An Analysis of "Boundary-Value Techniques" for Parabolic Problems*

By Alfred Carasso** and Seymour V. Parter

Abstract. Finite-difference methods for parabolic initial boundary problems are usually treated as marching procedures. However, if the solution reaches a known steady state value as $t \to \infty$, one may provide approximate values on a line $t = T$ for a preselected T suitably large. With this extra data, it is feasible to consider the use of elliptic boundary-value techniques for the numerical computation of such problems. In this report we give a complete analysis of this method for the linear second-order case with time-independent coefficients. We also discuss iterative methods for solving the difference equations. Finally, we give an example where the method fails.

1.1. Introduction. Consider the one-dimensional "heat equation" in a strip:

$$\partial u/\partial t = \partial^2 u/\partial x^2, \qquad 0 < x < 1, \quad t > 0 \tag{1.1}$$

subject to the Dirichlet conditions

$$\begin{aligned} u(x, 0) &= f(x), \qquad 0 \leqq x \leqq 1, \\ u(0, t) &= u(1, t) = 0, \qquad t \geqq 0. \end{aligned} \tag{1.2}$$

It is well known that, provided $f(x)$ is "smooth", there is a unique solution $u(x, t)$ and

$$|u(x, t)| \leqq K \exp[-\pi^2 t], \quad \text{where } K \text{ is a constant.} \tag{1.3}$$

Let $\Delta x = 1/(M + 1)$, M a positive integer, let $\Delta t > 0$, let $v_k^n \equiv v(k\Delta x, n\Delta t)$, and consider the following finite difference approximation of (1.1), (1.2):

$$\frac{v_k^{n+1} - v_k^{n-1}}{2\Delta t} = \frac{v_{k+1}^n - 2v_k^n + v_{k-1}^n}{\Delta x^2}, \qquad k = 1, \dots, M, \quad n = 1, 2, \dots$$

$$v_0^n = v_{M+1}^n = 0, \qquad n = 0, 1, 2, \dots \tag{1.4}$$

$$v_k^0 = f(k\Delta x), \qquad k = 0, 1, \dots, M + 1.$$

Rather than use these equations as a marching procedure, D. Greenspan recently, (see [10], [11]), proposed an alternative approach: Choose N large and solve the system

Received October 8, 1968, revised September 4, 1969.

AMS Subject Classifications. Primary 6568, 6562.

Key Words and Phrases. Boundary-value technique, parabolic equations, use of limiting steady-state, long times.

*Prepared under Contract Number NO0014-67-A-1028-0004 at the University of Wisconsin. "This research was supported by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government."

**Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico 87106.

$$\begin{aligned}&\frac{v_k^{n+1} - v_k^{n-1}}{2\Delta t} = \frac{v_{k+1}^n - 2v_k^n + v_{k-1}^n}{\Delta x^2}, \qquad k = 1, \ldots, M, \quad n = 1, \ldots, N. \\ (1.5) \qquad &v_0^n = v_{M+1}^n = 0, \qquad n = 0, 1, 2, \ldots, N + 1 \\ &v_k^0 = f(k\Delta x), \qquad v_k^{N+1} \equiv 0, \qquad k = 0, 1, \ldots, M + 1\end{aligned}$$

of MN linear equations in MN unknowns. Indeed, Greenspan suggested this method for a general class of parabolic problems, linear and nonlinear, and carried out several interesting computational experiments.

The scheme selected by Greenspan is the leap-frog scheme discussed in Richtmyer [15]. When used as a marching procedure with parabolic problems, this scheme leads to an improperly posed numerical problem as data on the line $t = \Delta t$ must be supplied, in addition to the usual data, in order to start the calculation. (This is why it is possible to use it as a boundary-value procedure.) However, even if this extra data were exactly known, the scheme would in general be useless as a marching procedure: it is *unconditionally* unstable and therefore always diverges whenever the solution to the analytic problem contains arbitrarily high frequencies. We will show, however, that as a boundary-value procedure, for linear problems with time-independent coefficients, the scheme is unconditionally uniformly convergent, and the rate of convergence is $O(\mu^2)$ as the "mesh-size" $\mu \to 0$, $T \to \infty$, under minimal smoothness of the solution. Indeed, for linear problems with time dependent coefficients and for mildly nonlinear problems, one has uniform convergence at the rate of $O(\Delta t^{3/2})$ as $\Delta t \to 0$, $T \to \infty$, $\Delta x = O(\Delta t)$, and at the rate of $O(\Delta t^2)$ for sufficiently smooth exponentially decaying solutions. These results will appear in a later report (see [4] also).

We also analyze an example with which Greenspan had computational difficulty and which points out one of the interesting features of the boundary value method. We then discuss the convergence of the usual iterative methods for solving the systems of linear equations which arise in this method. We observe that, unlike the case of systems of elliptic difference equations, line iterative methods may diverge even if the related point iterative methods converge.

As we have undertaken a very thorough study of this method, it is reasonable to comment on its merits and the meaning of the results at the conclusion. Thus, we include a short section of commentary.

1.2. Notation and Definitions. Let Δx, Δt be small increments in the variables x, t, and let $T = (N + 1)\Delta t$ where N is a positive integer. Let M be a positive integer so that $1 = (M + 1)\Delta x$. Introduce a mesh over $R_T \equiv \{(x, t) \mid 0 < x < 1, 0 < t < T\}$ by means of the lines $x = k\Delta x$, $k = 1, \ldots, M$, $t = n\Delta t$, $n = 1, \ldots, N$. We will be dealing with functions $v(x, t)$ defined at the mesh-points of R_T and we adopt the notation

$$(1.6) \qquad v_k^n \equiv v(k\Delta x, n\Delta t).$$

Denote by V^n the M component vector, or M-vector

$$(1.7) \qquad V^n = \{v_1^n, v_2^n, \ldots, v_M^n\}^T,$$

and let V be the "block" vector of MN components

$$(1.8) \qquad V = \{V^1, V^2, \ldots, V^N\}^T.$$

Let ξ denote an N-component vector

$$\xi = \{\xi^1, \xi^2, \dots, \xi^N\}^T. \tag{1.9}$$

We define the following norms and scalar products for complex-valued mesh functions:

For any two M-vectors, X^n, Y^n, let their scalar product be defined by

$$\langle X^n, Y^n \rangle = \Delta x \sum_{k=1}^{M} x_k^n \bar{y}_k^n \tag{1.10}$$

and let the corresponding norm be

$$\langle X^n, X^n \rangle = \Delta x \sum_{k=1}^{M} |x_k^n|^2 = \|X^n\|_2^2. \tag{1.11}$$

For N-vectors ξ, ψ define

$$[\xi, \psi] = \Delta t \sum_{n=1}^{N} \xi^n \psi^n, \tag{1.12}$$

and

$$\|\xi\|_{2,N}^2 = \Delta t \sum_{n=1}^{N} |\xi^n|^2. \tag{1.13}$$

We will also use the norms:

$$\|X^n\|_\infty = \operatorname*{Max}_{j=1,\cdots,M} \{|x_j^n|\}, \tag{1.14}$$

$$\|V\|_\infty = \operatorname*{Max}_{n=1,\cdots,N} \{\|V^n\|_\infty\}, \tag{1.15}$$

$$\|V\|_2^2 = \Delta t \sum_{n=1}^{N} \|V^n\|_2^2. \tag{1.16}$$

For any square matrix A of appropriate size we define

$$\|A\| = \operatorname*{Sup}_{\|X\|=1} \|AX\| \tag{1.17}$$

the supremum being taken over all complex vectors.

Given a function $u(x, t)$, we sometimes write $u(t_0)$ to denote the function of x obtained from u when t is fixed at the value t_0. Also $u^n(x)$ stands for $u(x, n\Delta t)$.

2. Abstract Problems of Parabolic Type. Let H be a separable Hilbert space of complex valued functions defined on the open interval $0 < x < 1$ with scalar product (u, v) and corresponding norm $\|u\|_H$. Let $\|u\|_\infty$ be the essential supremum norm for such functions, and assume that there exists a constant K such that

$$\|u\|_H \leqq K\|u\|_\infty \quad \text{for every } u \in H. \tag{2.1}$$

Let A be a linear operator with domain and range contained in H, and let b_0, b_1 be linear boundary operators acting at $x = 0$, $x = 1$, respectively. Consider the eigenvalue problem

$$\begin{aligned} Av &= \lambda v, \qquad 0 < x < 1, \\ b_0 v &= b_1 v = 0. \end{aligned} \tag{2.2}$$

We assume that the problem (2.2) has a complete set of *orthonormal* eigenfunctions $\{\phi_k\}$ corresponding to strictly *positive* eigenvalues $\{\lambda_k\}$ with the property that

$$\sum_k \frac{\|\phi_k\|_\infty}{\lambda_k} < \infty. \tag{2.3}$$

Let R be the strip $\{(x, t) \mid 0 < x < 1, t > 0\}$ in the (x, t) plane, and let f be a real valued function on R such that $f(t) \in H$, as a function of x, for each fixed t.

Let $\chi(x)$ be a real valued function on $[0, 1]$ belonging to H, and let $\psi_0(t)$, $\psi_1(t)$ be defined and real for $t \geqq 0$. Consider the following abstract initial boundary-value problem on R, associated with the linear operator A:

Find a real valued function $u(x, t)$ defined on R such that for each fixed t, $u(t) \in$ the domain of A as a function of x, and u is differentiable as a function of t, for each fixed x, and

$$\begin{aligned} \frac{\partial u}{\partial t} &= -Au + f, \qquad 0 < x < 1, \quad t > 0, \\ u(x, 0) &= \chi(x), \qquad 0 \leqq x \leqq 1, \\ b_0 u &= \psi_0(t), \qquad b_1 u = \psi_1(t), \qquad t > 0. \end{aligned} \tag{2.4}$$

We assume that the above problem has a unique solution $u(x, t)$ which reaches a known steady state value $u^*(x)$ as $t \to \infty$, in such a way that $\|u(t) - u^*\|_H \to 0$ as $t \to \infty$, and so we speak of problems of parabolic type. Our main concern in this section is to describe a uniformly convergent semidiscrete finite-difference approximation to this abstract problem.

