being given in two parts. The first part includes most of the basic material on time dependent equations including parabolic and hyperbolic problems, multi-dimensional problems, systems and dissipation and dispersion. The second part includes chapters on stability theory for initial–boundary value problems (the GKSO theory), numerical schemes for conservation laws, numerical solution of elliptic problems and an introduction to irregular regions and irregular grids. When I teach the course, I usually cover most of the first five chapters during the first semester. During the second semester I usually cover Chapters 6 and 7 (systems and dissipation and dispersion), Chapter 10 (elliptic equations) and selected topics from Chapters 8, 9 and 11. In other instances, I have covered Chapters 8 and 9 during the second semester, and on one occasion, I used a full semester to teach Chapter 9. Other people who have used the notes have covered parts of Chapters 1–7 and Chapter 10 in one semester. In either case, there seems to be sufficient material for at least two semesters of course work.

At the end of most of the chapters of the text and in the middle of several, we include sections which we refer to as "Computational Interludes." The original idea of these sections was to stop working on new methods, take a break from theory and compute for a while. These sections do include this aspect of the material, but as they developed, they also began to include more than just computational material. It is in these sections that we discuss results from previous homework problems. It is also in these sections that we suggest it is time for the students to try one of their new methods on one of the problems HW0.0.1–HW0.0.4 from the Prelude. There are also some topics included in these sections that did not find a home elsewhere. At times a more appropriate title for these sections might have been "etc.".

At this time I would like to acknowledge some people who have helped me with various aspects of this text. I thank Drs. Michael Kirby, Steve McKay, K. McArthur and K. Bowers for teaching parts of the text and providing me with feedback. I also thank Drs. Kirby, McArthur, Jay Bourland, Paul DuChateau and David Zachmann for many discussions about various aspects of the text. Finally, I thank the many students who over the years put up with the dreadfully slow convergence of this material from notes to text. Whatever the result, without their input the result would not be as good. And, finally, though all of the people mentioned above and others have tried to help me, there are surely still some typos and errors of thought (though, hopefully, many mistakes have been corrected for the Second Printing). Though I do so sadly, I take the blame for all of these mistakes. I would appreciate it if you would send any mistakes that you find to thomas@math.colostate.edu. Thank you.

J.W. Thomas

Contents

Contents of Part 1: Finite Difference Methods

8

Stability of Initial–Boundary–Value Schemes

8.1 Introduction

Since early in Chapter 1, we have been computing solutions to initial-boundary-value problems. In Chapter 2 we included some theory that could be used to prove convergence of schemes for solving initial-boundary-value problems. In Example 2.2.2 we used the definition of convergence to prove the convergence of the basic difference scheme for the heat equation with zero Dirichlet boundary conditions. For the same difference scheme, in Section 2.5.2 we noted that the consistency and stability analyses done earlier in the text along with the Lax Theorem for a bounded domain (Theorem 2.5.3) imply convergence. We also pointed out that we could directly apply the definitions of consistency and stability, and Theorem 2.5.3 to obtain convergence for a hyperbolic scheme.

As we started developing tools for proving convergence (via the Lax Theorem), we found that the methods for initial-boundary-value schemes based on Chapter 2 worked nicely when we had Dirichlet or Neumann boundary conditions and symmetric difference operators (Example 3.1.7, Example 3.1.9, Example 4.3.4, Section 4.4.3.3, etc.) but that they cannot be used when either the difference operator is not symmetric (Example 3.1.6, Example 3.1.8, and all of the schemes given in Chapter 5 and 6 for hyperbolic equations) or when the boundary condition makes the finite difference operator nonsymmetric (the example done in Section 3.2 with the mixed boundary condition). Thus, *at the moment we do not have sufficiently good methods for proving stability for initial-boundary-value schemes.*

Remark 2: Also, we do not apply the approximation (10.5.114) for just one value of j. We let $\lambda_{1,\omega}$ denote that average value of the approximations given by approximation (10.5.114) over all j (or over a sampling of the j's). If the approximate eigenvalue $\lambda_{1,\omega}$ does not appear to be converging as a function of k, the ω chosen is probably greater than ω_b.

Remark 3: We must realize that this approach depends on an asymptotic result, i.e., k must generally be large before (10.5.114) gives a good approximation of λ_1. As we see below in HW10.5.22, if the calculation is made too early, we get a bad approximation (maybe better than we presently have, but not good enough to warrant all of the work). In parts (a)(iii) and (a)(iv) of HW10.5.22 we see that the approach can give us a very good approximation of ω_b.

Remark 4: We notice in part (b) of HW10.5.22 that if the solution and initial guess (i.e., the initial error) are sufficiently trivial and contain only one of the eigenvectors, the technique will compute that eigenvalue associated with that eigenvector, and will do it well with only three iterations. However, we should understand that we have not found the largest eigenvalue of the Jacobi iterations matrix and cannot find ω_b. In general, if the eigenvector associated with the largest eigenvalue is not present in the eigenvector expansion of the initial error, the above procedure will not find an approximation to the largest eigenvalue (and hence will not compute ω_b). The procedure will find an approximation to the largest eigenvalue associated with one of the eigenvectors present in the eigenvector expansion of the initial error.

Of course, there are other approaches to finding approximations of ω_b. We have included one approach mainly to give a taste of how it might be done and to emphasize that it must be done. For a more complete discussion, which includes some computer programs, see [22], page 223.

HW 10.5.22 Consider the problem

$$\nabla^2 v = F(x,y), \quad (x,y) \in R = (0,1) \times (0,1)$$
$$v = 0, \quad (x,y) \text{ on } \partial R.$$

(a) For $F(x,y) = e^{x+y}$, determine an approximation of ω_b by
  (i) using $\left(\frac{w_{j_3} - w_{j_2}}{w_{j_2} - w_{j_1}}\right)$ for one point j.
  (ii) using the average of $\left(\frac{w_{j_3} - w_{j_2}}{w_{j_2} - w_{j_1}}\right)$ over 100 points.
  (iii) using $\left(\frac{w_{j_{10}} - w_{j_9}}{w_{j_9} - w_{j_8}}\right)$ for one point j.
  (iv) using the average $\left(\frac{w_{j_{10}} - w_{j_9}}{w_{j_9} - w_{j_8}}\right)$ over 100 points.

