O. B. Widlund*

On Difference Methods for Parabolic Equations and Alternating Direction Implicit Methods for Elliptic Equations†

Abstract: Some aspects are discussed of the development of the theory of difference approximations to parabolic equations in its relation to the basic development of difference methods for hyperbolic and elliptic equations by Courant, Friedrichs and Lewy, *Math. Ann.* 100, 32 (1928). The present paper also deals with the related problem of establishing the convergence of alternating direction implicit methods for elliptic problems.

Introduction

In their famous paper of 1928, Courant, Friedrichs, and Lewy[6(‡)] paid little attention to partial differential equations of parabolic type. They considered only a quite special difference approximation for a heat-conduction equation with constant coefficients. The simple structure of their difference equation enabled them to write down its solution as a sum and to show that this sum converges to the well-known solution of the differential equation when the mesh size goes to zero. Such an approach does not lend itself to straightforward generalizations to more complicated problems. In spite of this, their paper has greatly influenced the development of the theory of difference approximations to parabolic equations in that many of the ideas which they developed for hyperbolic and elliptic equations have proved useful in the study of parabolic problems. For a discussion of these ideas we refer to the papers by Lax[21] and Parter[27] in this issue.

In this paper we will discuss some aspects of the development of the theory and point to some mathematical techniques that are now available in the study of difference methods for parabolic equations and the related problem of establishing the convergence of alternating direction implicit methods for elliptic problems.

Initial value problems for linear equations

Let us first consider the initial value problem for a parabolic equation of second order and with one space variable,

$$\frac{\partial u}{\partial t} = a_0(x, t)\frac{\partial^2 u}{\partial x^2} + a_1(x, t)\frac{\partial u}{\partial x} + a_2(x, t)u + f(x, t),$$

$$-\infty < x < \infty, \qquad 0 \le t \le T_0 < \infty, \quad (1)$$

$$u(x, 0) = u_0(x), \qquad -\infty < x < \infty,$$

where $u_0(x)$ and $f(x, t)$ are given, bounded functions. The coefficients are sufficiently smooth bounded functions and a_0 is bounded from below by a strictly positive constant. Such equations have been extensively studied; cf. Ref. 13. Among other things it is known that the solution depends continuously on the initial value u_0 and the inhomogenous term f. Thus, there is a constant C independent of u_0 and f such that for $t \in [0, T_0]$,

$$\max_{x,t} |u(x, t)| \le C(\max_{x} |u_0(x)| + \max_{x,t} |f(x, t)|). \quad (2)$$

This is, of course, to be expected from physical considerations if we think of (1) as a mathematical model for heat flow. When an inequality like (2) is fulfilled for some differential equation and in some norm we will call such a problem well posed in that particular norm.

To set up difference approximations to (1) we first introduce lattices of mesh points,

$$R(h, \lambda) = \{(x, t): x = 0, \pm h, \pm 2h, \cdots ;$$
$$t = 0, k, 2k, \cdots, Nk = T_0; k = \lambda h^2\}.$$

For simplicity we assume that there always is a natural number N such that $Nk = T_0$. When we study the convergence of the solution of a difference scheme to that of a

* Department of Computer Sciences, University of Uppsala, Sweden.
† Prepared under Grant NSF GP-4436 while the author was visiting the University of California at Berkeley.
‡ Superscripts refer to the bibliography at the end of the paper.

differential equation we will consider a whole set of lattices $R(h_i, \lambda_i)$ for which h_i and $k_i = \lambda_i h_i^2$ go to zero. In many cases the mathematics of the problem is very much simplified by assuming that λ_i is a positive constant. In fact much of the theory has been developed only for uniformly bounded λ_i and it is not known to what extent it can be generalized. Why such an extension would be of great interest is discussed in the next two sections of the paper. In the rest of this section we will always suppose that k/h^2, or in the case of an equation of order $2m$ that k/h^{2m}, is constant.

