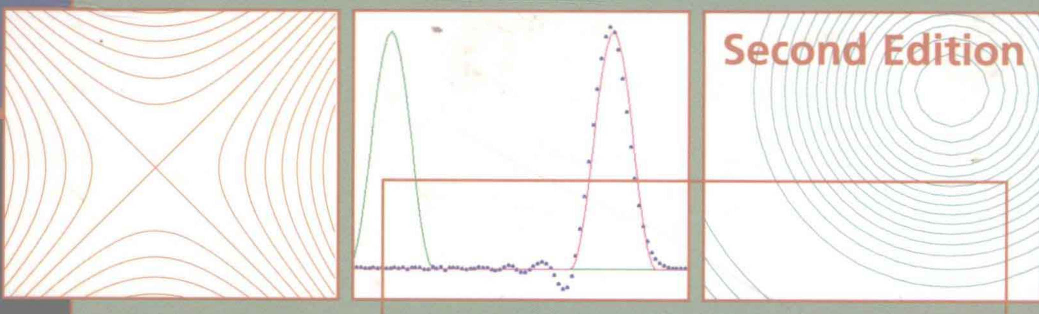
Finite Difference Schemes and Partial Differential Equations



Second Edition

John C. Strikwerda

siam

Finite Difference Schemes and Partial Differential Equations

Second Edition

John C. Strikwerda
University of Wisconsin–Madison
Madison, Wisconsin

siam

Society for Industrial and Applied Mathematics
Philadelphia

This SIAM edition is a second edition of the work first published by Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.

10 9 8 7 6 5 4 3 2

Library of Congress Cataloging-in-Publication Data

Strikwerda, John C., 1947-
Finite difference schemes and partial differential equations / John C. Strikwerda. — 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-89871-567-9
1. Differential equations, Partial—Numerical solutions. 2. Finite differences. I. Title.

QA374.S88 2004
518'.64—dc22

2004048714

Finite Difference Schemes and Partial Differential Equations

Preface to the Second Edition

I am extremely gratified by the wide acceptance of the first edition of this textbook. It confirms that there was a need for a textbook to cover the basic theory of finite difference schemes for partial differential equations, and I am pleased that this textbook filled some of that need.

I am very appreciative that SIAM has agreed to publish this second edition of the text. Many users of this textbook are members of SIAM, and I appreciate the opportunity to serve that community with this improved text

This second edition incorporates a number of changes, a few of which appeared in later printings of the first edition. An important modification is the inclusion of the notion of a stability domain in the definition of stability. The incompleteness of the original definition was pointed out to me by Prof. Ole Hald. In some printings of the first edition the basic definition was modified, but now the notion of a stability domain is more prevalent throughout the text.

A significant change is the inclusion of many more figures in the text. This has made it easier to illustrate several important concepts and makes the material more understandable. There are also more tables of computational results that illustrate the properties of finite difference schemes.

There are a few small changes to the layout requested by SIAM. Among these are that the end-of-proof mark has been changed to an open box, □, rather than the filled-in box used in the first edition.

I did not add new chapters to the second edition because that would have made the text too long and because there are many other texts and research monographs that discuss material beyond the scope of this text.

I offer my thanks to the many students who have taken my course using the textbook. They have encouraged me and given a great many suggestions that have improved the exposition. To them goes much of the credit for finding the typographical errors and mistakes that appeared in the first edition's text and exercises.

My special thanks is given to those former students, John Knox, Young Lee, Dongho Shin, and Suzan Stodder, for their many thoughtful suggestions.

John C. Strikwerda
March 2004

Preface to the First Edition

This text presents the basic theory of finite difference schemes applied to the numerical solution of partial differential equations. It is designed to be used as an introductory graduate text for students in applied mathematics, engineering, and the sciences, and with that in mind, presents the theory of finite difference schemes in a way that is both rigorous and accessible to the typical graduate student in the course. The two aims of the text are to present the basic material necessary to do scientific computation with finite difference schemes and to present the basic theory for understanding these methods.

The text was developed for two courses: a basic introduction to finite difference schemes for partial differential equations and an upper level graduate course on the theory related to initial value problems. Because students in these courses have diverse backgrounds in mathematics, the text presumes knowledge only through advanced calculus, although some mathematical maturity is required for the more advanced topics. Students taking an introduction to finite difference schemes are often acquainted with partial differential equations, but many have not had a formal course on the subject. For this reason, much of the necessary theory of partial differential equations is developed in the text.