(b) Repeat part (a) using $F(x,y) = \sin \pi x \sin 2\pi y$.

(c) Explain why the computation done in part (b) gives a bad approximation of ω_b.

10.6 Elliptic Difference Equations: Neumann Boundary Conditions

To this point in our consideration of elliptic boundary value problems, we have treated only Dirichlet boundary conditions. There are other boundary conditions we could consider other than Dirichlet and Neumann boundary conditions but these are clearly the two most common types of boundary conditions. We consider mixture boundary conditions and Robin boundary conditions in Section 10.8. In this section we will consider the numerical solution of elliptic partial differential equations with Neumann boundary conditions. For further results, analytic and numerical, see [21].

We consider the model problem

$$-\nabla^2 v = F \quad \text{in } R = (0,1) \times (0,1) \tag{10.6.1}$$
$$\frac{\partial v}{\partial n} = g \quad \text{on } \partial R. \tag{10.6.2}$$

It is easy to see that if v is a solution to problem (10.6.1)–(10.6.2), so is $v + c$ for any constant c. Hence, we know that we have a nonunique solution to problem (10.6.1)–(10.6.2). One might be tempted to say that because the problem does not have a unique solution, it cannot be an important problem. This is not the case. Problem (10.6.1)–(10.6.2) is easily as important as the analogous problem with Dirichlet boundary conditions. We must must be able to solve problems of this form and must be very careful to handle the numerical consequences of the nonuniqueness.

To talk about solutions to problem (10.6.1)–(10.6.2), one must prescribe conditions on F and g that will allow solutions to exist. We prove the following proposition.

Proposition 10.6.1 *If R is a Green's region and problem (10.6.1)–(10.6.2) has a solution v, then*

$$-\int_R F(x,y)\,dx\,dy = \int_{\partial R} g(x,y)\,ds. \tag{10.6.3}$$

For any constant c, $v+c$ will also be a solution to problem (10.6.1)–(10.6.2).

Proof: Before we start the proof, we define a **Green's region** to be a region in the plane sufficiently nice to satisfy the hypotheses for the first Green's formula,

$$\int_R v_1 \nabla^2 v_2\,dx\,dy = -\int_R (\nabla v_1, \nabla v_2)\,dx\,dy + \int_{\partial R} v_1 \frac{\partial v_2}{\partial n}\,ds,$$

where $(\nabla v_1, \nabla v_2)$ denotes the dot product of ∇v_1 and ∇v_2. If we let $v_1 = 1$ and $v_2 = v$, we obtain (10.6.3).

Remark 1: The converse of Proposition 10.6.1 is also true. See [21], page 154.

Remark 2: We shall refer to condition (10.6.3) as the **analytic compatibility condition.** It is also a conservation condition. The term on the left represents the amount of the material injected into or pumped out of the region, and the term on the right represents the amount of the material that flows into or out of the boundary of the region.

We wish to approximate the solution to problem (10.6.1)–(10.6.2) numerically. We will proceed much in the same way that we did for the analogous Dirichlet problem. We want an analogue to Theorem 10.3.3, and then we want to study the solution to the discrete problem. For convenience, we consider a grid on the region $R = [0,1] \times [0,1]$, $(x_j, y_k) = (j\Delta x, k\Delta y)$, $j = 0, \ldots, M$, $k = 0, \ldots, M$, where $\Delta y = \Delta x$. We difference the partial differential equation as we have before, and we approximate equation (10.6.1) by

$$-\frac{1}{\Delta x^2}\left(\delta_x^2 + \delta_y^2\right) u_{j\,k} = F_{j\,k} \quad j,k = 1, \ldots, M-1. \tag{10.6.4}$$

It is not as clear how we should approximate the boundary conditions (10.6.2). As was the case when we considered Neumann boundary conditions for parabolic equations, Sections 1.4 and 4.4.3.2, at least two of the obvious choices are to use the first or the second order approximations of the normal derivative. We treat each of these cases in the following two sections.

10.6.1 First Order Approximation

The treatment of the boundary conditions for Dirichlet boundary conditions was very routine and easy. It should not surprise us that the approximation of the boundary conditions might be more difficult and important when we consider Neumann boundary conditions. Using a first order approximation of the derivative in boundary condition (10.6.2) leaves us with the following discrete boundary conditions.

$$-\frac{u_{1\,k} - u_{0\,k}}{\Delta x} = g_{0\,k}, \quad k = 0, \ldots, M \tag{10.6.5}$$

$$\frac{u_{M\,k} - u_{M-1\,k}}{\Delta x} = g_{M\,k}, \quad k = 0, \ldots, M \tag{10.6.6}$$

$$-\frac{u_{j\,1} - u_{j\,0}}{\Delta x} = g_{j\,0}, \quad j = 0, \ldots, M \tag{10.6.7}$$

$$\frac{u_{j\,M} - u_{j\,M-1}}{\Delta x} = g_{j\,M}, \quad j = 0, \ldots, M \tag{10.6.8}$$

We should note that the negative signs in formulas (10.6.5) and (10.6.7) are due to the fact that the normal vector associated with the Neumann boundary condition is assumed to be an outward normal. Thus we must consider solving equations (10.6.4)–(10.6.8). This system of equations can

be written as

$$A\mathbf{u} = \mathbf{f} \tag{10.6.9}$$

where A is the $(M-1) \times (M-1)$ block matrix

$$A = \frac{1}{\Delta x^2}\begin{pmatrix} T_1 & -I & \Theta & \cdots & \\ -I & T & -I & \Theta & \cdots \\ & \ddots & \ddots & \ddots & \\ \cdots & \Theta & -I & T & -I \\ & \cdots & \Theta & -I & T_1 \end{pmatrix}, \tag{10.6.10}$$

T_1 and T are the $(M-1) \times (M-1)$ matrices

$$T_1 = \begin{pmatrix} 2 & -1 & 0 & \cdots & \\ -1 & 3 & -1 & 0 & \cdots \\ & \ddots & \ddots & \ddots & \ddots \\ & \cdots & 0 & -1 & 3 & -1 \\ & & \cdots & 0 & -1 & 2 \end{pmatrix}, \tag{10.6.11}$$

$$T = \begin{pmatrix} 3 & -1 & 0 & \cdots & & \\ -1 & 4 & -1 & 0 & \cdots & \\ & \ddots & \ddots & \ddots & \ddots & \cdots \\ & \cdots & 0 & -1 & 4 & -1 \\ & & \cdots & 0 & -1 & 3 \end{pmatrix}, \tag{10.6.12}$$