All useful difference approximations to (1) which are of one-step type have the form

$$\sum b_r(x, t, h) T^r v(x, t + k, h)$$
$$= \sum c_r(x, t, h) T^r v(x, t, h) + kf(x, t),$$
$$(x, t) \in R(h, \lambda),$$
$$v(x, 0, h) = u_0(x), \qquad x = 0, \pm h, \cdots. \qquad (3)$$

Here the sums are finite, the coefficients sufficiently smooth functions of x, t and h, and T is the translation operator defined by $T\varphi(x) = \varphi(x + h)$. To be able to compute $v(x, k, h)$, $v(x, 2k, h)$, etc. from (3) we must be able to invert the operator on the left-hand side. This is, of course, trivial if only one of the b_r's is different from zero. We call such methods *explicit*. In the opposite case, the *implicit* case, we must demand that the corresponding algebraic problem is finite dimensional.

In order to ensure that the difference equation will have something to do with the differential equation we demand that (3) is consistent (formally convergent) to (1); i.e., for all sufficiently smooth functions $\varphi(x, t)$,

$$\lim_{h\to 0} \left\{ \left(\frac{\partial \varphi}{\partial t} - a_0 \frac{\partial^2 \varphi}{\partial x^2} - a_1 \frac{\partial \varphi}{\partial x} - a_2 \varphi \right) - \frac{1}{k} \left[\sum b_r T^r \varphi(x, t + k) - \sum c_r T^r \varphi(x, t) \right] \right\} = 0.$$

By Taylor expansion it is easy to check whether or not a difference equation is consistent to a given differential equation. It is also very simple to construct many consistent difference equations to any given differential equation. But consistency alone does not guarantee that the difference between the solutions of (1) and (3) is small for small mesh sizes. On the contrary, as pointed out in the paper by Courant, Friedrichs and Lewy, we can expect many consistent difference equations to have an error growth which will make them completely useless for calculations of approximate solutions to the differential equation. As is shown in Ref. 29 we will get uniform convergence of the solution of (3) to the solution of the well-posed problem (1) if the difference method is not only consistent but also stable. Equation (3) is said to be stable if there is a constant C, independent of u_0, f and the mesh size such that

$$\max_{x,t} |v(x, t, h)| \leq C(\max_x |u_0(x)| + \max_{x,t} |f(x, t)|). \qquad (4)$$

Observe that (4) is the analogue to (2) for the difference equation. For a detailed discussion of stability, convergence, etc., we refer to Ref. 29.

We will now discuss two methods to prove the stability of consistent difference schemes. We first show that (4) holds if (a) the method is explicit, i.e. $b_0 = 1, b_r = 0, r \neq 0$; and (b) $c_r \geq 0$. For from (3) and (b) we immediately get

$$\max_x |v(x, t + k, h)| \leq \max_x \sum c_r(x, t, h) \cdot \max_x |v(x, t, h)| + k \max_x |f(x, t)|.$$

If we expand $c_r(x, t, h)$ in Taylor series in h and use the consistency we can show that

$$\sum c_r(x, t, h) \leq 1 + Ck,$$

where C is a constant independent of x, t and h. Thus

$$\max_x |v(x, nk, h)| \leq (1 + Ck)^n \max_x |u_0(x)| + (1 + Ck)^n nk \max_{x,t} |f(x, t)|$$
$$\leq \exp(CT_0)(\max_x |u_0(x)| + T_0 \max_{x,t} |f(x, t)|).$$

For a further discussion of this technique we refer to Refs. 8 and 14.

Condition (b) above is far from necessary and it is also hard to see how the technique could be extended to systems of parabolic equations. In 1952 Fritz John[17] gave a new sufficient stability condition, introducing a mathematical technique which has since been extended to treat more complicated parabolic problems. His paper is an outstanding contribution to the understanding of difference approximations to partial differential equations. John treated the explicit case and phrased his stability condition in terms of the characteristic polynomial

$$g(x, t, \theta) = \sum c_r(x, t, 0) \exp(ir\theta).$$

He proved that if

$$\max_{x,t} |g(x, t, \theta)| \leq \exp(-\alpha\theta^2), \quad |\theta| \leq \pi,$$
$$\alpha \text{ some constant} > 0, \qquad (5)$$

the difference equation is stable. (Observe that g is a 2π-periodic function in θ.)

