = O(h) . \tag{2.41}$$

Next, since u satisfies a Lipschitz condition, $|x - x_j| < h$ for all $x \in P_j$, and since $\|\nabla\cdot\alpha\|$ is uniformly bounded in Ω, we have

$$\frac{1}{A_j}\left|\int_{P_j}(\nabla\cdot\alpha)(u-u_j)\right| = O(h) . \tag{2.42}$$

Finally, since $\beta_{j,k}$ and $\beta_{j,B}$ are each evaluated at the midpoint of $\Gamma_{j,k}$ or $\Gamma_{j,B}$, respectively, we can use a Taylor series analysis, as in deriving equation (2.38), to obtain

$$\frac{1}{A_j}\left|\sum_k \int_{\Gamma_{j,k}}(\beta-\beta_{j,k})u_j + \sum_B \int_{\Gamma_{j,B}}(\beta-\beta_{j,B})u_j\right| = O(h) . \tag{2.43}$$

Combining (2.41), (2.42), and (2.43) in (2.40) we obtain

$$\left|(\nabla\cdot\alpha)u(x_j) - \sum_k \frac{L_{j,k}}{A_j}\beta_{j,k}u_j - \sum_B \frac{L_{j,B}}{A_j}\beta_{j,B}u_j\right| = O(h) . \tag{2.44}$$

Note that (2.44) holds for all j, not just for $j \in J_1$.

We can now substitute (2.39) and (2.44) in (2.33) to obtain

$$|Ku(x_j) - (K_h r_h u)_j| = O(h) \qquad \text{for all } j \in J_1. \tag{2.45}$$

We cannot obtain as good a bound for $|Ku(x_j) - (K_h r_h u)_j|$ when $j \in J_2$, although (2.44) holds, since $\Gamma_{j,k}$ is not in general bisected by the line between x_j and x_k. However, we can show that $|Ku(x_j) - (K_h r_h u)_j|$ is uniformly bounded for $j \in J_2$, which is adequate since the area of Ω_2 is of order h. The two inequalities which must be re-examined are (2.36) and (2.38). We now have, since u and (βu) satisfy Lipschitz conditions, that

$$\left|\int_{\Gamma_{j,k}} \beta_{j,k}(2u_{j,k} - (u_j + u_k))\right| = O(h^2)\,, \tag{2.46}$$

$$\left|\int_{\Gamma_{j,k}} \beta u - (\beta u)_{j,k}\right| = O(h^2)\,, \tag{2.47}$$

$$\left|\int_{\Gamma_{j,B}} \beta u - (\beta u)_{j,B}\right| = O(h^2)\,.$$

Using this, with the other results which still hold, we see that $|Ku(x_j) - (K_h r_h u)_j|$ is uniformly bounded for $j \in J_2$, as $h \to 0$. Using this, together with (2.45) in (2.32) we obtain

$$\|K_h w_h\|_h^2 = O(h^2) + O(h) \tag{2.48}$$

so that

$$\|K_h w_h\|_h = O(h^{1/2})\,. \tag{2.49}$$

The next step is to show that $\|M_h w_h\|_{B_h} = O(h)$. We have

$$\|M_h w_h\|_B = \|M_h r_h u\|_{B_h}$$

since $M_h u_h = 0$. Now

$$\|M_h r_h u\|_{B_h}^2 = \sum_j \sum_B L_{j,B}(\mu_{j,B}u_j - \beta_{j,B}(2u_B - u_j))^2\,.$$

However, using the fact that $\beta_{j,B}u_B = \mu_{j,B}u_B$,

$$|\mu_{j,B}u_j - \beta_{j,B}(2u_B - u_j)| = O(h)$$

since u is differentiable, and $\|\mu\|$ and $\|\beta\|$ are uniformly bounded. This shows that

$$\|M_h r_h u\|_{B_h}^2 = O(h^2)\,,$$

since $\sum_{j,B} L_{j,B}$ is simply the length of $\partial\Omega$. This proves that

$$\|M_h w_h\|_{B_h} = O(h)\,. \tag{2.50}$$

Using (2.49) and (2.50) in (2.31), we see that

$$\|w_h\|_h^2 = \|w_h\|_h O(h^{1/2}) + \|w_h\|_{B_h} O(h)\,. \tag{2.51}$$

From Lemma 2.4, $\|w_h\|_{B_h}$ must be bounded, since

$$\|w_h\|_{B_h} \leqq \|u_h\|_{B_h} + \|r_h u\|_{B_h}$$

$$\leqq \frac{1}{(\lambda_G \lambda_\mu)^{1/2}} \|r_h f\|_h + \|r_h u\|_{B_h}$$

which is certainly uniformly bounded as $h \to 0$. Likewise $\|w_h\|_h$ is bounded. So from (2.51) we have