I is the $(M-1) \times (M-1)$ identity matrix, Θ is the $(M-1) \times (M-1)$ zero matrix, and $\mathbf{f}$ is given by

$$\mathbf{f} = \mathbf{F} + \mathbf{b}_x + \mathbf{b}_y \tag{10.6.13}$$

where $\mathbf{F}$, $\mathbf{b}_x$, and $\mathbf{b}_y$ are the $L = (M-1)^2$–vectors

$$\mathbf{F} = [F_{1\,1} \cdots F_{M-1\,1}\, F_{1\,2} \cdots F_{M-1\,M-1}]^T, \tag{10.6.14}$$

$$\mathbf{b}_x = \frac{1}{\Delta x}[g_{0\,1}\, 0 \cdots 0\, g_{M\,1}\, g_{0\,2}\, 0 \cdots g_{M\,M-1}]^T \tag{10.6.15}$$

and

$$\mathbf{b}_y = \frac{1}{\Delta x}[g_{1\,0} \cdots g_{M-1\,0}\, 0 \cdots 0\, g_{1\,M} \cdots g_{M-1\,M}]^T. \tag{10.6.16}$$

See Figure 10.6.1 for the full matrix written out for the case of $M = 5$.

We include most of the important properties of matrix A and equation (10.6.9) in the following proposition.

$$\frac{1}{\Delta x^2}\left(\begin{array}{cccccccccccccccc|cccccccccccccccc}
2 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 3 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 3 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 2 & \mathbf{0} & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
-1 & 0 & 0 & \mathbf{0} & 3 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & -1 & 3 & \mathbf{0} & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & -1 & 0 & 0 & \mathbf{0} & 3 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 3 & \mathbf{0} & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & \mathbf{0} & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right)$$

FIGURE 10.6.1. Matrix A, as in equation (10.6.9), represents the matrix associated with solving difference equation (10.6.4) along with the first order approximation of the Neumann boundary conditions given by (10.6.5)–(10.6.8) with $M = 5$. A boldface zero represents a zero that is due to a boundary condition term.

Proposition 10.6.2 (1) $A\mathbf{1} = \boldsymbol{\theta}$.
(2) $\mathbf{1}$ *is the only vector in the null space of* A.
(3) *If* $\mathbf{u}^1$ *and* $\mathbf{u}^2$ *are any two solutions to equation (10.6.9), then there exists a constant* c *such that* $\mathbf{u}^1 = \mathbf{u}^2 + c\mathbf{1}$, *where* $\mathbf{1} = [1 \ \cdots \ 1]^T$.
(4) *Equation (10.6.9) has a solution if and only if*

$$-\Delta x^2 \sum_{k=1}^{M-1}\sum_{j=1}^{M-1} F_{jk} = \Delta x \sum_{j=1}^{M-1} [g_{j0} + g_{jM}] + \Delta x \sum_{k=1}^{M-1} [g_{Mk} + g_{0k}]. \tag{10.6.17}$$

Proof: To see that statement (1) is true, one need only compute $A\mathbf{1}$. The **null space** of a matrix A is defined to be $\{\mathbf{X} : A\mathbf{X} = \boldsymbol{\theta}\}$ and is denoted by $N(A)$. Statement (1) can be expressed as $\mathbf{1} \in N(A)$.

(2) If we eliminate the first row and column of the matrix A, we can apply Proposition 10.2.5 to the resulting submatrix to show that this submatrix is of full rank (the submatrix is invertible). As usual, it is easy to see that the submatrix is diagonally dominant. Likewise, it is easy to see that the first row of the submatrix satisfies the strict inequality necessary for Proposition 10.2.5. Technically, the irreducibility of the submatrix follows from the fact that any two grid points can be connected by a chain of neighboring points. It is also not difficult to use the characterization of irreducibility given immediately preceding Proposition 10.2.5 (if a change in any of the components of the right hand side will cause a change in the solution) to see that the submatrix is irreducible.

Since the submatrix formed by eliminating one row and one column is of full rank, $L - 1$ where $L = (M-1)^2$, then the rank of A is at least $L - 1$. Since we know that $\mathbf{1} \in N(A)$, then the rank of A is $L - 1$, and $\mathbf{1}$ is the only vector in $N(A)$.

(3) When we know that the rank of A is $L - 1$ and $\mathbf{1} \in N(A)$, we know that we can solve equation (10.6.9) for $L - 1$ of the variables in terms of one of the variables, i.e., all solutions of equation (10.6.9) are of the form

$$\mathbf{u} = \mathbf{u}_0 + c\mathbf{1}$$

where $\mathbf{u}_0$ is a fixed vector. Statement (3) then follows easily.

(4) Since $\mathbf{1} \in N(A)$, we know that equation (10.6.9) is solvable if and only if $(\mathbf{f}, \mathbf{1}) = 0$, [31], page 17. If we return to equations (10.6.13)–(10.6.16), we see that

$$\begin{aligned} 0 &= (\mathbf{f}, \mathbf{1}) \\ &= (\mathbf{F}, \mathbf{1}) + (\mathbf{b}_x, \mathbf{1}) + (\mathbf{b}_y, \mathbf{1}) \\ &= \sum_{j=1}^{M-1}\sum_{k=1}^{M-1} F_{jk} + \frac{1}{\Delta x}\sum_{k=1}^{M-1} [g_{0k} + g_{Mk}] + \frac{1}{\Delta x}\sum_{j=1}^{M-1} [g_{j0} + g_{jM}]. \end{aligned}$$

It is easy to see that this is the same as equation (10.6.17).

The results given in Proposition 10.6.2 show that equation (10.6.9) is not a nice equation. Considering the analytic analogue, equations (10.6.1)–(10.6.2), this should not surprise us. In fact, it is clearly good that $\mathbf{1} \in N(A)$ and that all solutions are of the form $\mathbf{u} = \mathbf{u}_0 + c\mathbf{1}$. Both of these facts show that the numerical problem (10.6.9) mimics the analytic problem (10.6.1)–(10.6.2) well. In addition, since the analytic problem had an analytic compatibility condition that must be satisfied for solvability, it is also logical that we have a **discrete compatibility condition**, (10.6.17), that must be satisfied for equation (10.6.9) to be solvable.