John's result have been extended by Aronson[1-3] and by the author.[35,36] They considered systems of partial differential equations that are parabolic in the sense of Petrowskii. (For details and definitions see Refs. 35 and 36.)

For any difference approximation, implicit or explicit, define a characteristic matrix which corresponds to the characteristic polynomial above. Denote the eigenvalues of that matrix by $\sigma_i(\mathbf{x}, t, \boldsymbol{\theta})$ where $\mathbf{x}$ and $\boldsymbol{\theta}$ are vectors with s components, where s is the number of space variables in our differential equation. It is shown in Refs. 35 and 36 that the difference scheme is stable both in the uniform norm and the L^2-norm if

$$\max_{x,t} |\sigma_i(\mathbf{x}, t, \boldsymbol{\theta})| \leq \exp\left(-\alpha \sum \theta_i^{2m}\right), \quad |\theta_i| \leq \pi,$$

$$\alpha \text{ some constant} > 0. \qquad (6)$$

This is the exact counterpart to (5). One of the main tools in the proof of this result is a matrix theorem by Kreiss[19] that gives necessary and sufficient conditions for L^2-stability for systems of difference equations with constant coefficients. For an extension of the results above to linear multistep difference schemes we refer to Refs. 35 and 36.

To what extent is (6) a necessary condition for stability and a reasonable behavior of the error? Is (6) a natural condition for consistent difference equations?

To answer the first question let us first note that we must require all σ_i to lie inside or on the unit circle, or else we violate the von Neumann condition and must expect a very rapid error growth. (Cf. Ref. 29.) Could $|\sigma_i| \leq 1$ be used as a general stability criterion instead of (6)? The answer is no, for it was shown in Ref. 36 that there are equations which fulfil this condition and still are wildly unstable. But, as is shown in Ref. 35, we can guarantee the fulfilment of (6), and thus the stability, for consistent approximations by the slightly more restrictive condition $|\sigma_i| < 1$, $\boldsymbol{\theta} \neq 0$, $|\theta_i| \leq \pi$. This condition is in fact very natural. The typical situation is, namely, that (6) is fulfilled for a consistent difference method for a neighborhood of $\boldsymbol{\theta} = 0$ and that we can adjust the value of k/h^{2m} so that $|\sigma_i| < 1$ for all $|\theta_i| \leq \pi$.

We conclude this section by pointing out that, once we have established the stability of a difference equation, it is quite simple to get an error bound if we have a differential equation with a sufficiently smooth solution. For details we again refer to Ref. 29.

The energy method

There are many interesting and important problems which cannot be treated with the theory of the previous section. Suppose, for example, that we want to design good difference methods for a mixed initial-boundary value problem for a parabolic equation and that the problem cannot be transformed into a pure initial value problem by some periodic extension. There is then good reason to expect that we will encounter many new difficulties, for it is well known from the theory of partial differential equations that the introduction of boundary values makes most problems more complicated. No general theory exists as yet which tells us how to choose convergent difference approximations to such problems. Some progress has, however, been made in the case of the one-space variable. We especially want to mention the work by Kreiss.[18,20] He showed how to construct stable differential approximations to properly posed differential equations. For this he used the energy method, which means among other things that his norm was an L^2-norm instead of the uniform norm discussed in the previous section.

It is well known from the theory of partial differential equations that it is often easier to prove results in the L^2-norm than in other norms. This is also true for the theory of difference approximations. The techniques such as summation by parts, etc., that are used in the energy method are direct counterparts to quite old tricks from the theory of partial differential equations. In many cases the energy method supplies not only estimates for the L^2-norm of the solution but also certain divided differences of the solution. By use of Sobolev inequalities these can give stability and error bounds in the maximum norm. See for example Refs. 22–26.