$$\|w_h\|_h = O(h^{1/4})\,. \tag{2.52}$$

However, if we use (2.52) in (2.51) we get $\|w_h\|_h = O(h^{3/8})$, or by repeating this procedure enough times,

$$\|w_h\|_h = O(h^\nu), \qquad \text{for any positive } \nu < 1/2. \tag{2.53}$$

Finally, we establish the convergence rate for $\|p_h u_h - u\|$. We have

$$\|p_h u_h - u\| \leqq \|w_h\|_h + \|p_h r_h u - u\|. \tag{2.54}$$

The last term can be estimated by

$$\|p_h r_h u - u\|^2 = \sum_j \int_{P_j} (u_j - u)^2 = O(h^2). \tag{2.55}$$

Using (2.53) and (2.55) in (2.54) we get

$$\|p_h u_h - u\| = O(h^\nu) + O(h) = O(h^\nu), \qquad \text{for any positive } \nu < 1/2. \tag{2.56}$$

This completes the proof of Theorem 2.1.

This finite-difference method can be applied to the Tricomi equation ([1], [3]). It is worthwhile noting that the solution obtained by the finite-difference solution of the symmetric positive form of the Tricomi equation consists of derivatives of the stream function, which corresponds to velocities in the physical problem. Hence, even though we have a convergence rate which is less than $O(h^{1/2})$, it is essentially equivalent to a convergence rate of $O(h^{3/2})$ if the original second-order equation were solved directly for the stream function.

If a rectangular mesh is used, we can partition the matrix A so as to be block tridiagonal. The matrix equation can then be solved by the block tridiagonal algorithm ([6] and [4, p. 196]). Schecter [6] shows that this algorithm is valid for any matrix with definite symmetric part. We have already shown that A has positive definite symmetric part. Schecter [6], also suggests an alternate procedure for reducing the computer storage requirements in solving the matrix equation.

An alternate method of solution may be possible in some cases. A may be decomposed as $A = D + S$ where D is Hermitian and S is skew symmetric. If the smallest eigenvalue, λ_D, of D is larger than the spectral radius, $\rho(S)$, of S, then $\|D^{-1}S\| < 1$. In this case we can use a simple iterative method. Let $u^{(0)}$ be arbitrary, and define $u^{(i)}$ recursively by $Du^{(i)} = -Su^{(i-1)} + f$. In this case $\lim_{i\to\infty} u^{(i)} = u$. In general, though, the eigenvalues of D will not be sufficiently large for this simple method to work. However, the original finite difference equations can be modified in some cases by the addition of a "viscosity" term, so as to obtain a convergent iterative procedure for the solution of the matrix equation. This will be discussed further in the next section.

We can consider the discrete analogue of a weak solution. Let V_h be the set of discrete functions, v_h, defined on $\overline{H}$ and satisfying $M_h{}^* v_h = 0$. For a discrete weak solution, u_h, we would then require that

$$(K_h{}^* v_h, u_h)_h = (v_h, r_h f)_h \qquad \text{for all } v \in V_h. \tag{2.57}$$

From the "first identity" (2.20) we have then

$$(v_h, r_h f)_h = (v_h, K_h u_h)_h + (v_h, M_h u_h)_{B_h} \qquad \text{for all } v \in V_h. \tag{2.58}$$

We see from this that $(K_h u_h)_j = f_j$ for all P_j which are not on the boundary, by choosing $(v_h)_j = 1$, and $(v_h)_k = 0$ for $k \neq j$. Because of the discrete nature of the equations we are not assured of u_h satisfying the boundary conditions. However,

conversely, if u_h satisfies $K_h u_h = r_h f$ and $M_h u_h = 0$ we see immediately that (2.57) must be satisfied.

Chu [2] has shown weak convergence of his finite-difference solution to a weak solution of a symmetric positive equation and Cea [7] has investigated generally the question of weak or strong convergence of approximate solutions to weak solutions of elliptic equations. Using these ideas, we can prove weak convergence of our finite-difference solutions to weak solutions of symmetric positive equations.

