A FREQUENCY ANALYSIS OF FINITE DIFFERENCE AND FINITE ELEMENT METHODS FOR INITIAL VALUE PROBLEMS

R. Vichnevetsky and F. De Schutter
Rutgers University, New Brunswick, N. J.

DCS-TR-36

Reprinted from: Advances in Computer Methods for Partial Equations. R. Vichnevetsky (editor), AICA Dept. Computer Science, Rutgers University, New Brunswick New Jersey

Dept. of Computer Science
Rutgers University
The State University of New Jersey
New Brunswick, New Jersey 08903
July, 1975.

ADVANCES IN COMPUTER METHODS FOR PARTIAL DIFFERENTIAL EQUATIONS
R. VICHNEVETSKY (EDITOR)
PUBL. A I C A - 1975

A FREQUENCY ANALYSIS OF FINITE DIFFERENCE AND FINITE ELEMENT METHODS FOR INITIAL VALUE PROBLEMS

R. Vichnevetsky and F. De Schutter
Rutgers University, New Brunswick, N.J.

1. INTRODUCTION

The popularity of finite element and spline methods has brought to the fore the need for criteria to compare the accuracy of computing algorithms. Several recent papers have addressed this problem, investigating the use of new analytical tools to achieve this goal (see e.g. Swartz and Wendroff (1974 a and b) Vichnevetsky (1973), Vichnevetsky, Tu and Steen (1974) Fix and Strang (1969)).
The approach used here consists in considering simple partial differential equations as typical models of more complex ones, and in showing that the error introduced by numerical approximations may be characterized, for those simple models by criteria which can be related to physical parameters of those equations. The mathematics are based on the expression of solutions as a sum of space-sinusoidal components for which errors can be expressed as a function of the frequency.

2. DERIVATION OF ALGORITHMS

Consider the equation

$$\frac{\partial u(x,t)}{\partial t} = X(u(x,t)); \; x \in D \; ; \; t \geq 0 \tag{1}$$

associated to an appropriate set of initial and boundary conditions, where $X(\cdot)$ is an operator containing derivatives with respect to x only. The synthesis of an important class of algorithms for the numerical solution of this equation may be viewed as a two-step process, consisting first in replacing (1) by the semi-discrete approximation

$$\sum_{\beta} A_{1,\beta} \cdot \frac{du_n}{dt} = \tilde{X}(u_n) \; ; \; u_n(t) \simeq u(n \cdot \Delta x, t) \qquad n = 1,2,\ldots\ldots \tag{2}$$

obtained by some form of finite element, spline or finite difference procedure, and then replacing the time-derivative in (2) by a discrete approximation in t, e.g. the general implicit method:

$$\sum_{\beta} A_{1,\beta} \left(\frac{u_n^{j+1} - u_n^{j}}{\Delta t} \right) = \theta \tilde{X}(u_n^{j+1}) + (1-\theta) \cdot \tilde{X}(u_n^{j}) \qquad u_n^{j} \simeq u_n(j \cdot \Delta t); \; 0 < \theta \leq 1 \tag{3}$$

(The left hand side may be expressed in operator notation. See (9) below).

We shall assume that Δt is in some sense small with respect to Δx. This will allow us to consider the errors due to the approximation contained in the spatial discretization (2) alone, and thereby compare the relative accuracy of the several competing processes used in the derivation of (2).
The formalism of derivation of algorithms of the form (2) which we intend to compare for accuracy is well known:

Finite element and B-spline methods consist in choosing a family of basis functions $\varphi_n(x)$ and in expressing an approximate solution as the linear combination:

$$u(x,t) \simeq u^*(x,t) \equiv \sum_m \varphi_m(x) \cdot v_m(t) \tag{4}$$

The coefficients v_m are implicitly determined by the collocation conditions

$$u^*(x_n,t) = u_n(t)$$

$$x_n = n \cdot \Delta x \; ; \; n = \ldots,-2,-1,0,1,2,\ldots\ldots$$

The equation residual is defined as:

$$\mathcal{R}(x,t) \equiv \frac{\partial u^*}{\partial t} - X(u^*) \tag{5}$$

and the semi-discrete approximation (2) is obtained by forcing it to be orthogonal to a set of weighting functions $w_m(x)$:

$$\int_D w_m(x) \cdot \mathcal{R} \, dx = 0 \qquad m = \ldots,-2,-1,0,1,2,\ldots\ldots \tag{6}$$

This is referred to as a weighted-residual condition, and is given Galerkin's name when the $w_m(x) \equiv \varphi_m(x)$. It also defines $u^*(x,t)$ as a weak solution of (1), with respect to the given weighting functions $w_m(x)$ In these methods, $u^*(x,t)$ is uniquely defined for all x . By contrast, finite difference methods are obtained by assuming a low-order polynomial approximation of $u(x,t)$, different around each point x_n (say $P_n(x,t)$), which is determined by simple collocation conditions. One then writes, point by point:

$$\frac{du_n}{dt} = X(P_n(x,t))_{x_n} \tag{7}$$

which is of the form (2) with $A_{1,\beta} = \delta(\beta)$ (the identity operator). Whenever A_1 is not diagonal, the semi-discretization (2) is called implicit, and explicit otherwise.