Let us illustrate the energy method with a simple parabolic equation,

$$\partial u/\partial t = \partial^2 u/\partial x^2, \qquad 0 < x < 1$$

$$u(x, 0) = u_0(x), \qquad u(0, t) = u(1, t) = 0.$$

We use the energy method to prove that

$$(\partial/\partial t) \int_0^1 [u(x, t)]^2 \, dx \leq 0, \quad (\partial/\partial t) \int_0^1 [u_x(x, t)]^2 \, dx \leq 0$$

so that the L^2-norm of both the solution and its first derivative with respect to x are non-increasing. To get the first inequality we multiply the equation by $u(x, t)$ and integrate by parts on the right-hand side, to get the second we multiply by $\partial^2 u/\partial x^2$ and use partial integration for the left-hand side. We can use the same technique to derive similar inequalities for the solutions of certain simple difference approximations to parabolic problems.

Just as in the theory of differential equations, this old-fashioned energy method enables us to settle certain questions for difference equations with coefficients which depend on the solution or which are not smooth. See, for example, Refs. 11, 25, 26.

Finally, we can sometimes use the energy method to prove stability and give error bounds without imposing restrictions between k and h. This gives us better error bounds than can be obtained by use of the theory of the previous section.

Alternating direction implicit methods

In order to compute approximate solutions to second-order parabolic equations with one space variable, many have chosen to use implicit schemes. The reason for this is that there are implicit methods that are stable for all values of k/h^2. Such methods are called *unconditionally stable*. This gives a great and desirable flexibility in the choice of the mesh size and we can, for example, use very long time steps. With few exceptions the explicit methods do not share this property. In the first instance it might, however, appear as if the inversion of the operator on the left-hand side of (3) might be quite time consuming. This is not necessarily so. For in many cases this operator corresponds to a positive definite tridiagonal matrix which can be inverted by Gaussian elimination using only a number of arithmetic operations which is proportional to the order of the matrix. Cf. Ref. 29.

In the two- or multi-dimensional case the situation will be somewhat different. If we write up the obvious generalization of implicit one-dimensional methods we have to invert a rather general matrix at each time step. No very fast method is available for this.

The situation improved very much when Peaceman and Rachford[28] invented the first alternating direction implicit method (ADI-method). Studying the first initial-boundary value problem

$$\partial u/\partial t = \partial^2 u/\partial x^2 + \partial^2 u/\partial y^2, \quad \text{for } (x, y) \text{ inside a unit square,} \tag{7}$$

with $u(x, y, 0) = u_0(x, y)$, $u(x, y, t) = \varphi(x, y, t)$ for (x, y) on the boundary of the square, they constructed an unconditionally stable method for which the operator on the left-hand side corresponds to the product of two matrices. These matrices are both similar via permutations to positive definite tridiagonal matrices and they can therefore be inverted very rapidly. The same idea has been used to construct many other ADI-methods. Cf. Refs. 9, 10, 12, 30.

It has turned out to be rather hard to analyze these ADI-methods for more general parabolic problems, more general regions and without restrictions on k/h^2. Considerable progress is, however, reported in Refs. 15, 16, 23, 24, and 31.

In fact we are very much interested in being able to prove results without restricting k/h^2. Now the most interesting idea, perhaps, in the paper of Peaceman and Rachford[28] is still to be mentioned. They noted that if the boundary values of (7) are independent of t we end up having the solution of Poisson's equation when $t \to \infty$. Using their difference method for the parabolic equation they then constructed an iterative method for the solution of a difference analogue to Poisson's equation. To do this they chose not only to proceed with very large time steps but also showed that if you vary the length of time steps between the different steps you can speed up the convergence of the iterative method considerably. Thus, if the time steps are cleverly chosen you can reduce the L^2-norm of the error by a factor

$$\left\{\frac{1 - [\tan(\pi h/2)]^{1/m}}{1 + [\tan(\pi h/2)]^{1/m}}\right\}^2, \tag{8}$$

in m time steps. It turns out that in this simple case this method is far superior to any other known iterative method if the number of mesh points is sufficiently large.