THEOREM 2.2. *For any $h > 0$, let H_h be a set of mesh points satisfying the requirements of Theorem* 2.1. *It is assumed that $\alpha \in C^2(\bar{\Omega})$. Let u_h be the unique solution to $K_h u_h = r_h f$, $M_h u_h = 0$.*

If $\{h_i\}_{i=1}^{\infty}$ is a positive sequence converging to zero, then $\{p_{h_i} u_{h_i}\}_{i=1}^{\infty}$ has a subsequence which converges weakly in H to a weak solution, u, of Eq. (1.5), *that is $(K^*v, u) = (v, f)$ for all $v \in V$.*

Furthermore, if u is a unique weak solution, then $\{p_{h_i} u_{h_i}\}_{i=1}^{\infty}$ converges weakly to u.

Proof. First we note that $\|p_h u_h\|$ is bounded, since $\|p_h u_h\| = \|u_h\|_h \leqq (1/\lambda_G)\|r_h f\|_h$, by Lemma 2.4. Hence, there is a subsequence of $\{p_{h_i} u_{h_i}\}$ that converges weakly to some $u \in \mathfrak{H}$. (See Theorem 4.41-B, Taylor [8].) For convenience of notation we will suppress the subscripts on the h.

We have, for all $v \in V$,

$$|(K_h^* r_h v, u_h)_h - (K^* v, p_h u_h)| \leqq (\|K_h^* r_h v - r_h K^* v\|_h + \|p_h r_h K^* v - K^* v\|)\|p_h u_h\| . \tag{2.59}$$

But $\|p_h r_h K^* v - K^* v\| \to 0$, and in Theorem 2.1 we can substitute K^* for K in equation (2.49) to show that $\|K_h^* r_h v - r_h K^* v\| \to 0$ (since $K_h w_h = r_h K u - K_h r_h u$). Since $\|p_h u_h\|$ is bounded,

$$\lim_{h \to 0} |(K_h^* r_h v, u_h)_h - (K^* v, p_h u_h)| = 0 .$$

However, since $K^* v \in \mathfrak{H}$, we know that $\lim_{h \to 0} (K^* v, p_h u_h) = (K^* v, u)$.

We have shown, then, that

$$\lim_{h \to 0} (K_h^* r_h v, u_h)_h = (K^* v, u) \quad \text{for all } v \in V . \tag{2.60}$$

The discrete "first identity," Eq. (2.20), gives

$$(K_h^* r_h v, u_h)_h + (M^* r_h v, u_h)_{B_h} = (r_h v, r_h f)_h . \tag{2.61}$$

Hence

$$|(K_h^* r_h v, u_h)_h - (r_h v, r_h f)_h| \leqq \|M^* r_h v\|_{B_h} \|u_h\|_{B_h} . \tag{2.62}$$

By Lemma 2.4 $\|u_h\|_{B_h} \leqq \|r_h f\|_h / (\lambda_G \lambda_\mu)^{1/2}$ which is bounded. Also, the proof of equation (2.50) shows that $\lim_{h \to 0} \|M^* r_h v\|_{B_h} = 0$, for all $v \in V$, so that

$$\lim_{h \to 0} |(K_h^* r_h v, u_h)_h - (r_h v, r_h f)_h| = 0 . \tag{2.63}$$

Further, it is obvious that

$$\lim_{h \to 0} (r_h v, r_h f)_h = (v, f) . \tag{2.64}$$

Combining (2.60), (2.63) and (2.64) gives

$$(K^*v, u) = (v, f) \quad \text{for all } v \in V ,$$

which completes the proof of the theorem.

3. Special Finite-Difference Scheme for Iterative Solution of Matrix Equation. As pointed out in Section 2, the matrix equation $Au = f$ can be solved by an iterative procedure if the eigenvalues of the diagonal coefficient matrix are sufficiently large compared to the eigenvalues of the off-diagonal coefficient matrix. Following the idea of Chu [2] we modify the finite-difference equation by adding a "viscosity" term which will have a diminishing effect on the finite-difference equations as $h \to 0$, and yet will assure the convergence of an iterative method. Unfortunately, the method is not applicable to every arrangement of mesh-points. In fact there are rather severe restrictions which must be met. The first requirement is that the difference in areas of adjacent mesh-regions be sufficiently small. This cannot be readily done along an irregular boundary, however, unless the boundary is modified. A problem arises if the boundary is modified. The boundary condition is given by $Mu = (\mu - \beta)u = 0$ on $\partial\Omega$. We need to extend M to be defined in a neighborhood of the boundary. It is possible to extend M continuously in a neighborhood of the boundary. However, if the direction of the boundary changes, β changes drastically, and we have no assurance that μ will be positive definite. The second requirement then is that M can be extended continuously over a neighborhood of the boundary, in such a way that μ will have positive definite symmetric part along the approximating boundary.