2.1. Semidiscrete Approximation to (2.4). Let $\Delta t > 0$ be a fixed "small" time increment. Let K_1 be a suitable positive constant. Choose T so that for some positive integer N we have

$$T = (N + 1)\Delta t \quad \text{and} \quad \|u(T) - u^*\|_H \leqq K_1 \Delta t^3. \tag{2.6}$$

Consider the following semidiscrete*** approximation to the analytic problem (2.4):

$$\begin{aligned} \frac{v^{n+1}(x) - v^{n-1}(x)}{2\,\Delta t} &= -Av^n(x) + f^n(x), \qquad n = 1, \ldots, N, \\ v^0(x) = \chi(x), &\qquad v^{N+1}(x) = u^*(x), \\ b_0 v^n = \psi_0^n, &\qquad b_1 v^n = \psi_1^n, \qquad n = 1, \ldots, N. \end{aligned} \tag{2.7}$$

The system of linear Eqs. (2.7) is an approximation to the analytic problem in the following sense:

*** Semidiscrete approximations, where only the time is discretized, have been considered from time to time in the literature. In Varga [17, p. 279], the author notes that such a procedure was used by Hartree and Womersley in 1937 to obtain a numerical solution to the heat equation; in Garabedian [9, p. 493], they are used to prove the *existence* of a solution to the heat equation and the author remarks on the connection with methods in the abstract theory of semigroups.

If $u(x, t)$ is the solution to (2.4), then u satisfies the equations

$$\begin{aligned} &\frac{\hat{u}^{n+1}(x) - \hat{u}^{n-1}(x)}{2\Delta t} = -A\hat{u}^n + f^n + \tau^n, \qquad n = 1,\dots,N,\\ &u^n = \hat{u}^n, \qquad n = 1,\dots,N,\\ &\hat{u}^0(x) = \chi(x), \qquad \hat{u}^{N+1}(x) = u^*(x),\\ &b_0\hat{u}^n = \psi_0^n, \qquad b_1\hat{u}^n = \psi_1^n, \qquad n = 1,\dots,N, \end{aligned} \tag{2.8}$$

where $\tau^n(x)$ is an error term. For $n = 1,\dots, N - 1$, $\tau^n(x)$ is the "truncation error" $-(\partial u/\partial t)^n + [(u^{n+1}(x) - u^{n-1}(x))/2\Delta t]$.

For $n = N$,

$$\tau^N(x) = \frac{u^*(x) - u^{N-1}(x)}{2\Delta t} - \left(\frac{\partial u}{\partial t}\right)^N = \frac{u(T) - u^{N-1}(x)}{2\Delta t} - \left(\frac{\partial u}{\partial t}\right)^N + \frac{u^* - u(T)}{2\Delta t}.$$

We will assume that u is such that

$$\|\tau^n\|_H \leqq K_4\,\Delta t^2, \qquad n = 1,\dots,N, \quad K_4 = \text{constant}. \tag{2.9}$$

For example, this condition will be satisfied if $u(x, t)$ has bounded continuous third-order time derivatives on R, and T is chosen so that $\|u(T) - u^*\|_H \leqq K_1\,\Delta t^3$.

Because $\|\tau^n\|_H \to 0$ as $\Delta t \to 0$, we say that (2.7) is *consistent* with the analytic problem. We rewrite (2.7) as

$$\begin{bmatrix} A & \sigma & & \bigcirc \\ -\sigma & \ddots & \ddots & \\ & \ddots & \ddots & \sigma \\ \bigcirc & & -\sigma & A \end{bmatrix} \begin{bmatrix} v^1(x) \\ \vdots \\ v^N(x) \end{bmatrix} = \begin{bmatrix} v^0(x)/2\Delta t + f^1(x) \\ f^2(x) \\ \vdots \\ f^N(x) - u^*(x)/2\Delta t \end{bmatrix}, \qquad 0 < x < 1, \tag{2.10}$$

with $b_0v^n = \psi_0^n$, $b_1v^n = \psi_1^n$, $n = 1, \dots, N$ and where $\sigma = 1/(2\Delta t)$. Since we assume that $u^*(x)$ is known a priori, the right-hand side of (2.10) is known. Having effectively replaced the problem (2.4) by a coupled system of linear equations for N functions of x, we must consider two questions:

(a) Does the system (2.10) have a solution? Is it unique?

(b) Does the solution of (2.10) converge to that of (2.4) as $\Delta t \to 0$? If so, in which norm and at what rate does this convergence take place? We will show the following:

THEOREM. *The system* (2.10) *has a unique solution* $V(\Delta t) = \{v^1(x), \dots, v^N(x)\}^T$. *Moreover, if* U *is the exact solution to* (2.4) *on the lines* $t = n\Delta t$, *i.e.* $U = \{u^1(x), u^2(x), \dots, u^N(x)\}^T$, *then*

$$\|V(\Delta t) - U\|_\infty \equiv \max_{n=1,\cdots,N} \|v^n - u^n\|_\infty \leqq K_0\,\Delta t^2,$$

so that $V(\Delta t)$ *converges uniformly to* U *at the rate of* $O(\Delta t^2)$ *as* $\Delta t \to 0$, $T \to \infty$.

We begin our analysis with the following key result.

Let $T = (N + 1)\Delta t$, $\lambda_j > 0$, $\sigma_j = 1/2\lambda_j\Delta t$ and consider the following $N \times N$ matrix $T_N(\sigma_j)$:

$$T_N(\sigma_j) = \begin{bmatrix} 1 & \sigma_j & & \bigcirc \\ -\sigma_j & \ddots & \ddots & \\ & \ddots & \ddots & \sigma_j \\ \bigcirc & & -\sigma_j & 1 \end{bmatrix}.$$

If we define

$$\|T_N\|_\infty = \operatorname*{Sup}_{\|X\|_\infty = 1} \|T_N X\|_\infty,$$

it is well known that then

$$\|T_N\|_\infty = \operatorname*{Max}_{i=1,\cdots,N} \sum_{j=1}^{N} |a_{ij}|,$$

where a_{ij} is the element in the ith row and jth column of the matrix T_N.

LEMMA 2.1. *$T_N(\sigma_j)$ is always invertible and $\|T_N^{-1}\|_\infty$ remains bounded independently of N, σ_j, as $N \to \infty$, $\Delta t \to 0$, $\lambda_j \to \infty$. In fact, if t_{sr} denotes an element in the sth row, rth column of $T_N^{-1}(\sigma_j)$, we have*

$$\text{(a)} \qquad |t_{sr}| \leqq 4(1 + \sigma_j^2)^{-1/2}\exp\left(-\frac{\lambda_j|s - r|\Delta t}{1 + 2\lambda_j\Delta t}\right)\left[1 - \exp\left(-\frac{2\lambda_j T}{1 + 2\lambda_j\Delta t}\right)\right]^{-1}$$

$$\text{(b)} \quad \sum_{r=1}^{N} |t_{sr}| \leqq 4(1 + 2\lambda_j\Delta t)(1 + \lambda_j\Delta t^2)^{-1/2}\left[1 - \exp\left(-\frac{2\lambda_j T}{1 + 2\lambda_j\Delta t}\right)\right]^{-1}$$
$$\times\left[2 + \Delta t - \exp\left(-\frac{\lambda_j s\Delta t}{1 + 2\lambda_j\Delta t}\right) - \exp\left(-\frac{\lambda_j(T - s\Delta t)}{1 + 2\lambda_j\Delta t}\right)\right]$$

Proof. We prove this lemma by explicitly computing $T_N^{-1}(\sigma_j)$.

The determinant Δ_N of T_N satisfies the recurrence relation

$$\Delta_{n+1} = \Delta_n + \sigma^2 \Delta_{n-1}, \qquad n = 1, \ldots, N - 1, \quad (\sigma = \sigma_j)$$

with

$$\Delta_1 = 1 \quad \text{and} \quad \Delta_0 = 1.$$

Hence, if $\alpha = \frac{1}{2} + \frac{1}{2}(1 + 4\sigma^2)^{1/2}$, $\beta = \frac{1}{2} - \frac{1}{2}(1 + 4\sigma^2)^{1/2}$, are the two roots of $x^2 - x - \sigma^2 = 0$, we see that

$$\Delta_N = (\alpha^{N+1} - \beta^{N+1})/(\alpha - \beta) \quad \text{on using } \Delta_1 = \Delta_0 = 1.$$

Now the cofactor of the element a_{ij} of T_N is

$$A_{ij} = (-)^{i+j}\Delta_{i-1}\Delta_{N-j}(-\sigma)^{j-i}, \quad \text{if } j > i,$$
$$= (-)^{i+j}\Delta_{j-1}\Delta_{N-i}(\sigma)^{i-j}, \quad \text{if } i > j$$

and both formulae hold if $i = j$.

If t_{sr} is the element in the sth row rth column of T_N^{-1}, we then have

$$|t_{sr}| = \frac{\Delta_{r-1}}{\Delta_N}\sigma^{s-r}\Delta_{N-s} \quad \text{if } s > r,$$

$$= \frac{\Delta_{s-1}}{\Delta_N}\Delta_{N-r}\sigma^{r-s} \quad \text{if } s \leqq r.$$

Since $\|T_N^{-1}\|_\infty = \mathrm{Max}_{s=1,\cdots,N} \sum_{r=1}^{N} |t_{sr}|$, we will first estimate $|t_{sr}|$ by means of these formulae and then proceed to estimate $\sum|t_{sr}|$.

Since α, β are roots of $x^2 - x - \sigma^2 = 0$ and $\alpha > 0$ whereas $\beta \leqq 0$, we have

$$\beta^2 = \sigma^2 + \beta \leqq \sigma^2, \qquad \alpha^2 = \sigma^2 + \alpha > \sigma^2.$$

Hence

$$\alpha - 1 = |\beta| \leqq \sigma < \alpha.$$

Consider first $|t_{sr}|$ for $s > r$. Using the formula for the determinants Δ_k,we obtain

$$\begin{aligned} |t_{sr}| &= \frac{1}{(1+4\sigma^2)^{1/2}} (\alpha^r - \beta^r) \frac{\sigma^s}{\sigma^r} \frac{\alpha^{N-s+1} - \beta^{N-s+1}}{\alpha^{N+1} - \beta^{N+1}} \\ &\leqq \frac{2\alpha^r}{(1+4\sigma^2)^{1/2}} \frac{\sigma^s}{\sigma^r} \frac{2\alpha^{N-s+1}}{\alpha^{N+1}(1 - |\beta|^{N+1}/\alpha^{N+1})} \\ &= \frac{4}{(1+4\sigma^2)^{1/2}} \left(\frac{\sigma}{\alpha}\right)^{s-r} \frac{1}{1 - (|\beta|/\alpha)^{N+1}}. \end{aligned}$$

Since $|\beta| = \alpha - 1$,

$$\left(\frac{|\beta|}{\alpha}\right)^{N+1} = \left(1 - \frac{1}{\alpha}\right)^{N+1} = \left(1 - \frac{2}{1 + (1+4\sigma^2)^{1/2}}\right)^{N+1}$$

and $(1 + 4\sigma^2)^{1/2} \leqq 1 + 2\sigma$ since $\sigma \geqq 0$. Hence

$$\left(1 - \frac{2}{1 + (1+4\sigma^2)^{1/2}}\right)^{N+1} \leqq \left(1 - \frac{1}{1+\sigma}\right)^{N+1} = \left(1 - \frac{(N+1)/(1+\sigma)}{N+1}\right)^{N+1},$$

which shows that $(|\beta|/\alpha)^{N+1} \leqq \exp(-(N+1)/(1+\sigma))$ using the well-known fact that for $0 \leqq x \leqq n$, $(1 - x/n)^n \leqq e^{-x}$.