The method can be generalized quite easily to other second-order elliptic problems with more complicated regions. However, except when certain operators commute, the theory of Douglas[7], which was published in a companion paper with Ref. 28, does not easily extend. As was shown by Birkhoff and Varga[4] this commutativity condition is very restrictive both for the coefficients and the region, which has to be rectangular. It is, however, known from experience that this iterative method often works very well in practice in non-commutative cases although divergence sometimes has been observed.[5]

Recently the author[37] was able to extend the theory to a non-commutative case. For a class of elliptic equations with sufficiently smooth coefficients and with Dirichlet data given on a rectangle, a variant of the Peaceman-Rachford method can be made to have as large an asymptotic rate of convergence as comparable commutative cases when the mesh size is sufficiently small. The variant of the original ADI-method results from the following consideration. Instead of setting up a difference method for $\partial u/\partial t = Pu$, where P is the elliptic operator, and then trying to compute the values of the solution when $t \to \infty$ we can as well use some other parabolic equation $c(x, y)\partial u/\partial t = Pu$, c strictly positive.

It turns out that by a proper choice of c we can make our iterative scheme rapidly convergent for small values of h and good choices of the time steps.

The proof of this result is essentially based on the energy method. If the more mathematically sophisticated techniques discussed in the second section of this paper could be extended to cases with general k and h, etc., we might hope that this would lead to a better understanding of these potentially extremely powerful iterative methods.

Concluding remarks

In this paper we have tried to convey some of the flavor of the field rather than to give its complete history. For surveys, details and references on the theory for parabolic equations we refer the reader to Refs. 8, 14, 29, 32, 33, and 38 and for additional material on alternating direction implicit methods for elliptic problems, to Refs. 5, 34 and 38.