Remark 1: Just as the analytic compatibility condition could be described as a conservation property, the discrete compatibility condition (10.6.17) can also be described as a conservation law for difference equations (10.6.4)–(10.6.8).

Remark 2: One problem we face is the fact that if we assume that F and g are nice enough to satisfy the analytic compatibility condition (10.6.3), this is not enough to imply that F and g will satisfy the discrete compatibility condition (10.6.17). If the integrals involved in the analytic compatibility condition are approximated by the appropriate numerical integration scheme, we see that

$$\begin{aligned}
0 &= \int_R F(x,y)\,dx\,dy + \int_{\partial R} g(x,y)\,ds \\
&= \int_R F(x,y)\,dx\,dy + \int_0^1 g(x,0)\,dx + \int_0^1 g(1,y)\,dy \\
&\quad + \int_1^0 g(x,1)\,dx + \int_1^0 g(0,y)\,dy \\
&\approx \sum_{k=1}^{M-1}\sum_{j=1}^{M-1} F_{j\,k}\Delta x\Delta x + \sum_{j=1}^{M-1} g_{j\,0}\Delta x + \sum_{k=1}^{M-1} g_{M\,k}\Delta x \\
&\quad + \sum_{j=1}^{M-1} g_{j\,M}\Delta x + \sum_{k=1}^{M-1} g_{0\,k}\Delta x + \mathcal{O}(\Delta x).
\end{aligned}$$

We note that the numerical approximation is not well done. If we consider rectangular regions centered at the grid points, the area integral omits a strip ($\mathcal{O}(\Delta x)$) around the region. The line integrals skip little chunks in each corner.

More importantly, we see that if F and g satisfy the analytic compatibility condition, F and g will generally only approximately satisfy the discrete compatibility condition. Because of this fact, *part (4) of Proposition 10.6.2 is generally not satisfied.* In addition, if we review our solution methods, we do not have any methods that are designed for solving singular equations. To explain what happens when the discrete compatibility condition is only approximately satisfied and obtain a system of equations that we can solve,

we consider a slight variation of system (10.6.9),

$$\bar{A}\bar{\mathbf{u}} = \bar{\mathbf{f}} \tag{10.6.18}$$

where

$$\bar{A} = \begin{pmatrix} A & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}, \quad \bar{\mathbf{u}} = \begin{bmatrix} \mathbf{u} \\ \lambda \end{bmatrix}, \text{ and } \bar{\mathbf{f}} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix}.$$

We are then able to prove the following proposition related to equation (10.6.18).

Proposition 10.6.3 (1) *System (10.6.18) is solvable.*
(2) *If the solution to system (10.6.18) is of the form* $\bar{\mathbf{u}}_0 = [\mathbf{u}_0 \quad 0]^T$, *then the discrete compatibility condition (10.6.17) is satisfied and* $\mathbf{u}_0$ *is a solution to system (10.6.9) such that* $(\mathbf{u}_0, \mathbf{1}) = 0$.
(3) *If the solution to system (10.6.18) is of the form* $\bar{\mathbf{u}}_0 = [\mathbf{u}_0 \quad \lambda]^T$, *where* $\lambda \neq 0$, *then* $\mathbf{u}_0$ *is a solution to the equation*

$$A\mathbf{u} = \mathbf{f} - \lambda\mathbf{1} \tag{10.6.19}$$

such that $(\mathbf{u}_0, \mathbf{1}) = 0$.

Proof: (1) From [31], page 17, we see that the range of the matrix A is orthogonal to $N(A^T)$. Since A is symmetric, $N(A^T) = N(A)$. The fact that the range of A is the span of the columns of A implies that $\mathbf{1}$ is independent of the columns of A. Then, since the rank of A is $L-1$, the rank of $[A \quad \mathbf{1}]$ is L. Since $[\mathbf{1}^T \quad 0]$ is independent of the rows of $[A \quad \mathbf{1}]$ (for essentially the same reason), the rank of $\bar{A}$ is $L+1$, i.e., $\bar{A}$ is of full rank and, hence, solvable for any right hand side.

(2) If the solution of equation (10.6.18) is in the form $\bar{\mathbf{u}} = [\mathbf{u}_0 \quad 0]^T$, then $\mathbf{u}_0$ satisfies $A\mathbf{u} = \mathbf{f}$ and $(\mathbf{u}_0, \mathbf{1}) = 0$ because of the form of equation (10.6.18). The discrete compatibility condition is satisfied because, as in Proposition 10.6.2, $\mathbf{1} \in N(A)$ implies that $(\mathbf{f}, \mathbf{1}) = 0$ (since equation (10.6.9) is solvable).

The proof of part (3) is a consequence of the hypotheses and the form of $\bar{A}$, $\bar{\mathbf{u}}$ and $\bar{\mathbf{f}}$.

Remark 1: In both of the parts (2) and (3) of Proposition 10.6.3 above, we have constrained our solutions to satisfy $(\mathbf{u}_0, \mathbf{1}) = 0$. The solutions to equation (10.6.9) (part (2) of the proposition) and (10.6.19) are of the form $\mathbf{u}_0 + c\mathbf{1}$ for some constant vector $\mathbf{u}_0$. Requiring that the solution be orthogonal to $\mathbf{1}$ forces c to be zero.

Remark 2: Part (3) of Proposition 10.6.3 first appears to be bad in that we cannot solve the problem that we want to solve. However, instead of being as bad as it first appears, the situation is quite nice. Because the discrete compatibility condition is satisfied only approximately, we know that equation (10.6.9) does not generally have a solution. Part (3) takes

care of the problem and gives the best solution possible—given the approximation of the boundary conditions. The solution to equation (10.6.19) is a solution to equation (10.6.9) where the array F_{jk} (the function F) has been replaced by

$$\tilde{F}_{jk} = F_{jk} - \lambda \mathbf{1}. \tag{10.6.20}$$

In other words, system (10.6.18) is smart enough to realize that we have the wrong F_{jk} (part of it could be due to g) and fixes it for us. As we shall see in the next theorem, λ is small.