Substituting $\sigma = \sigma_j = 1/2\lambda_j\Delta t$, $T = (N+1)\Delta t$, we obtain

$$\left(\frac{|\beta|}{\alpha}\right)^{N+1} \leqq \exp\left(-\frac{2\lambda_j T}{1 + 2\lambda_j\,\Delta t}\right).$$

Hence,

$$\frac{1}{1 - (|\beta|/\alpha)^{N+1}} \leqq \left[1 - \exp\left(-\frac{2\lambda_j T}{1 + 2\lambda_j\,\Delta t}\right)\right]^{-1}.$$

Let us now examine $(\sigma/\alpha)^{s-r}$. We have

$$\begin{aligned} (\sigma/\alpha)^{s-r} &= (1 - (\alpha - \sigma)/\alpha)^{s-r} \leqq (1 - 1/2\alpha)^{s-r} \quad \text{since } \alpha - \sigma \geqq \tfrac{1}{2} \\ &\leqq (1 - 1/(2 + 2\sigma))^{s-r} \quad \text{using } (1 + 4\sigma^2)^{1/2} \leqq 1 + 2\sigma. \end{aligned}$$

Hence, by a similar device,

$$\left(\frac{\sigma}{\alpha}\right)^{s-r} \leqq \exp\left(\frac{-\lambda_j(s-r)\Delta t}{1 + 2\lambda_j\Delta t}\right).$$

Now if $r \geqq s$, all formulae still hold with r and s interchanged. This concludes the proof of part (a) of the lemma. Let us now estimate $\sum_{r=1}^{N} |t_{sr}|$. We have

$$\sum_{r=1}^{N} |t_{sr}| \leqq 4\Delta t(\Delta t^2 + 4\sigma^2\Delta t^2)^{-1/2}\left[1 - \exp\left(\frac{-2\lambda_j T}{1 + 2\lambda_j\Delta t}\right)\right]^{-1}$$
$$\times \sum_{r=1}^{N} \exp\left(\frac{-\lambda_j|s - r|\Delta t}{1 + 2\lambda_j\Delta t}\right).$$

Let $\rho = 1/(1 + 2\lambda_j\Delta t)$, and consider

$$\Delta t \sum_{r=1}^{N} \exp[-\rho\lambda_j|s - r|\,\Delta t] = \Delta t \sum_{p=0}^{s-1} \exp[-\rho\lambda_j p\Delta t] + \Delta t \sum_{p=1}^{N-s} \exp[-\rho\lambda_j p\Delta t].$$

We may use a geometric argument (the integral test) to show

$$\Delta t \sum_{p=0}^{s-1} \exp[-\rho\lambda_j p\Delta t] = \Delta t + \Delta t \sum_{p=1}^{s-1} \exp[-\rho\lambda_j p\Delta t]$$
$$< \Delta t + \int_0^t \exp[-\rho\lambda_j u]du, \qquad t = s\Delta t,$$

and similarly

$$\Delta t \sum_{p=1} \exp[-\rho\lambda_j p\Delta t] < \int_0^{T-t} \mathrm{exk}[-\rho\lambda_j u]du, \qquad T = (N + 1)\Delta t.$$

Hence

$$\Delta t \sum_{r=1}^{N} \exp[-\rho\lambda_j|s - r|\,\Delta t] \leqq \frac{\Delta t + 2 - \exp[-\rho\lambda_j t] - \exp[-\rho\lambda_j(T - t)]}{\rho\lambda_j},$$

i.e., $\sum_{r=1}^{N} |t_{sr}|$ satisfies the estimate in part (b) of Lemma 2.1. Notice that as $\lambda_j \to \infty$,

$$1 - \exp\left(-\frac{2\lambda_j T}{1 + 2\lambda_j\Delta t}\right) \to 1 - e^{-T/\Delta t} = 1 - e^{-(N+1)}.$$

Clearly, the sum in part (b) is bounded as $\lambda_j \to \infty$, $\Delta t \to 0$, and the bound is independent of s. This proves the lemma.

Remark. In a subsequent discussion (in Section 4.1) we will also need the following result: Let $\lambda_1 \sim \beta\Delta t^2$ with β fixed $\neq 0$ as $\Delta t \to 0$. Then $\|T_N^{-1}(\sigma_1)\|_\infty$ remains bounded as $\Delta t \to 0$. We may see this as follows: since

$$|t_{sr}(\sigma_1)| \leqq 4(1 + \sigma_1^2)^{-1/2}\left[1 - \exp\left(\frac{-2\lambda_1 T}{1 + 2\lambda_1\Delta t}\right)\right]^{-1} \times \exp\left(\frac{-\lambda_1|s - r|\Delta t}{1 + 2\lambda_1\Delta t}\right)$$
$$\leqq 4(1 + \sigma_1^2)^{-1/2}\left[1 - \exp\left(\frac{-2\lambda_1 T}{1 + 2\lambda_1\Delta t}\right)\right]^{-1},$$

we have, on substituting $\lambda_1 = \beta\Delta t^2$ in the last expression,

$$|t_{sr}| \leqq 4\beta\Delta t^3\left[1 - \exp\left(\frac{-2\beta\Delta t^2 T}{1 + 2\beta\Delta t^3}\right)\right]^{-1},$$

and both the numerator and denominator of the last expression approach zero as $\Delta t \to 0$, if T is fixed. Differentiating with respect to Δt and using L'Hospital's rule, we obtain

$$\lim_{\Delta t\to 0}\frac{4\beta\Delta t^3}{1-e^{\frac{-2\beta\Delta t^2 T}{1+2\beta\Delta t^3}}}=\lim_{\Delta t\to 0}\frac{(12\beta\Delta t^2)(1+2\beta\Delta t^3)^2}{(4\beta T\Delta t-4\beta^2 T\Delta t^4)e^{\frac{-2\beta\Delta t^2 T}{1+2\beta\Delta t^3}}},$$

and, since the last expression tends to $3\Delta t/T = 3/(N+1)$ as $\Delta t \to 0$, we have

$$\sum_{r=1}^{N}|t_{sr}| \leqq (N+1)\operatorname*{Max}_{r,s}|t_{sr}| \to 3 \quad \text{as } \Delta t \to 0.$$

LEMMA 2.2. *The system of Eqs.* (2.10) *has a unique solution* $V(\Delta t)$.

Proof. Let M be the matrix of linear operators occurring in (2.10). In an obvious notation we may write (2.10) as

$$MV = F, \qquad b_0V = \psi_0, \qquad b_1V = \psi_1. \tag{2.11}$$

Observe that F is such that each of its components belongs to H. Because we have assumed the existence of a solution to (2.4), it is sufficient to prove that, given any G whose components $g^n(x)$ belong to H, $n = 1, \ldots, N$, the system

$$MV = G, \qquad b_0V = b_1V = 0$$

always has a unique solution. To do this, expand in the eigenfunctions of the problem (2.2) above. Set

$$v^n(x) = \sum_{j=1}^{\infty} c_j^n\phi_j, \qquad g^n(x) = \sum_{j=1}^{\infty} d_j^n\phi_j.$$

Then if $\sigma_j = 1/2\lambda_j\Delta t$, we obtain the following equations expressing the c_j^n in terms of the known d_j^n

$$\sigma_j(c_j^{n+1} - c_j^{n-1}) + c_j^n = \frac{d_j^n}{\lambda_j}, \qquad n = 1, \ldots, N, \quad j = 1, 2, \ldots$$

with

$$c_j^0 = c_j^{N+1} = 0 \quad \forall_j.$$

Hence if $T_N(\sigma_j)$ is the matrix of Lemma (2.1), we have

$$[T_N(\sigma_j)]\begin{bmatrix} c_j^1 \\ \vdots \\ c_j^N \end{bmatrix} = \frac{1}{\lambda_j}\begin{bmatrix} d_j^1 \\ \vdots \\ d_j^N \end{bmatrix}, \qquad j = 1, 2, \ldots. \tag{2.12}$$

Since $T_N(\sigma_j)$ is invertible for every j, (2.12) uniquely defines the c_j^n, so that the reduced problem above always has a unique solution. Q.E.D.

We are now ready to prove the convergence theorem of Section II.1.

Let $w^n = v^n - \hat{u}^n$, then $w^n(x)$ satisfies

$$w^0 = w^{N+1} = 0$$

$$\frac{w^{n+1} - w^{n-1}}{2\Delta t} = -Aw^n + \tau^n, \qquad b_0w^n = b_1w^n = 0, \quad n = 1, \ldots, N,$$

where $\tau^n(x) \in H$ and $\|\tau^n\|_H \leqq K_4\Delta t^2$. Setting

$$w^n = \sum_{j=1}^{\infty} c_j^n \phi_j, \qquad n = 1, \dots, N,$$

and

$$\tau^n = \sum_{j=1}^{\infty} d_j^n \phi_j, \qquad n = 1, \dots, N,$$

we have $|d_j^n| = |(\tau^n, \phi_j)| \leqq \|\tau^n\|_H \|\phi_j\|_H \leqq K_4 \Delta t^2$, and Eqs. (2.12) are satisfied for the c_j^n's.