3. ACCURACY ANALYSIS

One may perceive in the theoretical work which surrounds the numerical approximation of partial differential equations two different viewpoints. The first, which we call analytic, is primarily concerned with the study of the order of accuracy of methods, or accuracy in the small, on the general premise that if errors are known to decrease as some power of the mesh size, then any accuracy may theoretically be obtained by mesh re-

finement. How fine the mesh must be chosen is not predicted in this approach. The second, which we shall call applied, is the reverse viewpoint and may be stated as follows: given a problem, what is the maximum mesh size which will ensure that an a-priori accuracy requirement will be met. Answering this question leads to analyzing accuracy in the large. Whereas classical analyses based on Taylor series expansions are restricted to analyses in the small, the "frequency method" described in the next section lends itself well to analyses in the large as well as in the small. Illustrating this shall be one of the objectives of this paper.

4. THE FREQUENCY METHOD

We describe in this section the principles of a method of accuracy analysis which shall be used in evaluating the relative merits of the competing methods of approximation (2). It is based on a systematic use of Fourier series, and is referred to as the frequency method. The mathematics of this method are not new (see e.g. Fix and Strang [3, 12] and Thomée [14]). But as just remarked, the frequency method contains more information than has been exploited in previous studies.

Consider the case where $X(\cdot)$ is a linear, constant coefficient operator. The approximation $X(u_n)$ is then also linear, and we may rewrite (2) as

$$A_1 \cdot \frac{du_n}{dt} = A_2 \cdot u_n \tag{8}$$

where $A_1 \cdot$ and $A_2 \cdot$ are the operators:

$$A_1 \equiv \sum_\beta A_{1,\beta} \cdot E^\beta$$
$$A_2 \equiv \sum_\beta A_{2,\beta} \cdot E^\beta \tag{9}$$

(E is the classical space shift operator defined by $E^\beta \cdot u_n = u_{n+\beta}$)

We assume that formulae such as (2) have been multiplied by an appropriate normalizing factor such that:

$$\sum_\beta A_{1,\beta} = 1 \tag{10}$$

The normalized operator A_1 is then an averaging operator, and the appropriate constant factor is in general $1/\Delta x$. To the periodic initial condition

$$u(x,0) = e^{i\omega x} \tag{11}$$

corresponds, for (1) assumed linear, the solution

$$u(x,t) = e^{i\omega x + \hat{X}(\omega)\cdot t} \tag{12}$$

where $\hat{X}(\omega)$ is the spectral function (sometimes called symbol) of the operator $X(\cdot)$, defined by:

$$\hat{X}(\omega) \equiv \frac{X(e^{i\omega x})}{e^{i\omega x}} \tag{13}$$

We note that (12) remains periodic for $t > 0$.

The semi-discrete approximation (2) has, for the same initial condition (11) the analogous solution:

$$u_n(t) = e^{i\omega x_n + \hat{A}(\omega)\cdot t} \tag{14}$$

where $\hat{A}(\omega)$ is the spectral function of the operator $A \equiv A_1^{-1} \cdot A_2$. That is:

$$\hat{A}(\omega) = \frac{\hat{A}_2(\omega)}{\hat{A}_1(\omega)} \; ; \; \hat{A}_\ell(\omega) = \sum_\beta A_{\ell,\beta} e^{i\beta\cdot\omega\cdot\Delta x} \quad (\ell = 1, 2) \tag{15}$$

The approximate solution (14) also remains periodic for $t > 0$, but differs from (12) in its time evolution: the difference between the exponents $\hat{A}(\omega)$ and $\hat{X}(\omega)$ which describe these time evolutions is a measure of the error introduced by the semi-discrete approximation (8).

We thus define a spectral truncation error function as:

$$\hat{\mathcal{E}}(\omega) = \frac{\hat{A}(\omega) - \hat{X}(\omega)}{\hat{X}(\omega)} = \frac{\hat{A}(\omega)}{\hat{X}(\omega)} - 1 \tag{20}$$

It will be seen that in the case of simple parabolic and hyperbolic equations, $\hat{\mathcal{E}}(\omega)$ has the interpretation of a relative error on the physical parameters of these equations.