Let Ω_h be an approximation to Ω. Ω_h will have to meet several requirements to be specified later. H_h will denote a set of mesh-points associated with Ω_h and with maximum distance h between connected nodes, and $\overline{H}_h$ will denote $H_h \cup \{x_B\}$. The discrete inner product is given by

$$(u_h, v_h) = \sum_j A_j (u_h)_j \cdot (v_h)_j \tag{3.1}$$

with the A_j being the area of $P_j \subset \Omega_h$. Similarly, the "boundary" inner product is changed so that the lengths, $L_{j,B}$, are the lengths along $\partial\Omega_h$.

We define now two new finite-difference operators, $\overline{K}_h$ and $\overline{M}_h$, by

$$(\overline{K}_h u)_j = (K_h u)_j + \sum_k \sigma \frac{u_j - u_k}{l_{j,k}} + \sum_B \sigma \frac{u_j - u_B}{l_{j,B}} , \tag{3.2}$$

$$(\overline{M}_h u)_{j,B} = (M_h u)_{j,B} - \frac{\sigma A_j}{L_{j,B} l_{j,B}} (u_j - u_B) , \tag{3.3}$$

where σ is a positive number which must satisfy requirements to be specified later.

It will be useful to prove a slightly different version of the "second identity."

LEMMA 3.1. *If K is symmetric positive, then*

$$(u_h, \overline{K}_h u_h)_h + (u_h, \overline{M}_h u_h)_{B_h} = (u_h, Gu_h)_h + (u_h, \mu u_h)_{B_h} + \sum_{(j,k)} \frac{\sigma A_j}{l_{j,k}} (u_j - u_k)^2 \tag{3.4}$$

where $\sum_{(j,k)}$ indicates a sum over every (j, k) pair where x_j is connected to x_k.

Proof. Using the "second identity" for K_h and M_h, Eq. (2.22), we have

$$(u_h, \overline{K}_h u_h)_h + (u_h, \overline{M}_h u_h)_{B_h} = (u_h, Gu_h)_h + (u_h, \mu u_h)_{Bh} + \sum_j \sum_k \frac{\sigma A_j}{l_{j,k}} u_j \cdot (u_j - u_k) + \sum_j \sum_B \frac{\sigma A_j}{l_{j,B}} u_j \cdot (u_j - u_B) - \sum_j \sum_B \frac{\sigma A_j}{l_{j,B}} u_j \cdot (u_j - u_B) .$$

The last two terms cancel. For the other term we have

$$\sum_j \sum_k \frac{\sigma A_j}{l_{j,k}} u_j \cdot (u_j - u_k) = \sum_{(j,k)} \frac{\sigma A_j}{l_{j,k}} (u_j - u_k)^2$$

which completes the proof.

Lemma 3.1 immediately assures the existence and uniqueness of a solution for the special finite-difference scheme. Using $\overline{M}_h u_h = 0$ to eliminate u_B from $\overline{K}_h u_h = r_h f$, we obtain

$$(3.5) \quad \sum_k \left(L_{j,k}\beta_{j,k} - \frac{\sigma A_j}{l_{j,k}} I \right) u_k + \left(A_j G_j + \sum_k \frac{\sigma A_j}{l_{j,k}} I + \sum_B L_{j,B}\mu_{j,B} \right) u_j = A_j f_j$$

for all $x_j \in H_h$.

Let A be the matrix of coefficients of (3.5).

Lemma 3.2. *If K is symmetric positive, then $\overline{K}_h u_h = r_h f$, $\overline{M}_h u_h = 0$ has a unique solution on H_h.*

Proof. The hypothesis implies that

$$\langle u, Au \rangle = (u_h, \overline{K}_h u_h)_h + (u_h, \overline{M}_h u_h)_{B_h} .$$

By Lemma 3.1 A has positive definite symmetric part, and hence is nonsingular. Thus (3.5) defines u_h uniquely on H_h.

Also it will be noted that the "second identity" of Lemma 3.1 will give the same a priori bounds for $\|u_h\|_h$ and $\|u_h\|_{B_h}$ as given by (2.25) and (2.26).

We will now show that the special finite-difference scheme converges to a smooth solution, under a number of hypotheses given in the theorem. The theorem also includes all the hypotheses needed to assure convergence of the iterative matrix solution. Though quite a number of requirements are given, there are only two essential restrictions, namely, that the areas A_j must be nearly uniform, and that M can be specified on a modified boundary in such a way that μ remains positive definite.