The chief motivation for this text was the desire to present the material on time-dependent equations, Chapters 1 through 11, in a unified way that was accessible to students who would use the material in scientific and engineering studies. Chapters 1 through 11 contain much that is not in any other textbook, but more important, the unified treatment, using Fourier analysis, emphasizes that one can study finite difference schemes using a few powerful ideas to understand most of their properties. The material on elliptic partial differential equations, Chapters, 12, 13, and 14, is intended to be only an introduction; it should enable students to progress to more advanced texts and implement the basic methods knowledgably.

Several distinctive features of this textbook are:

- The fundamental concepts of convergence, consistency, and stability play an important role from the beginning.
- The concept of order of accuracy of a finite difference scheme is carefully presented with a single basic method of determining the order of accuracy of a scheme.
- Convergence proofs are given relating the order of accuracy of the scheme to that of the solution. A complete proof of the Lax–Richtmyer equivalence theorem, for the simple case of constant coefficient equations, is presented using methods accessible to most students in the course.
- Fourier analysis is used throughout the text to give a unified treatment of many of the important ideas.
- The basic theory of well-posed initial value problems is presented.
- The basic theory of well-posed initial-boundary value problems is presented for both partial differential equations and finite difference schemes.

A suggested one-semester introductory course can cover most of the material in Chapters 1, 2, 3, 5, 6, 7, 12, 13, and 14 and parts of Chapters 4 and 10. A more advanced course could concentrate on Chapters 9, 10, and 11.

In many textbooks on finite difference schemes, the discussion of the von Neumann stability condition does not make it clear when one may use the restricted condition and when one must use the general condition. In this text, theorems showing when the restricted condition may be used are stated and proved. The treatment given here was motivated by discussions with engineers and engineering students who were using the restricted condition when the more general condition was called for.

The treatment of accuracy of finite difference schemes is new and is an attempt to make the method for analyzing accuracy a rigorous procedure, rather than a grab-bag of quite different methods. This treatment is a result of queries from students who used textbook methods but were confused because they employed the wrong "trick" at the wrong time. Because many applications involve inhomogeneous equations, I have included the forcing function in the analysis of accuracy.

The convergence results of Chapter 10 are unique to this textbook. Both students and practicing computational engineers are often puzzled about why second-order accurate schemes do not always produce solutions that are accurate of second order. Indeed, some texts give students the impression that solutions to finite difference schemes are always computed with the accuracy of the scheme. The important results in Chapter 10 show how the order of accuracy of the scheme is related to the accuracy of the solution and the smoothness of the solution.

The material on Schur and von Neumann polynomials in Chapter 4 also appears in a textbook for the first time. Tony Chan deserves credit for calling my attention to Miller's method, which should be more widely known. The analysis of stability for multilevel, higher order accurate schemes is not practical without methods such as Miller's.

There are two topics that, regretfully, have been omitted from this text due to limitations of time and space. These are nonlinear hyperbolic equations and the multigrid methods for elliptic equations. Also, it would have been nice to include more material on variable grids, grid generation techniques, and other topics related to actual scientific computing. But I have decided to leave these embellishments to others or to later editions.

The numbering of theorems, lemmas, and corollaries is done as a group. That is, the corollary after Theorem 2.2.1 is numbered 2.2.2 and the next theorem is Theorem 2.2.3. The end of each proof is marked with the symbol ∎ and the end of each example is marked with the symbol □.

Many students have offered comments on the course notes from which this book evolved and they have improved the material immensely. Special thanks go to Scott Markel, Naomi Decker, Bruce Wade, and Poon Fung for detecting many typographical errors. I also acknowledge the reviewers, William Coughran, AT&T Bell Laboratories; Max Gunzberger, Carnegie-Mellon University; Joseph Oliger, Stanford University; Nick Trefethen, Massachusetts Institute of Technology; and Bruce Wade, Cornell University, for their helpful comments.