Bibliography

1. D. G. Aronson, "The stability of finite difference approximations to second order linear parabolic differential equations," *Duke Math. J.* **30**, 117–128 (1963).
2. D. G. Aronson, "On the stability of certain finite difference approximations to parabolic systems of differential equations," *Numer. Math.* **5**, 118–137 (1963); correction, *ibid.* **5**, 290 (1963).
3. D. G. Aronson, "Stability of implicit finite difference approximations for second order parabolic equations," Lecture at a symposium held at University of Maryland, May 3–8, 1965.
4. G. Birkhoff and R. S. Varga, "Implicit alternating direction methods," *Trans. A. M. S.* **92**, 13–24 (1959).
5. G. Birkhoff, R. S. Varga and D. M. Young, Jr., "Alternating-Direction-Implicit methods," *Advances in Computers*, vol. 3, Academic Press, New York, 1962.
6. R. Courant, K. Friedrichs, and H. Lewy, "On Partial Difference Equations of Mathematical Physics," *IBM Journal* **11**, 215–234 (1967). English translation of the original work, "Über die Partiellen Differenzengleichungen der Mathematischen Physik," *Math. Ann.* **100**, 32–74 (1928).
7. J. Douglas, Jr., "On the numerical integration of $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = \partial u/\partial$ by implicit methods," *J. Soc. Indust. Appl. Math.* **3**, 42–65 (1955).
8. J. Douglas, Jr., "A survey of numerical methods for parabolic differential equations," *Advances in Computers*, Vol. **2**, Academic Press, New York-London, 1961, pp. 1–52.
9. J. Douglas, Jr., "Alternating direction methods for three space variables," *Numer. Math.*, **4**, 41–63 (1962).
10. J. Douglas Jr. and J. E. Gunn, "A general formulation of alternating direction methods. Part I. Parabolic and hyperbolic problems," *Numer. Math.* **6**, 428–453 (1964).
11. J. Douglas, Jr. and B. F. Jones, "On predictor-corrector methods for nonlinear parabolic equations," *J. Soc. Indust. Appl. Math.* **11**, 195–204 (1963).
12. J. Douglas, Jr. and H. H. Rachford, "On the numerical solution of heat conduction problems in two and three space variables," *Trans. A. M. S.* **82**, 421–439 (1956).
13. A. Friedman, *Partial differential equations of parabolic type*, Prentice Hall, Englewood Cliffs, New Jersey, 1964.
14. G. E. Forsythe and W. R. Wasow, *Finite-difference methods for partial differential equations*, John Wiley and Sons, Inc., New York, 1960.
15. B. Hubbard, "Some locally one-dimensional difference schemes for parabolic equations in an arbitrary region," *Math. Comp.* **20**, 53–59 (1966).
16. B. Hubbard, "Alternating direction schemes for heat equation in a general domain," to appear.
17. F. John, "On integration of parabolic equations by difference methods, I," *Comm. Pure Appl. Math.* **5**, 155–211 (1952).
18. H. O. Kreiss, "Über die Lösung von Anfangsrandvertaufgaben für partielle Differentialgleichungen mit Hilfe von Differenzengleichungen," *Trans. Royal Inst. Techn.*, 166, (1960), Stockholm, Sweden.
19. H. O. Kreiss, "Über die Stabilitätsdefinition für Differenzengleichungen die partielle Differentialgleichungen approximieren," *BIT* **2**, pp. 153–181 (1962).
20. H. O. Kreiss, "Über implizite Differenzenmethoden für partielle Differentialgleichungen," *Numer. Math.* **5**, 24–47 (1963).
21. P. D. Lax, "Hyperbolic difference equations: a review of the Courant-Friedrichs-Lewy paper in the light of further developments," *IBM Journal* **11**, 235–238 (1967).
22. M. Lees, "A priori estimates for solutions of difference approximations to parabolic partial differential equations," *Duke Math. J.* **27**, 297–311 (1960).
23. M. Lees, "Alternating direction and semi-explicit difference methods for parabolic partial differential equations," *Numer. Math.* **3**, 398–412 (1961).
24. M. Lees, "A note on the convergence of alternating direction methods," *Math. Comp.* **16**, 70–75 (1962).
25. M. Lees, "A linear three-level difference scheme for quasilinear parabolic equations." To appear.
26. M. Lees, "An extrapolated Crank-Nicolson difference scheme for quasilinear parabolic equations." To appear.
27. S. V. Parter, "Elliptic Equations," *IBM Journal* **11**, 244–247 (1967).
28. D. W. Peaceman and H. H. Rachford, Jr. "The numerical solution of parabolic and elliptic differential equations," *J. Soc. Indust. Appl. Math.* **3**, 28–41 (1955).
29. R. D. Richtmyer, *Difference methods for initial value problems*, Interscience Publishers, Inc., New York 1957.
30. A. A. Samarskii, "An economical difference method for the solution of a multidimensional parabolic equation in an arbitrary region," *Ž. Vyčisl. Mat. i Mat. Fiz.* **2**, 787–811 (1962) (Russian); English translation, *U. S. S. R. Comput. Math. and Math. Phys.* **2**, 894–926 (1962).
31. A. A. Samarskii, "Locally one-dimensional difference schemes on non-uniform grids," (Russian), *Ž. Vyčisl. Mat. i Mat. Fiz.* **3**, 431–466 (1963).
32. V. K. Saul'yev, *Integration of equations of parabolic type by the method of nets*, Pergamon Press, Oxford, London, Edinburgh, New York, Paris, Frankfurt, 1964.
33. R. S. Varga, *Matrix Iterative Analysis*, Prentice Hall, Englewood Cliffs, New Jersey, 1962.
34. E. L. Wachspress, *Iterative Solution of Elliptic Systems and Applications to Neutron Diffusion Equations of Reactor Physics*, Prentice Hall, Englewood Cliffs, New Jersey, 1966.
35. O. B. Widlund, "On the stability of parabolic difference schemes," *Math. Comp.* **19**, 1–13 (1965).
36. O. B. Widlund, "Stability of parabolic difference schemes in the maximum norm," *Numer. Math.* **8**, 186–202 (1966).
37. O. B. Widlund, "On the rate of convergence of an alternating direction implicit method in a non-commutative case," *Math. Comp.* **20**, 500–515 (1966).
38. D. M. Young and T. G. Frank, "A survey of computer methods for solving elliptic and parabolic partial differential equations," *ICC Bulletin* **2**, No. 1 (1963).

Received August 16, 1966