Theorem 3.1. *Suppose that $u \in C^2(\bar{\Omega})$ satisfies $Ku = f$ on Ω, $Mu = 0$ on $\partial\Omega$, where K is symmetric positive. For any $h > 0$, let Ω_h be an approximation to Ω, and let H_h be a corresponding set of mesh points with maximum distance h between connected nodes, and also with $L_{j,k}$, $L_{j,B}$, and $|x - x_j|$ for $x \in P_j$ all less than h. It is assumed that the following hypotheses are satisfied:*

(i) *There exists $K_1 > 0$, independent of h, such that for every P_j we have $h^2/A_j < K_1$.*

(ii) *There exists $K_2 > 0$, independent of h, such that all P_j with any point at a distance greater than $K_2 h$ from $\partial\Omega$ are equal rectangles.*

(iii) *There exists $K_3 > 0$, independent of h, such that for all $x \in \partial\Omega_h$, the distance from x to $\partial\Omega$ is less than $K_3 h$.*

(iv) *There exists* $K_4 > 0$, *such that* M *can be extended so as to satisfy a uniform Lipschitz condition at all points at a distance less than* K_4 *from* $\partial\Omega$.

(v) Ω_h *is such that* $\mu = M + \beta$ *has positive definite symmetric part on* $\partial\Omega_h$.

(vi) *Let* W *be the set of points that are a distance less than* K_4 *from* $\partial\Omega$. *Then* α, G, *and* f *are all extended to be defined on* $\Omega \cup W$ *with* $\alpha \in C^2(\Omega \cup W)$ *and* G *positive definite on* $\Omega \cup W$.

(vii) *There exists* $K_5 > 0$, *independent of* h, *such that all points,* x_j, *associated with a boundary polygon,* P_j, *are in the polygon, and at a sufficient distance,* $l_{j,B}$, *from any boundary node,* x_B, *of* P_j *so that* $A_j \leqq K_5 L_{j,B} l_{j,B}$.

(viii) *Either* $\Omega_h \subset \Omega$ *or else* u *can be extended so that* $u \in C^2(\bar{\Omega}_h)$.

(ix) $\sigma > \eta K_1 \rho_B + d$, *where* $d > 0$ *and* ρ_B *is the supremum of the spectral radius of* $n \cdot \alpha(x)$ *for* $x \in \Omega \cup W$, *where* n *is any unit vector and* η *is the maximum number of nodes connected to any one node.*

(x) $|A_j/A_k - 1| < d\lambda_G(h')^2/(\eta^2\sigma^2 h)$, *for all connected nodes,* x_j *and* x_k, *where* λ_G *is the smallest eigenvalue of* G *in* $\bar{\Omega}_h$, *and* $h' = \min(l_{j,k})$.

(xi) *The length of* $\partial\Omega_h$ *is uniformly bounded.*

Let u_h *be the unique solution to* $\overline{K}_h u_h = r_h f$, $\overline{M}_h u_h = 0$; *then*

$$\|u_h - r_h u\| = O(h^\nu) \qquad \text{as } h \to 0\,, \qquad \text{for any positive } \nu < 1/2\,.$$

Proof. Letting $w_h = u_h - r_h u$, and using the "second identity," (3.4), we see that the inequality (2.31) is still valid for $\overline{K}_h$ and $\overline{M}_h$,

$$\|w_h\|_h^2 \leqq (1/\lambda_G)(\|w_h\|_h \|\overline{K}_h w_h\|_h + \|w_h\|_{B_h} \|\overline{M}_h w_h\|_{B_h})\,. \tag{3.6}$$

We have

$$\overline{K}_h w_h = r_h f - \overline{K}_h r_h u\,;$$

hence

$$\|\overline{K}_h w_h\|_h \leqq \|r_h K u - K_h r_h u\|_h + \|K_h r_h u - \overline{K}_h r_h u\|_h\,. \tag{3.7}$$

In checking the proof of Theorem 2.1 we see that $r_h K u - K_h r_h u$ is the same as $K_h w_h$ (Theorem 2.1), hence the bound of (2.49) holds for this term;

$$\|r_h K u - K_h r_h u\|_h = O(h^{1/2})\,. \tag{3.8}$$

For the other term we have

$$\|(\overline{K}_h - K_h) r_h u\|_h^2 = \sum_j A_j \sigma^2 \Big(\sum_k \frac{u_j - u_k}{l_{j,k}} + \sum_B \frac{u_j - u_B}{l_{j,B}} \Big)^2\,. \tag{3.9}$$