John C. Strikwerda
April 1989

Contents

Chapter 1

Hyperbolic Partial Differential Equations

We begin our study of finite difference methods for partial differential equations by considering the important class of partial differential equations called hyperbolic equations. In later chapters we consider other classes of partial differential equations, especially parabolic and elliptic equations. For each of these classes of equations we consider prototypical equations, with which we illustrate the important concepts and distinguishing features associated with each class. The reader is referred to other textbooks on partial differential equations for alternate approaches, e.g., Folland [18], Garabedian [22], and Weinberger [68]. After introducing each class of differential equations we consider finite difference methods for the numerical solution of equations in the class.

We begin this chapter by considering the simplest hyperbolic equation and then extend our discussion to include hyperbolic systems of equations and equations with variable coefficients. After the basic concepts have been introduced, we begin our discussion of finite difference schemes. The important concepts of convergence, consistency, and stability are presented and shown to be related by the Lax–Richtmyer equivalence theorem. The chapter concludes with a discussion of the Courant–Friedrichs–Lewy condition and related topics.

1.1 Overview of Hyperbolic Partial Differential Equations

The One-Way Wave Equation

The prototype for all hyperbolic partial differential equations is the one-way wave equation:

$$u_t + au_x = 0, \tag{1.1.1}$$

where a is a constant, t represents time, and x represents the spatial variable. The subscript denotes differentiation, i.e., $u_t = \partial u/\partial t$. We give $u(t, x)$ at the initial time, which we always take to be 0—i.e., $u(0, x)$ is required to be equal to a given function $u_0(x)$ for all real numbers x —and we wish to determine the values of $u(t, x)$ for positive values of t. This is called an *initial value problem*.

By inspection we observe that the solution of (1.1.1) is

$$u(t, x) = u_0(x - at). \tag{1.1.2}$$

(Actually, we know only that this is *a* solution; we prove later that this is the unique solution.)

The formula (1.1.2) tells us several things. First, the solution at any time t_0 is a copy of the original function, but shifted to the right, if a is positive, or to the left, if a is negative, by an amount $|a|t_0$. Another way to say this is that the solution at (t, x) depends only on the value of $\xi = x - at$. The lines in the (t, x) plane on which $x - at$ is constant are called *characteristics.* The parameter a has dimensions of distance divided by time and is called the speed of propagation along the characteristic. Thus the solution of the one-way wave equation (1.1.1) can be regarded as a wave that propagates with speed a without change of shape, as illustrated in Figure 1.1.

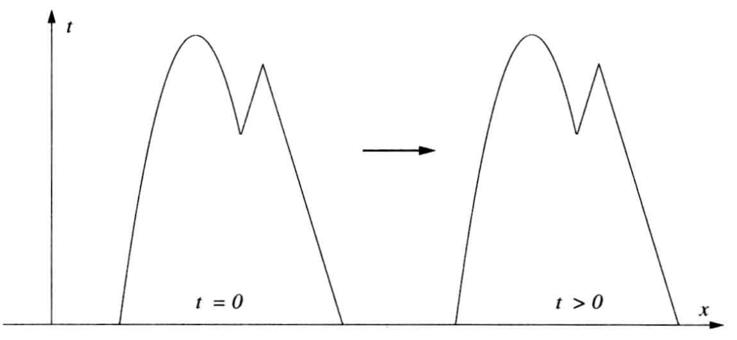


Figure 1.1. *The solution of the one-way wave equation is a shift.*

Second, whereas equation (1.1.1) appears to make sense only if u is differentiable, the solution formula (1.1.2) requires no differentiability of u_0. In general, we allow for discontinuous solutions for hyperbolic problems. An example of a discontinuous solution is a shock wave, which is a feature of solutions of nonlinear hyperbolic equations.

To illustrate further the concept of characteristics, consider the more general hyperbolic equation

$$\begin{aligned} u_t + au_x + bu &= f(t, x), \\ u(0, x) &= u_0(x), \end{aligned} \tag{1.1.3}$$

where a and b are constants. Based on our preceding observations we change variables from (t, x) to (τ, ξ), where τ and ξ are defined by

$$\tau = t, \qquad \xi = x - at.$$

The inverse transformation is then

$$t = \tau, \qquad x = \xi + a\tau,$$

and we define $\tilde{u}(\tau, \xi) = u(t, x)$, where (τ, ξ) and (t, x) are related by the preceding relations. (Both u and $\tilde{u}$ represent the same function, but the tilde is needed to distinguish between the two coordinate systems for the independent variables.) Equation (1.1.3) then becomes

$$\frac{\partial \tilde{u}}{\partial \tau} = \frac{\partial t}{\partial \tau} u_t + \frac{\partial x}{\partial \tau} u_x$$
$$= u_t + a u_x = -bu + f(\tau, \xi + a\tau).$$