Strictly speaking, it is implied that the domain in x is the whole real axis $(-\infty, +\infty)$. But the intent is of course to consider this periodic problem as a local model of equations of finite x-domain, much in the manner in which Von Neumann-type solutions are classically used to analyze numerical stability of difference methods. Non-linear problems are also included in the analysis, since one may consider linear equations as the model of locally linearized forms of non-linear equations.

5. PARABOLIC CASE

The simple heat equation:

$$\frac{\partial u}{\partial t} = \sigma \frac{\partial^2 u}{\partial x^2} \tag{22}$$

for which the spectral function $\hat{X}(\omega)$ is:

$$\hat{X}(\omega) = -\sigma \cdot \omega^2 \tag{23}$$

has the exact simusoidal solution

$$u(x,t) = e^{i\omega x - \sigma\omega^2 t} \tag{24}$$

By analogy, we express (14) as

$$u_n(t) = e^{i\omega x_n + \hat{A}(\omega)t} = e^{i(\omega x_n + Im(\hat{A}(\omega)t) - \sigma^*(\omega)\cdot\omega^2 t} \tag{25}$$

with

$$\sigma^*(\omega) = -\frac{Re(\hat{A}(\omega))}{\omega^2} \tag{26}$$

having the interpretation of a "numerical" diffusion constant, and $Im(\hat{A})\cdot t/\omega$ being a phase error of the approximation. It is easily shown that all symmetrical approximations in x lead to $Im(\hat{A}(\omega)) = 0$ (and we limit our analysis to these cases). Thus,

$$\sigma^*(\omega) = -\frac{\hat{A}(\omega)}{\omega^2} \tag{26-a}$$

The spectral error function (20) becomes:

$$\hat{\mathcal{E}}(\omega) = \frac{\sigma^*(\omega)}{\sigma} - 1 \tag{27}$$

and has the physical interpretation of a relative error on the diffusion constant of the equation introduced by its semi-discrete approximation(8].

Comparison of Several Schemes

Several schemes for the semi-discretization of the parabolic equation (22) are compared. These are:

(a) Standard 3 point and 5 point central difference schemes:

$$\frac{du_n}{dt} = \sigma\left(\frac{u_{n-1} - 2u_n + u_{n+1}}{\Delta x^2}\right) \tag{30}$$

and

$$\frac{du_n}{dt} = \sigma\left(\frac{-u_{n-2} + 16u_{n-1} - 30u_n + 16u_{n+1} - u_{n+2}}{12\cdot\Delta x^2}\right) \tag{31}$$

(b) Galerkin approximations with linear and quadratic B-Splines:

$$\frac{1}{6}\left(\frac{du_{n-1}}{dt} + 4\frac{du_n}{dt} + \frac{du_{n+1}}{dt}\right) = \sigma\left(\frac{u_{n-1} - 2u_n + u_{n+1}}{\Delta x^2}\right) \tag{32}$$

and

$$\frac{1}{120}\left(\frac{du_{n-2}}{dt} + 26\frac{du_{n-1}}{dt} + 66\frac{du_n}{dt} + 26\frac{du_{n+1}}{dt} + \frac{du_{n+2}}{dt}\right) = \sigma\left(\frac{u_{n-2} + 2u_{n-1} - 6u_n + 2u_{n+1} + u_{n+2}}{\Delta x^2}\right) \tag{33}$$

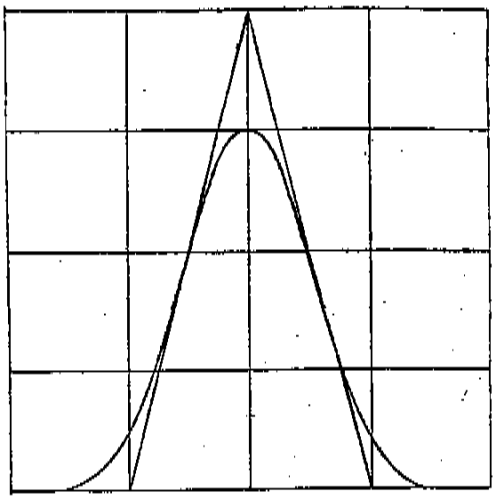

Figure 1: Linear and quadratic B-Splines.