Now that we understand what we are solving when we solve the discrete problem associated with problem (10.6.1)–(10.6.2), we state the following analogue of Theorem 10.3.3.

Theorem 10.6.4 *Let $v \in C^4(\overline{R})$ be a solution to Neumann problem (10.6.1)–(10.6.2). Let $\bar{\mathbf{u}} = [\mathbf{u}_0 \quad \lambda]^T$ be a solution to equation (10.6.18). Then*

$$|\lambda| = \mathcal{O}(\Delta x)$$
$$\|\mathbf{u}_0 - \mathbf{v}\|_\infty = \mathcal{O}(\Delta x).$$

Proof: See [21], page 69.

Remark 1: As in Theorem 10.3.3, the hypothesis that $v \in C^4$ implies that v has continuous derivatives up through order four. In addition, as in Theorem 10.3.3, the constants involved in the $\mathcal{O}$ notation depend on the derivatives of v. There is dependence on the fourth derivations due to the fact that difference equation (10.6.4) is a second order approximation to the partial differential equation (10.6.1). There is also dependence on the second derivatives of v because of the first order approximation of the boundary conditions.

Remark 2: From the proof of Theorem 10.6.4, it can be seen that the fact that the convergence above is only of first order in Δx and Δy is due to the treatment of the boundary conditions.

Remark 3: We might recall that in Example 2.3.4 when we used the one dimensional version of the first order approximations for the Neumann boundary condition used here, we found that the difference scheme was not norm consistent (the truncation error was not of first order, as one might expect). (We did find in Example 8.6.2 that we could prove that the scheme was convergent order Δx by the Osher result.) Yet here we find that we obtain first order convergence using these approximations. The difference is due to the time dependence involved in the norm consistency definition, Definition 2.3.2. Since we do not have to define consistency here that is compatible with the Lax Theorem, we have no problem with these boundary conditions.

HW 10.6.1 Use conservation methods similar to those used in Sections 1.6 and 1.6.1 to derive difference equations (10.6.4)–(10.6.8).

10.6.2 Second Order Approximation

Now that we have seen that we obtain first order convergence using the first order approximation of the boundary conditions (while the difference equation was a second order approximation to the partial differential equation), we might hope that if we use a second order approximation to the boundary conditions, we will obtain second order convergence. To show that this is the case, we essentially have to repeat everything done in the last section. We will include as much of the material that we feel is necessary and/or helpful, leaving the rest to your imagination or your own reading.

To use a second order approximation for the boundary conditions, as we did for parabolic equations in Chapter 4, we consider the difference equation at both the interior and boundary grid points, i.e., we consider

$$-\frac{1}{\Delta x^2}\left(\delta_x^2 + \delta_y^2\right) u_{jk} = F_{jk}, \quad j,k = 0,\ldots,M. \tag{10.6.21}$$

Of course, the problem with difference equation (10.6.21) is that it reaches to fictitious points outside of the domain. These points are eliminated by the use of the following second order approximation of the Neumann boundary condition.

$$-\frac{u_{1k} - u_{-1k}}{2\Delta x} = g_{0k}, \quad k = 0,\ldots,M \tag{10.6.22}$$
$$\frac{u_{M+1k} - u_{M-1k}}{2\Delta x} = g_{Mk}, \quad k = 0,\ldots,M \tag{10.6.23}$$
$$-\frac{u_{j1} - u_{j-1}}{2\Delta x} = g_{j0}, \quad j = 0,\ldots,M \tag{10.6.24}$$
$$\frac{u_{jM+1} - u_{jM-1}}{2\Delta x} = g_{jM}, \quad j = 0,\ldots,M. \tag{10.6.25}$$

Difference equation (10.6.21) along with boundary conditions (10.6.22)–(10.6.25) can be written as

$$A\mathbf{u} = \mathbf{f}, \tag{10.6.26}$$

where A is the $(M+1)\times(M+1)$ block tridiagonal matrix

$$A = \frac{1}{\Delta x^2}\begin{pmatrix} T & -2I & \Theta & \cdots & \\ -I & T & -I & \Theta & \cdots \\ & \ddots & \ddots & \ddots & \\ \cdots & \Theta & -I & T & -I \\ & \cdots & \Theta & -2I & T \end{pmatrix}, \tag{10.6.27}$$

T is the $(M+1) \times (M+1)$ matrix

$$T = \begin{pmatrix} 4 & -2 & 0 & \cdots & & \\ -1 & 4 & -1 & 0 & \cdots & \\ & \ddots & \ddots & \ddots & \ddots & \\ & \cdots & 0 & -1 & 4 & -1 \\ & & \cdots & 0 & -2 & 4 \end{pmatrix}, \tag{10.6.28}$$

I is the $(M+1) \times (M+1)$ identity matrix, Θ is the $(M+1) \times (M+1)$ zero matrix, and $\mathbf{f}$ is given by

$$\mathbf{f} = \mathbf{F} + \mathbf{b}_x + \mathbf{b}_y, \tag{10.6.29}$$

where $\mathbf{F}$, $\mathbf{b}_x$ and $\mathbf{b}_y$ are the $L = (M+1)^2$-vectors

$$\mathbf{F} = [F_{0\,0} \ \cdots \ F_{M\,0} \ F_{0\,1} \ \cdots \ F_{M\,M}]^T, \tag{10.6.30}$$

$$\mathbf{b}_x = \frac{2}{\Delta x}[g_{0\,0} \ 0 \ \cdots \ 0 \ g_{M\,0} \ g_{0\,1} \ 0 \ \cdots \ g_{M\,M}]^T \tag{10.6.31}$$

and

$$\mathbf{b}_y = \frac{2}{\Delta x}[g_{0\,0} \ \cdots \ g_{M\,0} \ 0 \ \cdots \ 0 \ g_{0\,M} \ \cdots \ g_{M\,M}]^T. \tag{10.6.32}$$

The full matrix A is given in Figure 10.6.2 for $M = 4$. Compare the form of A with that given for the scheme using the first order approximation of the Neumann boundary conditions given in Figure 10.6.1 (noting that for Figure 10.6.1 we used $M = 5$, whereas here we used $M = 4$ so that we could fit it on a page).