So we have

$$\frac{\partial \tilde{u}}{\partial \tau} = -b\tilde{u} + f(\tau, \xi + a\tau).$$

This is an ordinary differential equation in τ and the solution is

$$\tilde{u}(\tau, \xi) = u_0(\xi) e^{-b\tau} + \int_0^\tau f(\sigma, \xi + a\sigma) e^{-b(\tau - \sigma)}\, d\sigma.$$

Returning to the original variables, we obtain the representation for the solution of equation (1.1.3) as

$$u(t, x) = u_0(x - at) e^{-bt} + \int_0^t f\big(s, x - a(t - s)\big) e^{-b(t-s)}\, ds. \tag{1.1.4}$$

We see from (1.1.4) that $u(t, x)$ depends only on values of (t', x') such that $x' - at' = x - at$, i.e., only on the values of u and f on the characteristic through (t, x) for $0 \le t' \le t$.

This method of solution of (1.1.3) is easily extended to nonlinear equations of the form

$$u_t + a u_x = f(t, x, u). \tag{1.1.5}$$

See Exercises 1.1.5, 1.1.4, and 1.1.6 for more on nonlinear equations of this form.

Systems of Hyperbolic Equations

We now examine systems of hyperbolic equations with constant coefficients in one space dimension. The variable u is now a vector of dimension d.

Definition 1.1.1. *A system of the form*

$$u_t + A u_x + B u = F(t, x) \tag{1.1.6}$$

is hyperbolic if the matrix A is diagonalizable with real eigenvalues.

By saying that the matrix A is *diagonalizable,* we mean that there is a nonsingular matrix P such that PAP^{-1} is a diagonal matrix, that is,

$$PAP^{-1} = \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_d \end{pmatrix} = \Lambda.$$

The eigenvalues a_i of A are the characteristic speeds of the system. Under the change of variables $w = Pu$ we have, in the case $B = 0$,

$$w_t + \Lambda\, w_x = PF(t,x) = \tilde{F}(t,x)$$

or

$$w_t^i + a_i\, w_x^i = \tilde{f}^i(t,x),$$

which is the form of equation (1.1.3). Thus, when matrix B is zero, the one-dimensional hyperbolic system (1.1.6) reduces to a set of independent scalar hyperbolic equations. If B is not zero, then in general the resulting system of equations is coupled together, but only in the undifferentiated terms. The effect of the lower order term, Bu, is to cause growth, decay, or oscillations in the solution, but it does not alter the primary feature of the propagation of the solution along the characteristics. The definition of hyperbolic systems in more than one space dimension is given in Chapter 9.

Example 1.1.1. As an example of a hyperbolic system, we consider the system

$$\begin{aligned} u_t + 2u_x + v_x &= 0, \\ v_t + u_x + 2v_x &= 0, \end{aligned}$$

which can be written as

$$\begin{pmatrix} u \\ v \end{pmatrix}_t + \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}_x = 0.$$

As initial data we take

$$\begin{aligned} u(0,x) = u_0(x) &= \begin{cases} 1 & \text{if } |x| \le 1, \\ 0 & \text{if } |x| > 1, \end{cases} \\ v(0,x) &= 0. \end{aligned}$$

By adding and subtracting the two equations, the system can be rewritten as

$$\begin{aligned} (u+v)_t + 3(u+v)_x &= 0, \\ (u-v)_t + \ (u-v)_x &= 0 \end{aligned}$$

or

$$\begin{aligned} w_t^1 + 3w_x^1 &= 0, \qquad w^1(0,x) = u_0(x), \\ w_t^2 + w_x^2 &= 0, \qquad w^2(0,x) = u_0(x). \end{aligned}$$

The matrix P is $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ for this transformation. The solution is, therefore,

$$\begin{aligned} w^1(t,x) &= w_0^1(x - 3t), \\ w^2(t,x) &= w_0^2(x - t) \end{aligned}$$

or

$$u(t,x) = \tfrac{1}{2}(w^1 + w^2) = \tfrac{1}{2}\left[u_0(x-3t) + u_0(x-t)\right],$$

$$v(t,x) = \tfrac{1}{2}(w^1 - w^2) = \tfrac{1}{2}\left[u_0(x-3t) - u_0(x-t)\right].$$