By Lemma (2.1), $\|T_N^{-1}\|_\infty$ is bounded as $\Delta t \to 0$, $N \to \infty$, $\lambda_j \to \infty$. Hence

$$\operatorname*{Sup}_n |c_j^n| \leqq K_5 \Delta t^2 / \lambda_j.$$

Therefore,

$$\|w^n\|_\infty \leqq \sum_{j=1}^{\infty} |c_j^n| \|\phi_j\|_\infty \leqq K_5 \Delta t^2 \sum_{j=1}^{\infty} \frac{\|\phi_j\|_\infty}{\lambda_j}.$$

Since by assumption

$$\sum_j \frac{\|\phi_j\|_\infty}{\lambda_j} < \infty,$$

we have

$$\operatorname*{Sup}_n \|w^n\|_\infty \leqq K_6 \Delta t^2 \qquad K_6 = \text{constant},$$

and this proves the theorem.

Examples of such operators A are provided by regular Sturm-Liouville differential operators, operating in $H = L^2[0, 1]$. Thus for the problem†

$$(2.13) \qquad \begin{gathered} [a(x)u']' + b(x)u' - c(x)u + \lambda u = 0, \qquad 0 < x < 1, \\ u(0) = u(1) = 0 \end{gathered}$$

where $a(x) \geqq a_0 > 0$ and $c(x) \geqq 0$, it is known that the eigenvalues are real and form a countably infinite set, $\lambda_1 \leqq \lambda_2 \leqq \lambda_3 \leqq \dots$. Moreover $\lambda_1 > \inf_{0<x<1} c(x)$ (see [14, p. 37]).

It is a standard result that the eigenvalues of (2.13) can be characterized as the zeroes of an entire function [16, p. 190], and, as observed by Atkinson in [1], this function is of order at most 1/2 so that,

$$\sum_k \frac{1}{(\lambda_k)^{1/2+\varepsilon}} < \infty$$

for every $\varepsilon > 0$. Also, the *normalized* eigenfunctions may be shown to be uniformly bounded in the supremum norm, i.e.,

$$\|\phi_k\|_\infty \leqq \text{constant} \qquad [17, \text{p. } 335].$$

† A standard transformation, puts (2.13) in selfadjoint form.

We remark, however, that A may be a singular differential operator and still satisfy property (2.3). Thus consider in $L_2[0, 1]$, the problem

$$\begin{aligned} &Au \equiv -[(1 - x^2)u']' = \lambda u, \qquad 0 < x < 1, \\ &u(0) = 0, \qquad (1 - x)^2 u'(x) \to 0 \quad \text{as } x \uparrow 1. \end{aligned} \tag{2.14}$$

If

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n$$

are the Legendre polynomials for $n = 0, 1, 2, \ldots,$ then the eigenfunctions for this problem are

$$P_{2n+1}, \qquad n = 0, 1, 2, \ldots$$

corresponding to the eigenvalues

$$\lambda_n = (2n + 1)(2n + 2), \qquad n = 0, 1, 2, \ldots \qquad \text{(see [7])}.$$

The set $\{P_{2n+1}\}$ spans $L_2[0, 1]$, since the complete set of Legendre polynomials spans $L_2[-1, 1]$ and $P_n(x)$ is an even function if n is even. As defined above, the P_n are not normalized but satisfy $\|P_n\|_\infty = 1$ for $|x| \leqq 1$. However, if

$$\tau_n(x) = ((2n + 1)/4)^{1/2} P_n(x), \qquad n = 0, 1, 2, \ldots,$$

then the τ_n are orthonormal on (0, 1) and

$$\|\tau_n\|_\infty = ((2n + 1)/4)^{1/2},$$

thus

$$\sum_{n \text{ odd}} \frac{\|\tau_n\|_\infty}{\lambda_n} < \infty.$$

Finally, we remark that although we have emphasized one-dimensional problems, similar problems may be formulated in R^n with H, for example, being a Sobolev space of functions on some bounded domain Ω and A a uniformly elliptic operator of sufficiently high order.

3.1. Linear Parabolic Initial Boundary Problems: Fully Discrete Methods. We are concerned here with the numerical computation of problems of the following kind:

$$\begin{aligned} &\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left[a(x)\frac{\partial u}{\partial x}\right] + b(x)\frac{\partial u}{\partial x} - c(x)u + h(x, t), \qquad 0 < x < 1, t > 0 \\ &u(x, 0) = \chi(x), \qquad 0 \leqq x \leqq 1 \\ &u(0, t) = \phi_1(t), \qquad u(1, t) = \phi_2(t) \\ &\chi(0) = \phi_1(0), \qquad \chi(1) = \phi_2(0). \end{aligned} \tag{3.1}$$

We assume that $a(x) \geqq a_0 > 0$ and $c(x) \geqq 0$, a having three or more continuous derivatives and b one or more continuous derivatives in R. We assume further that $a, b, c, h, \chi, \phi_1, \phi_2$ are bounded and sufficiently smooth that a solution $u(x, t)$ exists having three continuous time derivatives and four continuous space derivatives. All of the above mentioned derivatives as well as u itself will be assumed bounded on R. (For existence, uniqueness and regularity theorems for parabolic equations, consult Friedman [8].) As in the previous section, the amount of smoothness that we assume will *suffice* for the truncation error τ^n to be of the order of $\Delta t^2 + \Delta x^2$ in the discrete L_2 norm and that is all we need. Thus, our assumptions may be weakened somewhat. We assume that h, ϕ_1, ϕ_2 reach a steady state as $t \to \infty$. Since $c(x) \geqq 0$, $u(x, t)$ converges to a steady value $u^*(x)$ (see Friedman [8, Chapter 6]) and we assume this convergence to be uniform in x. We may suppose that $u^*(x)$ is known without loss of generality. Indeed, we only require its values at mesh points, and these can be obtained with sufficient accuracy by existing numerical techniques, since $u^*(x)$ satisfies an inhomogeneous boundary-value problem for an ordinary differential equation. Finally, we assume that, given any $\varepsilon > 0$, it is possible to estimate how large T must be chosen so that

$$\|u(T) - u^*\|_\infty < \varepsilon$$

e.g. by means of asymptotic formulae.

3.2. Discrete Approximation to the Analytic Problem. Choose T so that for some positive integer N we have $T = (N + 1)\Delta t$ and

$$\|u(T) - u^*\|_2 \equiv \left\{\Delta x \sum_{k=1}^{M} |u(k\Delta x, T) - u^*(k\Delta x)|^2\right\}^{1/2} \leqq K\Delta t^3,$$

where K is a fixed positive constant independent of Δt. Introduce a mesh over R_T as in Section 1.2.

Our finite-difference approximation to (3.1) will be

$$\frac{v_k^{n+1} - v_k^{n-1}}{2\Delta t} = \frac{a_{k+1/2}(v_{k+1}^n - v_k^n) - a_{k-1/2}(v_k^n - v_{k-1}^n)}{\Delta x^2} + \frac{b_k(v_{k+1}^n - v_{k-1}^n)}{2\Delta t}$$

$$(3.2) \qquad - c_k v_k^n + h_k^n, \qquad n = 1, \ldots, N, \quad k = 1, \ldots, M$$

$$\text{with } v_k^0 = \chi(k\Delta x), \qquad v_k^{N+1} = u^*(k\Delta x), \qquad k = 0, 1, \ldots, M + 1,$$

$$v_0^n = \phi_1(n\Delta t), \qquad v_{M+1}^n = \phi_2(n\Delta t), \qquad n = 1, \ldots, N.$$

As before, this approximation is consistent with the problem (3.1), i.e., the exact solution u satisfies (3.2) if we add an error term τ_k^n on the right-hand side. For $n = 1, \ldots, N - 1$, τ_k^n is the truncation error due to replacing derivatives by finite-difference quotients. For $n = N$, there is an additional error due to prescribing $u^*(x)$ on the line $t = T$ instead of the exact solution $u(T)$. Since we chose T large enough, we have the estimate

$$\|\tau^n\|_2 \equiv \left\{\Delta x \sum_{k=1}^{M} |\tau_k^n|^2\right\}^{1/2} \leqq K_0(\Delta x^2 + \Delta t^2),$$

where K_0 is a fixed constant independent of Δx, Δt, and n. Let

$$\alpha_k = [a_{k+1/2} + a_{k-1/2}] + c_k \Delta x^2, \qquad \beta_k = -[a_{k+1/2} + b_k \Delta x/2]$$

and

$$\gamma_k = [b_k \Delta x/2 - a_{k-1/2}], \qquad k = 1, \ldots, M$$

and define the tridiagonal matrix L of order M by

$$L = \frac{1}{\Delta x^2}\begin{bmatrix} \alpha_1 & \beta_1 & & \bigcirc \\ \gamma_2 & \ddots & \ddots & \\ & \ddots & \ddots & \beta_{M-1} \\ \bigcirc & & \gamma_M & \alpha_M \end{bmatrix}$$

We may write the approximation (3.2) in the form

$$(V^{n+1} - V^{n-1})/2\Delta t + LV^n = F^n, \qquad n = 1, \dots, N, \tag{3.3}$$

where V^0, V^{N+1} are given and F^n is an M-vector containing the known lateral boundary data and the inhomogeneous term h_k^n. We may also write (3.3) in "block" form. Let M be the $MN \times MN$ block tridiagonal matrix

$$M = \begin{bmatrix} L & \sigma I & & \bigcirc \\ -\sigma I & \ddots & \ddots & \\ & \ddots & \ddots & \sigma I \\ \bigcirc & & -\sigma I & L \end{bmatrix},$$

where I is the $M \times M$ unit matrix and $\sigma = 1/2\Delta t$. Then we have

$$MV = F, \tag{3.4}$$

where F consists now of data at the four boundaries as well as the inhomogeneous term $h(x, t)$.

LEMMA 3.1. *There exists a nonsingular diagonal matrix D such that $D^{-1}LD = \hat{L}$ is a real symmetric matrix and $\|D\|_\infty$, $\|D^{-1}\|_\infty \leqq K_0 < \infty$ as $M \to \infty$, $\Delta x \to 0$, $(M + 1)\Delta x = 1$.*

Proof. See [4], [5].