Before we get too serious about equation (10.6.26), we notice that as in the previous section, $A\mathbf{1} = \boldsymbol{\theta}$ ($\mathbf{1} \in N(A)$) and $\mathbf{1}$ is the only vector in the null space of A. One of the differences from the case of the first order approximation is that now A is not symmetric. Generally, nonsymmetry tends to cause problems.

We next would like to find the discrete compatibility condition associated with equation (10.6.26). When A is nonsymmetric, we do not get the same result as we did for the case of the first order approximation. (Equation (10.6.9) is solvable if and only if $(\mathbf{f}, \mathbf{1}) = 0$.) In that result, we were using the fact that the matrix A was symmetric. However, the result given in [31], page 17, is that *equation (10.6.26) is solvable if and only if* $(\mathbf{f}, \mathbf{u}^*) = 0$ *for all* $\mathbf{u}^* \in N(A^T)$. Using this result, we can proceed as we did in the case of the first order approximation and obtain the following proposition.

Proposition 10.6.5 (1) $A\mathbf{1} = \boldsymbol{\theta}$ *and* $N(A) = \{\mathbf{1}\}$.
(2) $A^T\mathbf{u}^* = \boldsymbol{\theta}$ *where* $\mathbf{u}^*$ *is the* $L = (M-1)^2$*-vector*

$$\mathbf{u}^* = [\mathbf{u}^1 \ \mathbf{u}^2 \ \cdots \ \mathbf{u}^2 \ \mathbf{u}^1]^T,$$

$$\frac{1}{\Delta x^2}\left(\begin{array}{ccccc|ccccc|ccccc|ccccc|ccccc}
4 & -2 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 4 & -1 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 4 & -1 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 4 & -1 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -2 & 4 & \mathbf{0} & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
-1 & 0 & 0 & 0 & \mathbf{0} & 4 & -2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -2 & 4 & \mathbf{0} & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & \mathbf{0} & 4 & -2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -2 & 4 & \mathbf{0} & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & \mathbf{0} & 4 & -2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -2 & 4 & \mathbf{0} & 0 & 0 & 0 & -1 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & \mathbf{0} & 4 & -2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & -1 & 4 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & -1 & 4 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & -1 & 4 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & -2 & 4
\end{array}\right)$$

FIGURE 10.6.2. Matrix A, as in equation (10.6.26), represents the matrix associated with solving difference equation (10.6.4) along with the second order approximation of the Neumann boundary conditions given by (10.6.22)–(10.6.25) with $M = 4$. A boldface zero represents a zero that is due to a boundary condition term.

made up of the $(M-1)$*-vectors*

$$\mathbf{u}^1 = \left[\frac{1}{4}\ \frac{1}{2}\ \cdots\ \frac{1}{2}\ \frac{1}{4}\right]^T,$$

and

$$\mathbf{u}^2 = \left[\frac{1}{2}\ 1\ \cdots\ 1\ \frac{1}{2}\right]^T,$$

and $N(A^T) = \{\mathbf{u}^*\}$.
(3) *If* $\mathbf{u}^\diamondsuit$ *and* $\mathbf{u}^\heartsuit$ *are any two solutions to equation (10.6.26), then there exists a constant* c *such that* $\mathbf{u}^\diamondsuit = \mathbf{u}^\heartsuit + c\mathbf{1}$.
(4) *Equation (10.6.26) has a solution if and only if*

$$-\Delta x^2 \sum_{k=0}^{M}\sum_{j=0}^{M} s_j s_k F_{jk} = \Delta x \sum_{j=0}^{M} \frac{1}{2} s_j \left[g_{j0} + g_{jM}\right] + \Delta x \sum_{k=0}^{M} \frac{1}{2} s_k \left[g_{Mk} + g_{0k}\right], \tag{10.6.33}$$

where $s_0 = \frac{1}{2}$, $s_M = \frac{1}{2}$ *and* $s_j = 1$ *for* $j = 1, \ldots, M-1$.

Proof: The proofs of (1), (2) and (3) follow the same approach used for parts (1), (2) and (3) of Proposition 10.6.2.

The proof of part (4) follows from the fact mentioned earlier that equation (10.6.26) is solvable if and only if $(\mathbf{f}, \mathbf{u}^*) = 0$ for all $\mathbf{u}^* \in N(A^T)$ and the fact that $N(A^T) = \{\mathbf{u}^*\}$, given in part (2) in this proposition. Equation (10.6.33) is the same as $(\mathbf{f}, \mathbf{u}^*) = 0$, where $\mathbf{u}^*$ is as defined in part (2).

We can then proceed as we did in the last section and consider the variation of equation (10.6.26),

$$\bar{A}\bar{\mathbf{u}} = \bar{\mathbf{f}}, \tag{10.6.34}$$

where

$$\bar{A} = \begin{pmatrix} A & \mathbf{u}^* \\ \mathbf{1}^T & 0 \end{pmatrix}, \quad \bar{\mathbf{u}} = \begin{bmatrix} \mathbf{u} \\ \lambda \end{bmatrix}, \text{ and } \bar{\mathbf{f}} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix}.$$

We are then able to prove the following proposition.

Proposition 10.6.6 (1) *System (10.6.34) is solvable.*
(2) *If the solution to system (10.6.34) is of the form* $\bar{\mathbf{u}}_0 = [\mathbf{u}_0\ \ 0]^T$, *then the discrete compatibility condition (10.6.33) is satisfied, and* $\mathbf{u}_0$ *is a solution to system (10.6.26) such that* $(\mathbf{u}_0, \mathbf{1}) = 0$.
(3) *If the solution to system (10.6.34) is of the form* $\bar{\mathbf{u}}_0 = [\mathbf{u}_0\ \ \lambda]^T$, *where* $\lambda \neq 0$, *then* $\mathbf{u}_0$ *is a solution to the equation*

$$A\mathbf{u} = \mathbf{f} - \lambda\mathbf{u}^* \tag{10.6.35}$$

such that $(\mathbf{u}_0, \mathbf{1}) = 0$.