Let J_1 denote the set of subscripts for those P_j which are equal rectangles and let J_2 denote the rest of the subscripts. When $j \in J_1$ we have only the term $\sum_k (u_j - u_k)/l_{j,k}$ to consider. Because of the rectangular arrangement of points we can use a Taylor series analysis to show that

$$\Big| \sum_k \frac{u_j - u_k}{l_{j,k}} \Big| = O(h)$$

so that

$$\sum_{j \in J_1} A_j \sigma^2 \Big(\sum_k \frac{u_j - u_k}{l_{j,k}} \Big)^2 = O(h^2)\,. \tag{3.10}$$

On the other hand, when $j \in J_2$ we cannot do as well. However, we note that both $(u_j - u_k)/l_{j,k}$ and $(u_j - u_B)/l_{j,k}$ are uniformly bounded since u has a bounded derivative. Also, by hypothesis (ii), $\sum_{j\in j_2} A_j = O(h)$, so that

$$\sum_{j\in J_2} A_j\sigma^2\left(\sum_k \frac{u_j - u_k}{l_{j,k}} + \sum_B \frac{u_j - u_B}{l_{j,B}}\right)^2 = O(h) . \tag{3.11}$$

It is assumed, of course, that the number of nodes connected to any one node is bounded as $h \to 0$.

Now, using (3.10) and (3.11) in (3.9) we have

$$\|(\overline{K}_h - K_h)r_h u\|_h = O(h^{1/2}) . \tag{3.12}$$

Taking this together with (3.8) in (3.7) finally

$$\|\overline{K}_h w_h\|_h = O(h^{1/2}) . \tag{3.13}$$

It is necessary now to obtain a bound for $\|\overline{M}_h w_h\|_{B_h}$. Since $\overline{M}_h w_h = -\overline{M}_h r_h u$, we have

$$\|\overline{M}_h w_h\|_{B_h} \leqq \|M_h r_h u\|_{B_h} + \|(\overline{M}_h - M_h)r_h u\|_{B_h} . \tag{3.14}$$

We have

$$\|M_h r_h u\|_B^2 = \sum_j \sum_B L_{j,B}(\mu_{j,B} - \beta_{j,B}(2u_B - u_j))^2 .$$

We can establish a bound, since

$$\begin{aligned}|\mu_{j,B} - \beta_{j,B}(2u_B - u_j)| \leqq\ & |\mu_{j,B}(u_j - u_B)| \\ & + |(\mu_{j,B} - \beta_{j,B})u_B| + |\beta_{j,B}(u_j - u_B)| .\end{aligned}$$

The first and last term on the right are of order h, since u is differentiable and $\|\mu\|$ and $\|\beta\|$ are bounded. By hypothesis (iv) M satisfies a Lipschitz condition, and so does u. Since the distance from x_B to $\partial\Omega$ is less than $K_3 h$ by (iii) and $Mu = 0$ on $\partial\Omega$, we see that $|(\mu_{j,B} - \beta_{j,B})u_B| = O(h)$. Since, by (xi), $\sum_j \sum_B L_{j,B}$ is uniformly bounded, we have

$$\|M_h r_h u\|_{B_h} = O(h) . \tag{3.15}$$

Also, by using (vii)

$$\|(\overline{M}_h - M_h)r_h u\|_B^2 \leqq \sum_j \sum_B L_{j,B}K_5^2\sigma^2(u_j - u_B)^2 = O(h^2) . \tag{3.16}$$

This shows that

$$\|\overline{M}w_h\|_{B_h} = O(h) . \tag{3.17}$$

We check now to see that $\|w_h\|_h$ and $\|w_h\|_{B_h}$ are bounded. We have, using the a priori bound for $\|u_h\|_h$,

$$\|w_h\|_h \leqq (1/\lambda_G)\|r_h f\|_h + \|r_h u\|_h \tag{3.18}$$

which must be bounded since f and u are. In the same manner, $\|w_h\|_{Bh}$ must be bounded. Using this fact together with (3.13) and (3.17) in (3.6) we have

$$\|w_h\|_h = O(h^{1/4}) . \tag{3.19}$$

Using now (3.19) in (3.6) we get $\|w_h\|_h = O(h^{3/8})$ and by repeating the process as many times as needed we get

$$\|w_h\|_h = O(h^\nu) \qquad \text{for any positive } \nu < 1/2 . \tag{3.20}$$

This completes the proof of Theorem 3.1.