Remark. The change of variables $X = DZ$ is the discrete analog of the transformation

$$u(x) = \exp\left(-\frac{1}{2}\int_0^x \frac{b(t)}{a(t)}\,dt\right)v(x)$$

which puts the linear differential operator

$$\mathscr{L}[u] \equiv -(au')' - bu' + cu, \qquad 0 < x < 1$$
$$u(0) = u(1) = 0$$

into the selfadjoint form

$$\hat{\mathscr{L}}[v] \equiv -(av')' + [c + \tfrac{1}{2}b' + \tfrac{1}{4}b^2/a]v, \qquad 0 < x < 1,$$
$$v(0) = v(1) = 0.$$

LEMMA 3.2. *The eigenvalues of L are strictly positive and remain bounded away from zero as $M \to \infty$, $\Delta x \to 0$, $(M + 1)\Delta x = 1$. Let $\{\lambda_j\}_{j=1}^M$ be the eigenvalues of L arranged in increasing order. Then there exists a positive integer j_0, independent of M as $M \to \infty$, such that for all $j_0 < j \leqq M$, we have*

$$c_1 j^2 \leqq \lambda_j \leqq c_2 j^2 \quad \textit{where } c_1, c_2$$

are positive constants.

Finally, let V^j *be the eigenvector of* $\hat{L}$ *corresponding to the eigenvalue* λ_j *and normalized so that*

$$\Delta x \sum_{k=1}^{M} |v_k^j|^2 = 1.$$

Then, there exists a positive constant K_0 *and a positive integer* j_1, *independent of* M, *such that for all* $j_1 < j \leqq M$,

$$\|V^j\|_\infty = \operatorname*{Sup}_{k=1,\cdots,M} |v_k^j| \leqq K_0(j)^{1/2}.$$

Proof. In the selfadjoint case (i.e. $b(x) \equiv 0$) these results are to be found in Bückner [3]. In this more general case, the lemma follows from the discrete maximum principle, and from Lemma 3.1 together with Bückner's argument. Another proof may be found in [4] and [5], which proceeds via a discrete analog of the *Sturm comparison theorem.*

THEOREM. *Let* $\mu^2 = (\Delta t^2 + \Delta x^2)$ *and let* $\{V^n\}_{n=1}^N$ *be the solutions of Eqs.* (3.3), *or equivalently of* (3.4). *Let* $\{U^n\}_{n=1}^N$ *be the vector obtained from evaluations of* $u(x, t)$ *at the mesh points. Finally, assume*

$$\|\tau^n\|_2 \leqq K_1 \mu^2. \tag{3.5}$$

Then, there is a constant K_2 *such that*

$$\|V^n - U^n\|_\infty \leqq K_2 \mu^2.$$

Proof. The argument is very similar to the proof of the main theorem of the preceding section. Let $W = V - U$. With D the diagonal matrix which symmetrizes L, let $X^n = D^{-1}W^n$ and substitute into (3.3), to obtain

$$\frac{X^{n+1} - X^{n-1}}{2\Delta t} + \hat{L}X^n = D^{-1}\tau^n, \qquad n = 1, \ldots, N, \quad X^0 = X^{N+1} = 0. \tag{3.6}$$

Since $\hat{L}$ is real symmetric, it has a complete set of orthonormal eigenvectors Z^j, $j = 1, \ldots, M$. We may solve by expanding in terms of the Z^j. Thus if

$$X^n = \sum_{j=1}^{M} c_j^n Z^j, \qquad D^{-1}\tau^n = \sum_{j=1}^{M} d_j^n Z^j,$$

we obtain on substituting into (3.6), MN equations for the coefficients c_j^n:

$$\frac{c_j^{n+1} - c_j^{n-1}}{2\Delta t} + \lambda_j c_j^n = d_j^n, \qquad n = 1, \ldots, N, \quad j = 1, \ldots, M, \tag{3.7}$$

where

$$c_j^0 = c_j^{N+1} = 0, \qquad j = 1, \ldots, M.$$

Let $\sigma_j = 1/2\lambda_j \Delta t$; then

$$[T_N(\sigma_j)]\begin{bmatrix} c_j^1 \\ \vdots \\ c_j^N \end{bmatrix} = \frac{1}{\lambda_j}\begin{bmatrix} d_j^1 \\ \vdots \\ d_j^N \end{bmatrix}, \qquad j = 1, \dots, M, \tag{3.8}$$

where $T_N(\sigma_j)$ is the $N \times N$ matrix of Lemma 2.1. The proof now follows from Lemma 3.2 and the fact that

$$\sum_{j=1}^{\infty} (j)^{-3/2} < \infty.$$

4.1. An Example. Consider now the problem

$$\begin{aligned} &u_t = u_{xx} + \pi^2 u + \sin \pi x \cos t, \qquad 0 < x < 1, \quad t > 0 \\ &u(x, 0) = 0, \qquad 0 \leqq x \leqq 1, \qquad u(0, t) = u(1, t) = 0, \qquad t \geqq 0. \end{aligned} \tag{4.1}$$

This problem has the unique solution $u = \sin \pi x \sin t$. It differs from the class of problems considered in the previous sections in that $c(x)$ is negative and the related Sturm-Liouville problem has the eigenvalue $\lambda = 0$. Nevertheless, since $u \equiv 0$ at $t = \pi$ and at $t = 2\pi$, we may select *either* of these lines as the line $t = T$ and prescribe the *exact* solution $u \equiv 0$ on $t = T$ in our difference approximation to (4.1). Thus, if H is the tridiagonal matrix of order M given by

$$H = \frac{1}{\Delta x^2}\begin{bmatrix} 2 & -1 & & \bigcirc \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ \bigcirc & & -1 & 2 \end{bmatrix} \tag{4.2}$$

with eigenvalues $0 < h_1 < h_2 < \dots < h_M$ and if W^1 is the M-vector $w_k^1 = \sin k\pi \Delta x$, $k = 1, \dots, M$, our approximation may be written as

$$\begin{aligned} &\frac{V^{n+1} - V^{n-1}}{2\Delta t} + (H - \pi^2 I)V^n = W^1 \cos n\Delta t, \qquad n = 1, \dots, N \\ &V^0 = V^{N+1} = 0. \end{aligned} \tag{4.3}$$

On expanding in eigenvectors of H, we easily see that (4.3) has the unique solution

$$V^n = c^n W^1,$$

where the c^n's satisfy

$$\begin{aligned} &\frac{c^{n+1} - c^{n-1}}{2\Delta t} + (h_1 - \pi^2)c^n = \cos n\Delta t, \qquad n = 1, \dots, N \\ &c^0 = c^{N+1} = 0. \end{aligned} \tag{4.4}$$

The computation of this example was attempted by Greenspan in [10] with $T = 2\pi$. However, he was not able to solve the system of difference equations by point successive over-relaxation for any value of ω. Apart from that, the above example has another interesting property: As we shall see, it makes a difference whether one selects $T = \pi$ or $T = 2\pi$. With $T = \pi$, the unique solution V of the system

(4.3) (even though it remains uniformly bounded as Δx, $\Delta t \to 0$, with $\Delta x = O(\Delta t)$) *does not converge to the analytic solution U, unless $N \to \infty$ through even integers.*

We begin with a few observations. The systems of difference equations occurring in (3.4) and (4.3) are special cases of the system

$$QV = F, \tag{4.5}$$

where Q is a block tridiagonal matrix of the form

$$Q = \begin{bmatrix} \Lambda & \sigma I & & \bigcirc \\ -\sigma I & \ddots & \ddots & \\ & \ddots & \ddots & \sigma I \\ \bigcirc & & -\sigma I & \Lambda \end{bmatrix} \tag{4.6}$$

with $\sigma = 1/2\Delta t$ and Λ a nonsingular $M \times M$ matrix with distinct real eigenvalues, λ_j, $j = 1, \ldots, M$.

LEMMA 4.1. *Let the eigenvalues of Λ be ordered so that $|\lambda_1| \leqq |\lambda_2| \leqq \cdots \leqq |\lambda_M|$, and let Y^j, $j = 1, \ldots, M$ be the corresponding eigenvectors. For fixed s, j, define the M vector $X^n_{s,j}$ by*

$$X^n_{s,j} = i^n \left[\operatorname{Sin} s\left(\frac{n\pi}{N+1}\right)\right] Y^j, \qquad n = 1, \ldots, N. \tag{4.7}$$

Let $X_{s,j}$ be the block vector

$$X_{s,j} = \{X^1_{s,j}, X^2_{s,j}, \ldots, X^N_{s,j}\}, \qquad s = 1, \ldots, N; \quad j = 1, \ldots, M. \tag{4.8}$$

Let

$$\mu_s = \frac{i}{\Delta t} \cos s\left(\frac{\pi}{N+1}\right), \qquad s = 1, \ldots, N, \tag{4.9}$$

then the $X_{s,j}$ are eigenvectors of Q corresponding to the eigenvalues

$$\Theta_{s,j} = (\lambda_j + \mu_s), \qquad s = 1, \ldots, N; \quad j = 1, \ldots, M, \tag{4.10}$$

respectively.

Proof. Direct verification.

LEMMA 4.2. *Let Q be the matrix of* (3.4), *i.e. with $\Lambda = L$. Then*

$$\|Q^{-1}\|_2 \leqq \text{constant as } \Delta x,\ \Delta t \to 0. \tag{4.11}$$

For the system (4.3), *i.e. with $\Lambda = H - \pi^2 I$, we have*

$$\|Q^{-1}\|_2 \to \infty, \quad \text{as } \Delta t \to 0,\ N \to \infty \text{ through odd integers} \tag{4.12}$$

$$\|Q^{-1}\|_2 \leqq \text{constant as } \Delta t \to 0,\ N \to \infty \text{ through even integers.} \tag{4.13}$$

Proof. For the system (3.4), (4.11) follows from the fact that because of Lemma 3.1, there exists a nonsingular diagonal matrix P of order MN such that $P^{-1}QP$ is a normal matrix, and $\|P\|_2$, $\|P^{-1}\|_2 \leqq$ constant, and from Lemmas 4.1 and 3.2. For the system (4.3) we first note that the eigenvalues h_j of H are given by

$$h_j = \frac{4}{\Delta x^2} \operatorname{Sin}^2 \frac{j\pi\Delta x}{2}, \qquad j = 1, \dots, M, \tag{4.14}$$

(see [2, p. 66]) and an elementary calculation shows that

$$\pi^2 + O(\Delta x^2) = h_1 < \pi^2. \tag{4.15}$$

Since $\mu_{((N+1)/2)} = 0$ whenever N is odd, the eigenvalue of Q which is smallest in absolute value for N odd is $\lambda_1 = h_1 - \pi^2 < 0$, and hence

$$\|Q^{-1}\|_2 = \frac{1}{|\lambda_1|} = O\left(\frac{1}{\Delta x^2}\right) \to \infty \tag{4.16}$$

as $\Delta t \to 0$, N odd, $\Delta x = O(\Delta t)$.