These formulas show that the solution consists of two independent parts, one propagating with speed 3 and one with speed 1. □

Equations with Variable Coefficients

We now examine equations for which the characteristic speed is a function of t and x. Consider the equation

$$u_t + a(t,x)u_x = 0 \tag{1.1.7}$$

with initial condition $u(0,x) = u_0(x)$, which has the variable speed of propagation $a(t,x)$. If, as we did after equation (1.1.3), we change variables to τ and ξ, where $\tau = t$ and ξ is as yet undetermined, we have

$$\begin{aligned}\frac{\partial \tilde{u}}{\partial \tau} &= \frac{\partial t}{\partial \tau}u_t + \frac{\partial x}{\partial \tau}u_x \\ &= u_t + \frac{\partial x}{\partial \tau}u_x.\end{aligned}$$

In analogy with the constant coefficient case, we set

$$\frac{dx}{d\tau} = a(t,x) = a(\tau,x).$$

This is an ordinary differential equation for x giving the speed along the characteristic through the point (τ, x) as $a(\tau, x)$. We set the initial value for the characteristic curve through (τ, x) to be ξ. Thus the equation (1.1.7) is equivalent to the system of ordinary differential equations

$$\begin{aligned}\frac{d\tilde{u}}{d\tau} &= 0, & \tilde{u}(0,\xi) &= u_0(\xi), \\ \frac{dx}{d\tau} &= a(\tau,x), & x(0) &= \xi.\end{aligned} \tag{1.1.8}$$

As we see from the first equation in (1.1.8), u is constant along each characteristic curve, but the characteristic determined by the second equation need not be a straight line. We now present an example to illustrate these ideas.

Example 1.1.2. Consider the equation

$$u_t + x\,u_x = 0,$$

$$u(0,x) = \begin{cases} 1 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.}\end{cases}$$

Corresponding to the system (1.1.8) we have the equations

$$\frac{d\tilde{u}}{d\tau} = 0, \qquad \frac{dx}{d\tau} = x, \qquad x(0) = \xi.$$

The general solution of the differential equation for $x(\tau)$ is $x(\tau) = ce^{\tau}$. Because we specify that ξ is defined by $x(0) = \xi$, we have $x(\tau) = \xi\, e^{\tau}$, or $\xi = xe^{-t}$. The equation for $\tilde{u}$ shows that $\tilde{u}$ is independent of τ, so by the condition at τ equal to zero we have that

$$\tilde{u}(\tau, \xi) = u_0(\xi).$$

Thus

$$u(t, x) = \tilde{u}(\tau, \xi) = u_0(\xi) = u_0(xe^{-t}).$$

So we have, for $t > 0$,

$$u(t, x) = \begin{cases} 1 & \text{if } 0 \le x \le e^t, \\ 0 & \text{otherwise.} \ \square \end{cases}$$

As for equations with constant coefficients, these methods apply to nonlinear equations of the form

$$u_t + a(t, x)u_x = f(t, x, u), \tag{1.1.9}$$

as shown in Exercise 1.1.9. Equations for which the characteristic speeds depend on u, i.e., with characteristic speed $a(t, x, u)$, require special care, since the characteristic curves may intersect.

Systems with Variable Coefficients

For systems of hyperbolic equations in one space variable with variable coefficients, we require uniform diagonalizability. (See Appendix A for a discussion of matrix norms.)

Definition 1.1.2. *The system*

$$u_t + A(t, x)\, u_x + B(t, x)u = F(t, x) \tag{1.1.10}$$

with

$$u(0, x) = u_0(x)$$

is hyperbolic if there is a matrix function $P(t, x)$ *such that*

$$P(t, x)\, A(t, x)\, P^{-1}(t, x) = \Lambda(t, x) = \begin{pmatrix} a_1(t, x) & & 0 \\ & \ddots & \\ 0 & & a_d(t, x) \end{pmatrix}$$

is diagonal with real eigenvalues and the matrix norms of $P(t, x)$ *and* $P^{-1}(t, x)$ *are bounded in* x *and* t *for* $x \in R$, $t \ge 0$.