Proof: The proof follows much as the proof of Proposition 10.6.3. The only difference is that we must use the complete result (we do not have a symmetric matrix) that $N(A) \oplus R(A^T) = \mathbb{R}^L$ and $N(A^T) \oplus R(A) = \mathbb{R}^L$ (i.e., the null space of A is independent of the range of A^T, the span of the rows of A; and the null space of A^T is independent to the range of A, the span of the columns of A).

We note that as in the case of the first order approximation, if the discrete compatibility condition is not satisfied (as it generally will not be), system (10.6.34) adjusts the right hand side the appropriate amount. We also note that the discrete compatibility condition (10.6.33) is a $\mathcal{O}(\Delta x^2)$ approximation of the analytic compatibility condition (10.6.3).

We now obtain the following convergence result, analogous to Theorems 10.3.3 and 10.6.4.

Theorem 10.6.7 *Let* $v \in C^4(\overline{R})$ *be a solution to Neumann problem (10.6.1)-(10.6.2). Let* $\bar{\mathbf{u}} = [\mathbf{u}_0\ \lambda]^T$ *be a solution to equation (10.6.34). Then*

$$|\lambda| = \mathcal{O}(\Delta x^2)$$
$$\|\mathbf{u}_0 - \mathbf{v}\|_\infty = \mathcal{O}(\Delta x^2).$$

Proof: See [21], page 71.

Remark: We should note that the nonsymmetry in the matrix associated with the second order approximation can be approached by another method. If we define the $(M+1) \times (M+1)$ diagonal matrices

$$D_1 = \begin{pmatrix} \frac{1}{4} & 0 & \cdots & & \\ 0 & \frac{1}{2} & 0 & \cdots & \\ & \ddots & & & \\ & \cdots & 0 & \frac{1}{2} & 0 \\ & & \cdots & 0 & \frac{1}{4} \end{pmatrix},$$

$$D_2 = 2D_1 = \begin{pmatrix} \frac{1}{2} & 0 & \cdots & & \\ 0 & 1 & 0 & \cdots & \\ & \ddots & & & \\ & \cdots & 0 & 1 & 0 \\ & & \cdots & 0 & \frac{1}{2} \end{pmatrix}$$

and the $(M+1) \times (M+1)$ block diagonal matrix

$$D = \begin{pmatrix} D_1 & \Theta & \cdots & & \\ \Theta & D_2 & \Theta & \cdots & \\ & \ddots & & & \\ & \cdots & \Theta & D_2 & \Theta \\ & & \cdots & \Theta & D_1 \end{pmatrix},$$

then the matrix DA is symmetric. Hence, instead of considering equation (10.6.26), we consider equation

$$DA\mathbf{u} = D\mathbf{f}. \tag{10.6.36}$$

The same results used in Section 10.6.1 can be used with equation (10.6.36) to obtain solvability conditions, etc. The results obtained in this manner are equivalent to those given in Propositions 10.6.5 and 10.6.6.

10.6.3 Second Order Approximation on an Offset Grid

In Section 2.3 we showed that when we have Neumann boundary conditions, it is more logical to use an offset grid, i.e., a grid where the boundary points fall halfway between the grid points instead of on the grid points, for example, if we consider the grid

$$\begin{aligned} G = \{(x_j, y_k) : x_j &= (j-1)\Delta x + \Delta x/2,\ j = 0, \ldots, M, \\ y_k &= (k-1)\Delta x + \Delta x/2,\ k = 0, \ldots, M\} \end{aligned} \tag{10.6.37}$$

where $\Delta x = \Delta y = 1/(M-1)$. A picture of the grid is given in Figure 10.6.3. Note that there are no grid points on the boundaries of the region and the points associated with $j = 0$, $j = M$, $k = 0$, and $k = M$ are fictitious points outside of $[0,1] \times [0,1]$. As we have done before, we consider the equation

$$\frac{1}{\Delta x^2}\left(\delta_x^2 + \delta_y^2\right) u_{jk} = F_{jk}, \quad j,k = 1, \ldots, M-1. \tag{10.6.38}$$

Using a centered approximation for the derivatives applied at the boundary (which are not grid points), we get the following approximations of the Neumann boundary conditions.

$$-\frac{u_{1k} - u_{0k}}{\Delta x} = g_{1/2\,k}, \quad k = 1, \ldots, M-1 \tag{10.6.39}$$

$$\frac{u_{Mk} - u_{M-1\,k}}{\Delta x} = g_{M-1/2\,k}, \quad k = 1, \ldots, M-1 \tag{10.6.40}$$

$$-\frac{u_{j1} - u_{j0}}{\Delta x} = g_{j\,1/2}, \quad j = 1, \ldots, M-1 \tag{10.6.41}$$

$$\frac{u_{jM} - u_{j\,M-1}}{\Delta x} = g_{j\,M-1/2}, \quad j = 1, \ldots, M-1. \tag{10.6.42}$$

We note specifically that evaluating g at $j = \frac{1}{2}$, $j = M - \frac{1}{2}$, $k = \frac{1}{2}$, and $k = M - \frac{1}{2}$ centers the normal derivative evaluation on the boundary of the domain.

If we then proceed as we have done in the last two sections, use equations (10.6.39)–(10.6.42) to eliminate u_{j0}, u_{jM}, u_{0k}, and u_{Mk} from the equations given in (10.6.38), and write this system as a matrix equation, we get the same system that we got in Section 10.6.1. Hence, equations

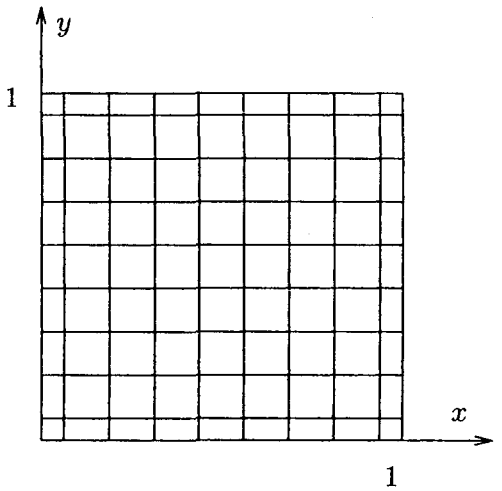

FIGURE 10.6.3. An example of a two dimensional offset grid.