On the other hand, if N is even, the eigenvalue of Q smallest in absolute value is

$$\lambda_1 \pm \frac{i}{\Delta t} \operatorname{Sin} \frac{\pi}{2(N+1)} \quad \text{and} \quad \Delta t = \frac{\pi}{N+1} \quad \text{or} \quad \frac{2\pi}{N+1}$$

depending on whether we choose $T = \pi$ or $T = 2\pi$ in problem (4.1). In either case, $\|Q^{-1}\|_2$ remains bounded.

LEMMA 4.3. *Let S be the skew-symmetric $N \times N$ matrix*

$$S = \frac{1}{2\Delta t} \begin{bmatrix} 0 & 1 & & \bigcirc \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ \bigcirc & & -1 & 0 \end{bmatrix},$$

then S has the distinct eigenvalues

$$\mu_s = \frac{i}{\Delta t} \operatorname{Cos} \frac{s\pi}{N+1}, \qquad s = 1, \dots, N \tag{4.17}$$

with corresponding eigenvectors $\xi_s = \{\xi_s^n\}$, $s = 1, \dots, N$, where

$$\xi_s^n = i^n \operatorname{Sin} \frac{ns\pi}{N+1}, \qquad n = 1, \dots, N. \tag{4.18}$$

If N is odd, $\mu_{((N+1)/2)}$ is the only zero eigenvalue and the corresponding eigenvector may be taken to be

$$\psi_{((N+1)/2)} = \delta\{1, 0, 1, 0, \dots, 1\}^T \tag{4.19}$$

with δ chosen so that $\|\psi_{(N+1)/2}\|_{2,N} = 1$.

Proof. Direct verification.

THEOREM. *Let $V(\Delta t) = \{V^n\}_{n=1}^N$ be the solution of* (4.3) *and let $\Delta t = \gamma\,\Delta x$, where γ is a positive constant as Δx, $\Delta t \to 0$, and let U be the solution to* (4.1) *at the mesh points. Then*

(1) *$\|V(\Delta t)\|_\infty$ remains bounded as $\Delta t \to 0$.*

(2) *V does not converge to U, if $T = \pi$ unless $N \to \infty$ through even integers.*

(3) *Even though $\|Q^{-1}\|_2 \to \infty$ as $\Delta t \to 0$, N odd, $\|V(\Delta t) - U\|_\infty \to 0$ as $\Delta t \to 0$ provided $T = 2\pi$.*

Proof. Let $E^n = U^n - V^n$. Then E^n satisfies

$$\begin{aligned} &\frac{E^{n+1} - E^{n-1}}{2\Delta t} + (H - \pi^2 I)E^n = \tau^n, \qquad n = 1, \ldots, N \\ &E^0 = E^{N+1} = 0 \end{aligned} \tag{4.20}$$

where τ^n is the truncation error and $\|\tau^n\|_2 = O(\Delta t^2)$.

Expanding in the orthonormal eigenvectors of H, we are led to a system of MN linear equations for the Fourier coefficients c_j^n of E^n, in terms of the coefficients d_j^n of τ^n, viz.

$$[T_N(\sigma_j)] \begin{bmatrix} c_j^1 \\ \vdots \\ c_j^N \end{bmatrix} = \frac{1}{h_j - \pi^2} \begin{bmatrix} d_j^1 \\ \vdots \\ d_j^N \end{bmatrix}, \qquad j = 1, \ldots, M, \tag{4.21}$$

where T_N is the matrix of Lemma 2.1 with

$$\sigma_j = \frac{1}{2\Delta t(h_j - \pi^2)}.$$

From the fact that the eigenvalues h_j of H are distinct and $h_j \to j^2\pi^2$ as $\Delta x \to 0$, j fixed, we have

$$h_j - \pi^2 \geqq h_0 > 0 \quad \text{for all } j \geqq 2$$

if Δx is sufficiently small. Moreover, since $\sin^2\theta/\theta^2 \geqq 4/\pi^2$ for $0 \leqq \theta \leqq \pi/2$, we have from (4.14),

$$h_j - \pi^2 \geqq 2j^2 \quad \text{for } j \text{ sufficiently large.} \tag{4.22}$$

Furthermore, if $\tilde{W}^j$, $j = 1, \ldots, M$ are the orthonormal eigenvectors of H, then it is well known that $\|\tilde{W}^j\|_\infty \leqq$ constant.

Using the estimate

$$\max_{n=1,\cdots,N} |c_j^n| \leqq \|T_N^{-1}(\sigma_j)\|_\infty O(\Delta t^2)/(h_j - \pi^2), \qquad j = 1, \ldots, M$$

obtained by inverting (4.21) and using Lemma 2.1, we have

$$\|E^n\|_\infty \leqq \sum_{j=1}^{M} |c_j^n| \|\tilde{W}^j\|_\infty \leqq O(\Delta t^2)\left(\sum_{j=1}^{M} \frac{\|\tilde{W}^j\|_\infty}{h_j - \pi^2}\right)$$

or

$$\|E^n\|_\infty \leqq \frac{O(\Delta t^2)\|\tilde{W}^1\|_\infty}{h_1 - \pi^2} + O(\Delta t^2),$$

since $\sum_{j=2}^{M} \|\tilde{W}^j\|_\infty/(h_j - \pi^2)$ is bounded independently of M. Thus

$$\sup_{n=1,\cdots,N} \|E^n\|_\infty \leqq \text{constant}, \tag{4.22}$$

because $h_1 - \pi^2 = O(\Delta x^2)$, and we assume $\Delta x = O(\Delta t)$. Since the exact solution U is bounded, it follows from (4.22) that

$$\|V(\Delta t)\|_\infty \leqq \text{constant} \quad \text{as } \Delta t \to 0.$$

This proves the first part.

Let us now examine the convergence of $V(\Delta t)$ to U. Let I_N be the $N \times N$ unit matrix.

Let $\mathbf{c} = \{c^n\}_{n=1}^N$ and $\mathbf{p} = \{p^n\}_{n=1}^N$, where $p^n = \cos n\Delta t$. Then, Eq. (4.4) takes the form

$$[S + (h_1 - \pi^2)I_N]\mathbf{c} = \mathbf{p}, \quad \text{where } S \text{ is the matrix of Lemma 4.3.} \tag{4.23}$$

Let N be odd. Whether $T = \pi$ or $T = 2\pi$, we have

$$[\mathbf{p}, \psi_{(N+1)/2}] = \Delta t \sum_{n=1}^N \cos n\Delta t \psi_{(N+1)/2}^n = 0.$$

Hence, if we solve (4.23) by expanding in the orthonormal eigenvectors ψ_s of S, we see immediately that the solution $\mathbf{c}$ satisfies

$$[\mathbf{c}, \psi_{(N+1)/2}] = 0. \tag{4.24}$$

Suppose now that

$$\|V(\Delta t) - U\|_2 \to 0 \quad \text{as } N \to \infty. \tag{4.25}$$

Since $U^n = \sin n\Delta t W^1$, this means that

$$\Delta t \sum_{n=1}^N |c^n - \sin n\Delta t|^2 \to 0 \quad \text{as } N \to \infty. \tag{4.26}$$

However, if $T = \pi$, $\sin t$ is positive on $(0, \pi)$ and

$$\Delta t \sum_{n=1}^N \psi_{((N+1)/2)}^n \sin n\Delta t \geqq \beta > 0$$

and therefore, using (4.24),

$$\Delta t \sum_{n=1}^N (\sin n\Delta t - c^n)\psi_{((N+1)/2)}^n \geqq \beta > 0.$$

By Schwarz's inequality

$$0 < \beta \leqq \|\psi_{((N+1)/2)}\|_{2,N} \left\{\Delta t \sum_{n=1}^N |c^n - \sin n\Delta t|^2\right\}^{1/2},$$

so that (4.26) is impossible, if $T = \pi$ and N is odd. In fact, $V(\Delta t)$ cannot converge to U in any of the previously defined norms, since this would imply (4.25).

On the other hand, if $T = 2\pi$, N odd, then we have

$$\Delta t \sum_{n=1}^N \psi_{(N+1)/2}^n \sin n\Delta t = 0. \tag{4.27}$$

Let $\mathbf{b}$ be the N vector $\{b^n\}$ with

$$b^n = \frac{\Delta t}{\sin \Delta t} \sin n\Delta t, \qquad n = 1, \ldots, N.$$

Then it is easily verified that $\mathbf{b}$ satisfies $S\mathbf{b} = \mathbf{p}$.

Using (4.23), we then have

$$S(\mathbf{b} - \mathbf{c}) = (h_1 - \pi^2)\mathbf{c}. \tag{4.28}$$

Expanding $\mathbf{b} - \mathbf{c}$ in the orthonormal eigenvectors ψ_s of S, we have

$$\mathbf{b} - \mathbf{c} = \sum_{s=1}^{N} a_s \psi_s.$$

Observe that by (4.24) and (4.17), we have $a_{(N+1)/2} = 0$. Since $S(\mathbf{b} - \mathbf{c}) = \sum_s a_s \mu_s \psi_s$, we have from (4.28)

$$\|S(\mathbf{b} - \mathbf{c})\|_{2,N}^2 = \sum_s |a_s|^2 |\mu_s|^2 = (h_1 - \pi^2)^2 \|\mathbf{c}\|_{2,N}^2 \leqq K\Delta t^4, \tag{4.29}$$

where K is a constant, because $\|\mathbf{c}\|_{2,N}$ is bounded and $(h_1 - \pi^2) = O(\Delta t^2)$.

Also, for $s = 1, \ldots, N$, $s \neq (N + 1)/2$, the eigenvalues μ_s of S which are smallest in absolute value are given by

$$\mu = \pm \frac{i}{\Delta t} \sin \frac{\Delta t}{2} \quad \text{since } \Delta t = \frac{2\pi}{N + 1}. \tag{4.30}$$

Therefore, using (4.29)

$$\frac{\sin^2 \dfrac{\Delta t}{2}}{\Delta t^2} \operatorname*{Max}_s |a_s|^2 \leqq \sum_{s=1}^{N} |a_s|^2 |\mu_s|^2 \leqq K\Delta t^4,$$

i.e.,

$$\operatorname*{Max}_s |a_s|^2 \leqq K_1 \Delta t^4.$$

Consequently,

$$\|\mathbf{b} - \mathbf{c}\|_{2,N}^2 = \sum_{s=1}^{N} |a_s|^2 \leqq K_1 \Delta t^3 \sum_{s=1}^{N} \Delta t \leqq 2\pi K_1 \Delta t^3 \tag{4.31}$$

and hence

$$\Delta t \operatorname*{Max}_n |b^n - c^n|^2 \leqq \Delta t \sum_{n=1}^{N} |b^n - c^n|^2 \leqq 2\pi K_1 \Delta t^3. \tag{4.32}$$

Thus

$$\operatorname*{Max}_n |b^n - c^n| \leqq K_2 \Delta t. \tag{4.33}$$

Now,

$$\|V^n - U^n\|_\infty \leqq \|W^1\|_\infty \{|b^n - c^n| + |b^n - \sin n\Delta t|\}, \tag{4.34}$$

from which we obtain

$$\|V - U\|_\infty \to 0 \quad \text{as } \Delta t \to 0. \tag{4.35}$$

Thus $V(\Delta t)$ converges uniformly to U if N is odd provided $T = 2\pi$.