(c) The Stormer-Numerov maximal order three point formula (see Birkhoff and Gulati, ref [ ])

$$\frac{1}{12}\left(\frac{du_{n-1}}{dt} + 10\frac{du_n}{dt} + \frac{du_{n+1}}{dt}\right) = \sigma\left(\frac{u_{n-1} - 2u_n + u_{n+1}}{\Delta x^2}\right) \tag{34}$$

(d) Quadratic and Hermite cubic finite elements (see ref[12,16]).

The corresponding expressions obtained for $\hat{\mathcal{E}}(\omega)$ are listed in the following table, and are illustrated in Figure 2.

METHOD	$\hat{\mathcal{E}}(\omega) = \frac{\sigma^*(\omega)}{\sigma} - 1$
3-Point finite differences	$\left(\frac{\sin(\omega\Delta x/2)}{\omega\Delta x/2}\right)^2 - 1$
Linear finite elements	$\left(\frac{\sin(\omega\Delta x/2)}{\omega\Delta x/2}\right)^2 \frac{3}{2+\cos(\omega\Delta x)} - 1$
Stormer-Numerov method	$\left(\frac{\sin(\omega\Delta x/2)}{\omega\Delta x/2}\right)^2 \frac{6}{5+\cos(\omega\Delta x)} - 1$
5-Point finite differences	$\frac{15 - 16\cos(\omega\Delta x) + \cos(2\cdot\omega\Delta x)}{6\cdot(\omega\Delta x)^2} - 1$
Quadratic B-Splines	$\frac{20(2 - 3\cos(\omega\Delta x) - \cos(2\cdot\omega\Delta x))}{(\omega\Delta x)^2\cdot(33 + 26\cdot\cos(\omega\Delta x) + \cos(2\cdot\omega\Delta x))} - 1$
Cubic Hermite finite elements	$\frac{B - \sqrt{B^2 - 4AC}}{2\cdot A} - 1$ $A = (312 + 106\cos(\omega\Delta x))\cdot(8 - 6\cos(\omega\Delta x)) - (26\sin(\omega\Delta x))^2$ $B = (-1008(8 - 6\cos(\omega\Delta x))\cdot(\cos(\omega\Delta x) - 1)$ $- 28\cdot(312 + 106\cdot\cos(\omega\Delta x))(\cos(\omega\Delta x) - 4)$ $+ 4563\sin^2(\omega\Delta x))/(\omega\Delta x)^2$ $C = 28224(\cos(\omega\Delta x) - 1)(\cos(\omega\Delta x) - 4)/(\omega\Delta x)^4$ $- (84\sin(\omega\Delta x)/(\omega\Delta x)^2)^2$
Quadratic finite elements	$\frac{4(15 + 2\cdot\cos(\omega\Delta x) - \sqrt{D})}{(\omega\Delta x)^2\cdot(5 - \cos(\omega\Delta x))} - 1$ $D = (15 + 2\cdot\cos(\omega\Delta x))^2 + 16(\cos(\omega\Delta x) - 1)(5 - \cos(\omega\Delta x))$

TABLE 1: Spectral error function for semi-discretisations of the equation $\frac{\partial u}{\partial t} = \sigma\frac{\partial^2 u}{\partial x^2}$

13. "BEST" THREE-POINTS DISCRETIZATION

Maximizing $(\omega\cdot\Delta x)_{.01}$, as in section 10, produces in this case the three point semi-discretization:

$$\frac{1}{5.55}\left(\frac{du_{n-1}}{dt}+3.53\frac{du_n}{dt}+\frac{du_{n+1}}{dt}\right) = -c\left(\frac{-u_{n-1}+u_{n+1}}{2\cdot\Delta x}\right) \qquad (62)$$

for which

$$(\omega\cdot\Delta x)_{.01} \simeq 1.66$$

which, again, gives an accuracy in the large which is about 50 percent larger than for the 3-point linear finite element discretization(59) which is of maximal order in this case. The spectral error functions of those two 3-point algorithms are shown for comparison in fig. 5.

14. A NUMERICAL ILLUSTRATION

A concrete comparison of the three semi-discretizations (57), (59) and (62) has been obtained by computing numerically step responses using these schemes. The results illustrated in fig 6 clearly show the improved quality of these numerical results as $(\omega\cdot\Delta x)_{.01}$ increases (From $(\omega\cdot\Delta x)_{.01}=.25$ in the 3-point finite difference case to 1.66 in the "best" case.)