For the iterative solution of the matrix equation $Au = f$ we will split A into a block diagonal part D, and off diagonal part B. (We will suppress the subscript h on the finite-difference solution u_h.) Thus, from (3.5), the jth block of D is an $r \times r$ matrix,

$$D_j = A_j G_j + \sum_k \frac{\sigma A_j}{l_{j,k}} I + \sum_B L_{j,B}\mu_{j,B}$$

and a typical block element of B is

$$B_{j,k} = L_{j,k}\beta_{j,k} - \frac{\sigma A_j}{l_{j,k}} I$$

and $A = D + B$. The iterative method is given by

$$u^{(i+1)} = -D^{-1}Bu^{(i)} + D^{-1}f$$

where $u^{(0)}$ is arbitrary. The hypotheses of Theorem 3.1 assure the convergence of $u^{(i)}$ to u.

THEOREM 3.2. *For any $h > 0$, let Ω_h and H_h satisfy the hypotheses of Theorem 3.1. Let $u^{(0)}$ be an arbitrary vector defined on H_h, and let $\{u^{(i)}\}_{i=0}^{\infty}$ be a sequence defined recursively by*

$$u^{(i+1)} = -D^{-1}Bu^{(i)} + D^{-1}f .$$

Then $\lim_{i\to\infty} u^{(i)} = u$, where $Au = f$.

Proof. By the contraction mapping theorem it is sufficient to show that $\|D^{-1}B\| < 1$ for some matrix norm. Let v be an arbitrary vector defined on H_h, and let $w = D^{-1}Bv$. Since $Dw = Bv$, we have $\langle w, Dw\rangle = \langle w, Bv\rangle$, or

$$\begin{aligned} &\sum_j w_j \cdot \left(A_j G_j + \sum_k \frac{\sigma A_j}{l_{j,k}} I + \sum_B L_{j,B}\mu_{j,B} \right) w_j \\ &\leqq \frac{1}{2} \sum_j \sum_k w_j \cdot \left(\frac{\sigma A_j}{l_{j,k}} I - L_{j,k}\beta_{j,k} \right) w_j \\ &\quad + \frac{1}{2} \sum_j \sum_k v_k \cdot \left(\frac{\sigma A_j}{l_{j,k}} I - L_{j,k}\beta_{j,k} \right) v_k . \end{aligned} \tag{3.21}$$

This last inequality follows from the fact that

$$\langle w, Hv\rangle \leqq \tfrac{1}{2}\langle w, Hw\rangle + \tfrac{1}{2}\langle v, Hv\rangle$$

for any positive definite Hermitian matrix. We see that $(\sigma A_j)/(l_{j,k})I - L_{j,k}\beta_{j,k}$ is positive definite, since

$$\sigma A_j / l_{j,k} \geqq L_{j,k}\rho(\beta_{j,k}) \tag{3.22}$$

by (i) and (ix). By rearranging the terms of (3.21) so as to have all the w terms on the left and all the v terms to the right, we obtain

$$\sum_j w_j \cdot \left(A_j G_j + \sum_B L_{j,B} \mu_{j,B} \right) w_j + \frac{1}{2} \sum_j \sum_k w_j \cdot \left(\frac{\sigma A_j}{l_{j,k}} I + L_{j,k} \beta_{j,k} \right) w_j \tag{3.23}$$
$$\leqq \frac{1}{2} \sum_j \sum_k v_j \cdot \left(\frac{\sigma A_k}{l_{j,k}} I + L_{j,k} \beta_{j,k} \right) v_j .$$

The last expression was obtained by interchanging j and k, since

$$\beta_{j,k} = -\beta_{k,j} .$$

We can write (3.23) in the following form.

$$\sum_j w_j \cdot \left(A_j G_j + \sum_B L_{j,B} \mu_{j,B} \right) w_j + \frac{1}{2} \sum_j \sum_k w_j \cdot \left(\frac{\sigma A_j}{l_{j,k}} I + L_{j,k} \beta_{j,k} \right) w_j \tag{3.24}$$
$$\leqq \frac{1}{2} \sum_j \sum_k v_j \cdot \left(\frac{\sigma A_j}{l_{j,k}} I + L_{j,k} \beta_{j,k} \right) v_j + \frac{1}{2} \sum_j \sum_k \frac{\sigma}{l_{j,k}} (A_k - A_j) v_j^2$$

or

$$\langle w, Xw \rangle + \langle w, Yw \rangle \leqq \langle v, Yv \rangle + \langle v, Zv \rangle \tag{3.25}$$

where X, Y, and Z are matrices defined by (3.24).