If N is even, S has no zero eigenvalues and

$$\min_s |\mu_s| = O(1) \quad \text{as } \Delta t \to 0. \tag{4.36}$$

Hence as before

$$\max_{s=1,\cdots,N} |a_s|^2 \leqq K_1 \Delta t^4 \tag{4.37}$$

which implies uniform convergence whether $T = \pi$ or 2π. This completes the proof of the theorem.

We will see later, however, that whether $T = \pi$ or 2π and whether N is even or odd, it is not possible to solve the system of difference equations (4.3) by either the point Jacobi or the point successive over-relaxation method. We conclude this section with an observation on the Moore-Penrose pseudo-inverse (or general reciprocal), of a matrix [see [12]], in relation to the semidiscrete approximation for the analytic problem (4.1).

If we discretize only the time variable in (4.1), as was done in Section 2, we obtain the system

$$\begin{aligned} &\frac{v^{n+1}(x) - v^{n-1}(x)}{2\Delta t} = \frac{\partial^2 v^n}{\partial x^2} + \pi^2 v^n + \sin \pi x \cos n\Delta t, \qquad n = 1, \ldots, N, \\ &\text{with} \qquad v^n(0) = v^n(1) = 0, \\ &\text{and} \qquad v^0(x) = v^{N+1}(x) = 0. \end{aligned} \tag{4.38}$$

Clearly, any solution of the above system must have the form

$$v^n(x) = c^n \sin \pi x, \qquad n = 1, \ldots, N, \tag{4.39}$$

where c^n's satisfy the equation

$$S\mathbf{c} = \mathbf{p} \tag{4.40}$$

in the previously defined notation.

Now let N be odd so that S is singular. Since $\mathbf{p}$ is orthogonal to the null space of S, there always exists a solution to the last equation and, in fact, all solutions of $S\mathbf{c} = \mathbf{p}$ have the form

$$\mathbf{c} = \mathbf{b} + \beta \psi_{((N+1)/2)}, \tag{4.41}$$

where β is an arbitrary constant and where $\mathbf{b}$ is the vector $\mathbf{b} = \{b^n\}_{n=1}^N$ with

$$b^n = \frac{\Delta t}{\sin \Delta t} \sin n\Delta t, \qquad n = 1, \ldots, N. \tag{4.42}$$

The "pseudo-inverse" of S defines a unique solution of $S\mathbf{c} = \mathbf{p}$ by the requirement that $\mathbf{c}$ be orthogonal to the null space of S.

Suppose now that $T = \pi$. Then, as previously noted, $[\mathbf{b}, \psi_{(N+1)/2}]$ is positive so that the solution obtained via the pseudo-inverse must be such that

$$\mathbf{c} = \mathbf{b} + \beta \psi_{(N+1)/2} \quad \text{with } |\beta| \geqq \beta_0 > 0 \tag{4.43}$$

and with this $\mathbf{c}$, $v^n(x) = c^n \sin \pi x$ does not converge to $\sin \pi x \sin n\Delta t$. On the other hand, if $T = 2\pi$, then $[\mathbf{b}, \psi_{((N+1)/2)}] = 0$ and the pseudo-inverse gives the correct solution

$$\mathbf{c} = \mathbf{b}.$$

4.2. Solutions of the Difference Equations by Iterative Methods. In the iterative solution of linear equations, one distinguishes between point iterative and block iterative methods. The systems of linear equations which arise in the numerical solution of elliptic boundary-value problems are usually such that block iterative methods are more efficient than point iterative methods, i.e., they have a larger asymptotic rate of convergence [see Varga [17]]. Such is not the case for the system (4.5) above.

In the SLOR method, P and N are defined as follows:

$$(4.44) \qquad P = \frac{1}{\omega}[D + \omega E], \qquad N = \frac{1}{\omega}[(1 - \omega)D - \omega F],$$

where ω is a nonzero real parameter and D, E, F are the following block matrices:

$$D = \begin{bmatrix} \Lambda & & O \\ & \ddots & \\ O & & \Lambda \end{bmatrix}, \quad F = \begin{bmatrix} 0 & \sigma I & & O \\ & \ddots & \ddots & \\ & & \ddots & \sigma I \\ O & & & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 0 & & & O \\ -\sigma I & \ddots & & \\ & \ddots & \ddots & 0 \\ O & & -\sigma I & \end{bmatrix}$$

so that $Q = D + E + F$.

The choice $\omega = 1$ in the SLOR method is known as the *line Gauss-Seidel method.* The *line Jacobi method* corresponds to the splitting $Q = P' - N'$, where $P' = D$ and $N' = -(E + F)$ and $(P')^{-1}N'$ is called the line Jacobi matrix.

The following results are known for matrices such as Q which are so-called consistently ordered 2-cyclic matrices (see Varga [17] and D. Young [18]).

(a) If the SLOR method converges, then $0 < \omega < 2$.

(b) Let ρ be an eigenvalue of $P^{-1}N$, the SLOR matrix; if χ satisfies

$$(4.45) \qquad (\rho + \omega - 1)^2 = \chi^2\omega^2\rho, \qquad \omega \neq 0,$$

then χ is an eigenvalue of the line Jacobi matrix. Conversely, if χ is an eigenvalue of the line Jacobi matrix and if ρ satisfies (4.45), then ρ is an eigenvalue of the SLOR matrix. Hence, if the line Jacobi method converges, so does the line Gauss-Seidel and vice versa.

(c) Starting from (4.45) and using conformal mapping arguments, D. Young [18] has proved the following:

THEOREM. *There exists an ω such that the* SLOR *method converges if and only if all the eigenvalues χ of the line Jacobi matrix satisfy* $|\mathrm{Re}(\chi)| < 1$.

If $\beta > 0$ and if no eigenvalue of the line Jacobi matrix is contained in the closed exterior of the ellipse

$$[\mathrm{Re}(\chi)]^2 + [\mathrm{Im}(\chi)]^2/\beta^2 = 1,$$

and if $0 < \omega \leqq 2/(1 + \beta)$, *the* SLOR *method converges.*

Let us apply these results to our situation.

Since $Q = D + E + F$ has the eigenvalues $\mu_s + \lambda_j$, it follows that

$$\chi_{s,j} = \mu_s/\lambda_j, \qquad s = 1, \ldots, N, \quad j = 1, \ldots, M$$

are the eigenvalues of the line Jacobi matrix $D^{-1}(E + F)$. Hence if χ is the spectral radius of $D^{-1}(E + F)$, we have

$$|\chi| = \cos\left(\frac{\pi}{N + 1}\right)\Big/(|\lambda_1|\Delta t) \gtrsim O(1/\Delta t) \quad \text{as } \Delta t \to 0, \tag{4.46}$$

so that for all Δt sufficiently small the line Jacobi and Gauss-Seidel methods diverge for the matrix Q. On the other hand, since $D^{-1}(E + F)$ has only pure imaginary eigenvalues, Young's theorem shows that if

$$\beta = \frac{(1 + \varepsilon)\cos\left(\dfrac{\pi}{N + 1}\right)}{|\lambda_1|\,\Delta t} \quad \text{for any } \varepsilon > 0,$$

then the SLOR method converges for all $0 < \omega \leqq 2/(1 + \beta)$, i.e., for

$$0 < \omega \leqq \frac{2|\lambda_1(\Delta t)|\,\Delta t}{|\lambda_1(\Delta t)|\,\Delta t + (1 + \varepsilon)\cos\dfrac{\pi}{N + 1}}.$$

Point Iterative Methods for the "Model Problem" $\Lambda = H$. We consider now point iterative methods for the case $\Lambda = H$ corresponding to the heat equation. We will assume that Δt, Δx approach zero in such a way that $\Delta t = \gamma\Delta x$ where γ is a positive constant.

We will show that there always exists an interval $0 < \omega < \omega_3$ such that the point successive over relaxation method converges, but *that the point Jacobi* (and hence the point Gauss-Seidel) method *converges if and only if* $\gamma \geqq \gamma_c$, where γ_c is a constant which depends on the range of the space variable x in the analytic problem.

In the point Jacobi method, Q is again split so that $Q = P' - N'$ where now P' is the matrix obtained from Q by deleting all but the main diagonal elements of Q. If L and U are respectively the lower and upper triangular parts of N', the point successive over-relaxation method corresponds to the splitting $Q = P - N$ with

$$P = \frac{1}{\omega}[P' + \omega L], \qquad N = \frac{1}{\omega}[(1 - \omega)P' - \omega U]. \tag{4.47}$$

Moreover, the convergence results (a), (b), (c) stated for line iterative methods remain valid if we replace line by point.

Consider first the eigenvalues of $(P')^{-1}N'$, given by

$$x_{s,j} = (h_j + \mu_s - d)/d, \qquad s = 1, \ldots, N, \quad j = 1, \ldots, M, \tag{4.48}$$

where $d = 2/\Delta x^2$ are the constant diagonal elements of H.