(10.6.38)–(10.6.42) can be written in matrix form as equation (10.6.9). We should recall that this is the same type of situation that we found in Example 2.3.5 (where we found that exactly the same numerical scheme that was not consistent with respect to the usual grid, and hence could not be proved convergent by the Lax Theorem, was consistent with respect to the offset grid, and could be proved convergent by the Lax Theorem).

The next obvious steps that should be taken are to prove analogues of Propositions 10.6.2 and 10.6.3 and Theorem 10.6.4. Clearly, since the system of equations that we consider on the offset grid is the same system considered in Section 10.6.1, it is not necessary to prove analogues to Propositions 10.6.2 and 10.6.3. One difference between the approach taken in Section 10.6.1 and here is that the discrete compatibility condition with respect to the offset grid appears to be a better approximation of the analytic compatibility condition (i.e., we actually integrate over the entire region) than in Section 10.6.1. Of course, because we consider exactly the same matrix equation for the offset grid as we did in Section 10.6.1, the discrete compatibility conditions are the same (the discrete compatibility condition was due to the equation being solved, not the grid).

The major difference between considering equation (10.6.9) as an approximation to our problem on the usual grid versus on the offset grid is the convergence as Δx approaches zero. Since the approximation of the boundary conditions is now a second order approximation, we obtain the following convergence theorem.

Theorem 10.6.8 *Let the function $v \in C^4(\overline{R})$ be a solution to Neumann problem (10.6.1)–(10.6.2). Let $\bar{\mathbf{u}} = [\mathbf{u}_0\ \lambda]^T$ be a solution to equation (10.6.18) with respect to the offset grid (10.6.37). Then*

$$|\lambda| = \mathcal{O}(\Delta x^2)$$
$$\|\mathbf{u}_0 - \mathbf{v}\|_\infty = \mathcal{O}(\Delta x^2).$$

Hence, we see the dichotomy. From Theorem 10.6.4 we see that the solution of equation (10.6.18) converges first order in Δx to the solution of problem (10.6.1)–(10.6.2) when we consider equation (10.6.18) to approximate problem (10.6.1)–(10.6.2) on the usual grid. However, the same solution of equation (10.6.18) converges to the solution of problem (10.6.1)–(10.6.2) second order in Δx when we consider equation (10.6.18) to approximate problem (10.6.1)–(10.6.2) on the offset grid. Clearly, it pays to choose carefully the grid to be used.

10.7 Numerical Solution of Neumann Problems

10.7.1 Introduction

As we did with the Dirichlet problem, now that we know that the solution to the discretized Neumann problem converges to the solution of the analytic Neumann problem and we know that though the discrete problem is not uniquely solvable, it is uniquely solvable to an additive constant, we turn to the task of solving the discrete problems. At this point we know enough about our discrete problems to know that the numerical solution of these problems will not be as straightforward as it was for their Dirichlet counterpart. A summary of our situation is as follows.

- We wish to solve either system (10.6.9) or (10.6.26) but know that neither of these systems has a unique solution.
- Most often—when the discrete compatibility condition is not satisfied because it is only an approximation to the analytic compatibility condition—a solution to systems (10.6.9) and (10.6.26) does not exist.
- The situation is not as bad as it seems. Systems (10.6.18) and (10.6.34) are uniquely solvable and give approximate solutions to the discrete Neumann problem.

Thus, it seems clear that the most obvious approach is to consider solving equations (10.6.18) or (10.6.34). Of course, we do not want to try to solve either equation (10.6.18) or (10.6.34) directly, so we consider iterative solvers. If we inspect both of these systems, we find the following facts.

- If the matrix A is symmetric, then matrix $\bar{A}$ in equation (10.6.18) is symmetric. Even though matrix $\bar{A}$ in equation (10.6.34) is not symmetric, $\bar{A}$ can easily be symmetrized using the matrix D defined in the Remark, page 383.
- Neither of the matrices $\bar{A}$ are positive definite (due to the zero on the diagonal).
- Neither of the matrices $\bar{A}$ are consistently ordered.
- The matrix $\bar{A}$ inherits most of its eigenvalues and eigenvectors from the matrix A. If $\mu \neq 0$ is an eigenvalue of A and $\mathbf{x}$ the eigenvector associated with μ, then μ is an eigenvalue of $\bar{A}$ associated with the eigenvector $[\mathbf{x}\ \ 0]^T$.
- The zero eigenvalue of A (and its associated eigenvector $\mathbf{1}$) corresponds to the eigenvalue $\mu = \sqrt{L}$ and eigenvector $\mathbf{x} = [\mathbf{1}^T\ \ \sqrt{L}]^T$ of $\bar{A}$ (where $L = (M-1)^2$ and $L = (M+1)^2$ for the cases of first and second order approximation, respectively).

Thus *we cannot apply Jacobi, Gauss-Seidel, or SOR relaxation schemes to solve equations (10.6.18) or (10.6.34)* (if we write $\bar{A}$ as $L+D+U$, D^{-1} will not exist).

10.7.2 Residual Correction Schemes

A different approach is to try to apply the iterative schemes directly to equation (10.6.9) or (10.6.26). It should be clear that we can apply the residual correction algorithms to these problems. Because A is not invertible, it is not clear that the iterations will not converge to an approximate solution of either (10.6.9) or (10.6.26). To be specific, *we consider solving equation (10.6.26)* and assume that $\mathbf{f}$ satisfies $(\mathbf{f}, \mathbf{u}^*) = 0$ (the discrete compatibility condition is satisfied).

It is not difficult to see that since $\mathbf{1}$ is an eigenvector of A associated with the zero eigenvalue,

$$R_J\mathbf{1} = (I - BA)\mathbf{1} = \mathbf{1} - \boldsymbol{\theta} = (1)\mathbf{1},$$

i.e., $\lambda = 1$ is an eigenvalue of the matrix R_J. Hence, *the spectral radius of R_J will not be less than one.*

This is not terrible. *Suppose that R is an iteration matrix associated with a residual correction scheme for solving equation (10.6.26), the eigenvalues of R satisfy*

$$\lambda_1 = 1 > |\lambda_2| \geq \cdots |\lambda_L|,$$

the eigenvector associated with $\lambda_1 = 1$ is $\mathbf{x}_1 = \mathbf{1}$, and the eigenvectors of the matrix R are independent. An analysis similar to that used in Section 10.5.1