We have already shown that Y is positive definite (using (3.22)); hence we can define a norm by

$$\|v\|_Y^2 = \langle v, Yv \rangle . \tag{3.26}$$

We will show that $D^{-1}B$ is a strict contraction in the Y norm. First we will need some inequalities. We have

$$\langle w, Xw \rangle > \lambda_G \|w\|_h^2 . \tag{3.27}$$

By (i) and (ix) we have

$$\langle w, Yw \rangle \leqq (\eta\sigma/h') \|w\|_h^2 . \tag{3.28}$$

Also $\langle w, Yw \rangle$ can be bounded below by using (i) and (ix):

$$\langle v, Yv \rangle \geqq (d/2h) \|v\|_h^2 . \tag{3.29}$$

Finally, we have

$$\langle v, Zv \rangle \leqq \Lambda(\eta\sigma/2h') \|v\|_h^2 \tag{3.30}$$

where $\Lambda = \max |A_k/A_j - 1|$, for all connected nodes, x_j and x_k. From the definition (3.26), and using (3.27) and (3.28) we have

$$\langle w, Xw \rangle + \langle w, Yw \rangle > (1 + \lambda_G h'/\eta\sigma) \|w\|_Y^2 . \tag{3.31}$$

On the other hand from (3.29) and (3.30)

$$\langle v, Yv \rangle + \langle v, Zv \rangle \leqq \left[1 + \frac{\eta\sigma\Lambda}{d} \left(\frac{h}{h'} \right) \right] \|v\|_Y^2 . \tag{3.32}$$

Substituting (3.31) and (3.32) in (3.25) we have

$$\|w\|_Y^2 < \left(\frac{1 + \frac{\eta\sigma\Lambda}{d} \left(\frac{h}{h'} \right)}{1 + \lambda_G h'/\eta\sigma} \right) \|v\|_Y^2 . \tag{3.33}$$

Since $w = D^{-1}Bv$, and v is arbitrary, we see that $\|D^{-1}B\|_Y < 1$ since

$$\Lambda < \frac{d\lambda_G h'}{\eta^2 \sigma^2} \left(\frac{h'}{h} \right) \tag{3.34}$$

by hypothesis (x). This completes the proof of Theorem 3.2.

Of course, if Ω_h can be selected so that all the A_j are equal, then hypothesis (x) is satisfied, and

$$\|D^{-1}B\|_Y < \frac{1}{(1 + (\lambda_G h'/\eta\sigma))^{1/2}}. \tag{3.35}$$

In the special case where all the P_j are equal rectangles, $\eta = 4$, so that

$$\|D^{-1}B\|_Y < \frac{1}{(1 + (\lambda_G h'/4\sigma))^{1/2}}. \tag{3.36}$$

4. Concluding Remarks. The Tricomi equation can be expressed in symmetric positive form. In [3] a Tricomi equation with a known analytical solution was solved numerically as an illustration of the numerical results which can be obtained. There was strong convergence to the analytical solution, but pointwise divergence. However, smoothing of the solution produced satisfactory numerical results.

5. Acknowledgement. I would like to express my appreciation to Professor Milton Lees for his guidance in this work.

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

→ K. O. Friedrichs, "Symmetric positive linear differential equations," *Comm. Pure Appl. Math.*, v. 11, 1958, pp. 333–418. MR **20** #7147.

2. C. K. Chu, *Type-Insensitive Finite Difference Schemes*, Ph.D. Thesis, New York University, 1958.

3. T. Katsanis, *Numerical Techniques for the Solution of Symmetric Positive Linear Differential Equations*, Ph.D. Thesis, Case Institute of Technology, 1967.

4. R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Princeton, N. J., 1962. MR **28** #1725.

5. R. H. MacNeal, "An asymmetrical finite difference network," *Quart. Appl. Math.*, v. 11, 1953, pp. 295–310. MR **15**, 257.

6. S. Schecter, "Quasi-tridiagonal matrices and type-insensitive difference equations," *Quart. Appl. Math.*, v. 18, 1960/61, pp. 285–295. MR **22** #5133.

7. J. Céa, "Approximation variationelle des problèmes aux limites," *Ann. Inst. Fourier* (Grenoble), v. 14, 1964, fasc. 2, pp. 345–444. MR **30** #5037.

8. A. E. Taylor, *Introduction to Functional Analysis*, Chapman & Hall, London; Wiley, New York, 1958. MR **20** #5411.