If χ is the spectral radius of $(P')^{-1}N'$ then

$$\chi^2 = \operatorname*{Max}_{s,j} \frac{(2 - h_j\,\Delta x^2)^2 + \Delta x^4|\mu_s|^2}{4} \tag{4.49}$$

and the maximum is attained for $s = j = 1$. Hence, if $\Delta t = \gamma\Delta x$,

$$\chi^2 = \left(2 - 4\sin^2\frac{\pi\Delta x}{2}\right)^2 + \frac{\Delta t^2}{\gamma^4}\cos^2\left(\frac{\pi}{N + 1}\right). \tag{4.50}$$

By Taylor's theorem, we have

$$(4.51)\qquad (2 - 4\sin^2\pi\Delta x/2) = 2\cos\pi\Delta x = 2 - \pi^2 + \pi^4\Delta x^4/12 + O(\Delta x^6).$$

Hence

$$(4.52)\qquad \begin{aligned}(2 - 4\sin^2\pi\Delta x/2)^2 &= 4 - 4\pi^2\Delta x^2 + \tfrac{4}{3}\pi^4\Delta x^4 + O(\Delta x^6)\\ &= 4 - 4\pi^2\gamma^2\Delta t^2/\gamma^4 + \tfrac{4}{3}\pi^4\Delta t^4/\gamma^4 + O(\Delta t^6)\end{aligned}$$

on using $\Delta t = \gamma\Delta x$. Therefore

$$(4.53)\qquad \chi^2 = 1 - \Delta t^2\frac{[4\pi^2\gamma^2 - \cos^2(\pi/(N+1))]}{4\gamma^4} + \frac{4}{3}\frac{\pi^4\Delta t^4}{\gamma^4} + O(\Delta t^6).$$

This shows that the point Jacobi method converges for all sufficiently small Δt if and only if

$$(4.54)\qquad \gamma = \Delta t/\Delta x \geqq 1/2\pi,$$

and the same is true of the point Gauss-Seidel method.

The eigenvalues $\chi_{s,j}$ of $(P')^{-1}N'$ satisfy

$$(4.55)\qquad [\mathrm{Im}(\chi_{s,j})]^2 \leqq \Delta t^2/(4\gamma^4)$$

$$(4.56)\qquad [\mathrm{Re}(\chi_{s,j})]^2 \leqq 1 - \pi^2\Delta t^2/\gamma^2 + O(\Delta t^4).$$

Hence

$$(4.57)\qquad \frac{1}{1 - [\mathrm{Re}(\chi_{s,j})]^2} \leqq \frac{\gamma^2}{\pi^2\Delta t^2}\left[\frac{1}{1 + O(\Delta t^2)}\right],$$

and therefore

$$(4.58)\qquad \frac{[\mathrm{Im}(\chi_{s,j})]^2}{1 - [\mathrm{Re}(\chi_{s,j})]^2} \leqq \frac{1}{4\pi^2\gamma^2}[1 + O(\Delta t^2)].$$

Consequently, given any $\varepsilon > 0$, $\exists\ \delta(\varepsilon)$ such that if $0 < \Delta t < \delta$

$$(4.59)\qquad \frac{[\mathrm{Im}(\chi_{s,j})]^2}{1 - [\mathrm{Re}(\chi_{s,j})]^2} < \frac{1 + \varepsilon}{4\pi^2\gamma^2}.$$

Hence if $\beta^2 = (1 + \varepsilon)/4\pi^2\gamma^2$, we have

$$(4.60)\qquad [\mathrm{Re}(\chi_{s,j})]^2 + [\mathrm{Im}(\chi_{s,j})]^2/\beta^2 < 1.$$

We see then that even if (4.54) is not satisfied, Young's theorem shows that the point successive over-relaxation method converges for all ω such that

$$(4.61)\qquad 0 < \omega \leqq 2\Bigg/\left(1 + \left(\frac{1 + \varepsilon}{4\pi^2\gamma^2}\right)^{1/2}\right).$$

Point Iterative Methods for the System (4.3). Suppose now that $\Lambda = H - \pi^2 I$. In this case the eigenvalues of $(P')^{-1}N'$ are given by

$$(4.62)\qquad \chi_{s,j} = \frac{h_j - \pi^2 + \mu_s - d}{d}, \qquad s = 1, \ldots, N, \quad j = 1, \ldots, M,$$

where now $d = 2/\Delta x^2 - \pi^2$. Hence

$$\max_{s,j} |\mathrm{Re}(\chi_{s,j})| = \frac{\left(2 - 4\sin^2 \dfrac{\pi\Delta x}{2}\right)}{2 - \pi^2\Delta x^2} = \frac{2\cos \pi\Delta x}{2 - \pi^2\Delta x^2}\,. \tag{4.63}$$

Since

$$\cos \pi\Delta x = 1 - \pi^2\Delta x^2/2 + \pi^4\Delta x^4/24 + O(\Delta x^6), \tag{4.64}$$

we have

$$\begin{aligned}\frac{\cos \pi\Delta x}{1 - \pi^2\Delta x^2/2} &= \left[1 - \frac{\pi^2\Delta x^2}{2} + \frac{\pi^4\Delta x^4}{24} + O(\Delta x^6)\right] \\ &\qquad \times \left[1 + \frac{\pi^2\Delta x^2}{4} + \frac{\pi^4\Delta x^4}{4} + O(\Delta x^6)\right] \\ &= 1 + \pi^4\Delta x^4/24 + O(\Delta x^6) > 1\end{aligned} \tag{4.65}$$

if Δx is sufficiently small.

Consequently the point successive over-relaxation method diverges for every ω by Young's theorem. In particular, the Gauss-Seidel method (and therefore the point Jacobi method) diverges.

5. Remarks.

(a) *Merits of the Boundary-Value Method.* It is impossible to comment fully on the merits of any method. On the one hand the advantages or disadvantages are to some extent determined by the existing computational hardware. Thus, suppose that in our present situation line iterative methods were advantageous, as in the case of elliptic problems, and that one had access to a large multi-processing parallel computer. Then the method analyzed here would be extremely worthwhile. At the present time, and with the existing approaches to the matrix inversion problem, one must be less enthusiastic.

On the other hand, advantages or disadvantages of a method also depend on the computational requirements of the "customer." Suppose that one wishes to perform such a "long time" calculation and be certain of the error. In that case, the usual marching procedures, e.g., the Crank-Nicolson, suffer from the possible growth of round-off error. In the method described here for the problem (3.1), one has an estimate (uniform in t and $\mu = (\Delta t^2 + \Delta x^2)^{1/2}$) of the form

$$\|M^{-1}\| \leqq \text{constant}, \tag{5.1}$$

where M is the matrix of (3.4), (see Lemma 4.2). Hence, one may obtain an a posteriori error estimate by simply computing residuals. As a matter of fact, the error in many marching procedures for (3.1) grows linearly with the time even in the absence of round-off error. Such is not the case here.

(b) *Further Comments.* Aside from the potential usefulness of the boundary-value procedure, the results obtained here are of independent interest. Thus for the abstract problem of Section 2.1, we have shown that if the operator A satisfies (2.3), L^2 consistency of the approximating semidiscrete problem is sufficient to guarantee *uniform* convergence. Presumably one may consider more general abstract problems; however, the example of Section 4.1 indicates the importance of requiring the operator A not to have zero as an eigenvalue. This example shows in fact that it is not sufficient to ask that the analytic problem have a unique solution, nor even that the fully-discrete approximate problem have a unique solution which remains bounded uniformly in Δt!! For problems such as (4.1), and more generally for problems where A has nonpositive eigenvalues, one may put $v = e^{-kt}u$ and consider instead the problem

$$\partial v/\partial t = -(A + kI)v + e^{-kt}f, \tag{5.2}$$

where $k > 0$ is chosen so that the spectrum of $A + kI$ lies in the open right half-plane.

Department of Computer Sciences
University of Wisconsin
Madison, Wisconsin 53706

1. F. V. Atkinson, *Discrete and Continuous Boundary Problems*, Mathematics in Science and Engineering, vol. 8, Academic Press, New York, 1964. MR **31** #416.

2. R. E. Bellman, *Introduction to Matrix Analysis*, McGraw-Hill, New York, 1960. MR **23** #A153.

3. H. Bückner, "Über Konvergenzsätze, die sich bei der Anwendung eines Differenzenverfahrens auf ein Sturm-Liouvillesches Eigenwertproblem ergeben," *Math. Z.*, v. 51, 1948, pp. 423–465. MR **11**, 58.

4. A. Carasso, *An Analysis of Numerical Methods for Parabolic Problems Over Long Times*, Ph.D. Thesis, University of Wisconsin, Madison, Wis., 1968.

5. A. Carasso, "Finite-difference methods and the eigenvalue problem for nonselfadjoint Sturm-Liouville operators," *Math. Comp.*, v. 23, 1969, pp. 717–729.

6. E. A. Coddington & N. Levinson, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955. MR **16**, 1022.

7. R. Courant & D. Hilbert, *Methods of Mathematical Physics*. Vol. I, Interscience, New York, 1953. MR **16**, 426.

8. A. Friedman, *Partial Differential Equations of Parabolic Type*, Prentice-Hall, Englewood Cliffs, N. J., 1964, MR **31** #6062.

9. P. R. Garabedian, *Partial Differential Equations*, Wiley, New York, 1964. MR **28** #5247.

10. D. Greenspan, *Approximate Solution of Initial-Boundary Parabolic Problems by Boundary Value Techniques*, MRC Technical Summary Report No. 782, August 1967, U. S. Army Mathematics Research Center, University of Wisconsin, Madison, Wis.

11. D. Greenspan, *Lectures on the Numerical Solution of Linear, Singular, and Nonlinear Differential Equations*, Prentice-Hall, Englewood Cliffs, N. J., 1968. MR **38** #2958.

12. A. S. Householder, *The Theory of Matrices in Numerical Analysis*, Blaisdell, Waltham, Mass., 1964. MR **30** #5475.

13. W. E. Milne, *Numerical Calculus, Approximations, Interpolation, Finite Differences, Numerical Integration, and Curve Fitting*, Princeton Univ. Press, Princeton, N. J., 1949. MR **10**, 483

14. M. H. Protter & H. F. Weinberger, *Maximum Principles in Differential Equations*, Prentice-Hall, Englewood Cliffs, N. J., 1967. MR **36** #2935.

15. R. D. Richtmyer, *Difference Methods for Initial-Value Problems*, 2nd ed., Interscience Tracts in Pure and Appl. Math., no. 4, Interscience, New York, 1957; 2nd ed., with K. W. Morton, 1967. MR **20** #438; MR **36** #3515.

16. R. V. Southwell, *Relaxation Methods in Theoretical Physics*. Vol. II: *A Continuation of the Treatise Relaxation Methods in Engineering Science*, Clarendon Press, Oxford, 1956. MR **18**, 677.

17. R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1962. MR **28** #1725.

18. D. M. Young, "Iterative methods for solving partial difference equations of elliptic type," *Trans. Amer. Math. Soc.*, v. 76, 1954, pp. 92–111. MR **15**, 562.