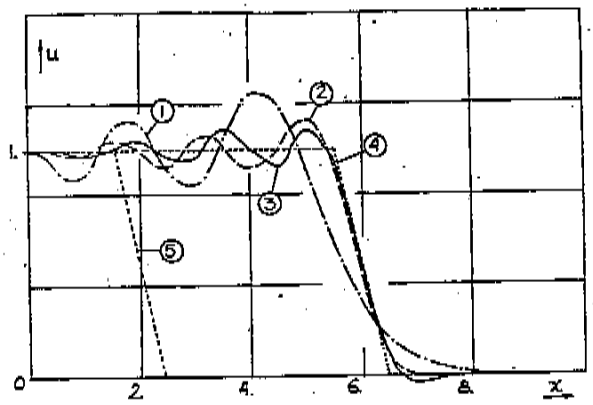

Figure 6: Solution of the equation $\partial u/\partial t + c\,\partial u/\partial x = 0$ *by different 3-Point semi-discretizations.*

1. *Finite differences.*
2. *Linear finite elements.*
3. *"Best" approximation..*
4. *Exact solution.*
5. *Initial condition.*

15. ACKNOWLEDGEMENTS :

Part of this work has been supported by the National Science Foundation of Belgium and by the U.S. Dept. of Interior, Office of Water Resources Research, Grant No. OWRR A-045 N.J.

16. REFERENCES

[1] BIRKHOFF, G. and GULATI, S. (1974) "Optimal few-point discretizations of linear source problems" SIAM J. Numer. Anal., September.

[2] COLLATZ, L. (1960) "The numerical treatment of differential equations" Springer-Verlag.

[3] FIX, G. and STRANG, G. "Fourier analysis of the finite element method in Ritz-Galerkin theory" Studies in Appl. Math. Vol. 48, pp. 265-273.

[4] KREISS, H. and OLIGER, J. (1972) "Comparison of accurate methods for the integration of hyperbolic equations" Tellus, 24 (1972).

[5] LAX, P.D. (1957) "Hyperbolic systems of conservation laws II" Comm. Pure and Appl. Math., Vol. 10, pp. 537-566.

[6] LAX, P.D. and WENDROFF, B. (1960) "Systems of conservations laws" Comm. Pure and Appl. Math., Vol. XIII, pp. 217-237.

[7] LAX, P.D. and WENDROFF, B. (1964) "Difference schemes for hyperbolic equations with high order of accuracy" Comm. Pure and Appl. Math., Vol. 17, p. 381.

[8] RICHTMYER, R. D. and MORTON, K. W. Difference Methods for Initial Value Problems Interscience Publishers, New York, 1967.

[9] SCHOENBERG, I. J. (1946) "Contributions to the problem of approximation of equidistant data by analytic functions, Parts A and B" Quart. Appl. Math., 4(1946) pp. 45-99, and pp. 112-141.

[10] SCHOENBERG, I. J. (1973) "Cardinal spline interpolation" Regional Conference Series in Applied Mathematics SIAM, Vol. 12, 1973.

[11] SCHULTZ, M. H. (1973) Spline Analysis, Prentice Hall, Inc., Englewood Cliffs, N. J.

[12] STRANG, G. and FIX, G.J. (1973) An Analysis of the Finite Element Method, Prentice-Hall, Inc., Englewood Cliffs, N. J.

[13] SWARTZ, B. and WENDROFF, B. (1974) "The relative efficiency of finite difference and finite element methods. I: hyperbolic problems and splines" SIAM J. Num. Anal., October.

[14] THOMÉE, V. and WENDROFF, B. (1974) "Convergence estimates for Galerkin methods for variable coefficient initial value problems" SIAM J. Num. Anal., Vol. 11, No. 5, October.

[15] VICHNEVETSKY, R. (1973) "Physical criteria in computer methods for partial differential equations" Proceedings, 7th AICA International Congress, Prague 1973. Reprinted in Proc. of AICA, Vol. XVI, No. 1, Jan. 1974, Brussels.

[16] VICHNEVETSKY, R. (1975) "Introduction to Finite Element Methods for Initial Value Problems" Course Notes, Rutgers University, Dept. of Computer Science, New Brunswick, N. J.

[17] VICHNEVETSKY, R., TU, K. W. and STEEN, J. A. (1974) "Quantitative error analysis of numerical methods for partial differential equations" Proceedings, Eighth Annual Princeton Conference on Information Science and Systems. Princeton University, March 1974.

[18] WENDROFF, B. (1960) "On central difference equations for hyperbolic systems" J. Soc. Indust. Appl. Math., Vol. 8, p. 549.

$$\frac{1}{120}\left(\frac{du_{n-2}}{dt}+26\frac{du_{n-1}}{dt}+66\frac{du_n}{dt}+26\frac{du_{n+1}}{dt}+\frac{du_{n+2}}{dt}\right)$$
$$=-c\left(\frac{-u_{n-2}-10u_{n-1}+10u_{n+1}+u_{n+2}}{24\cdot\Delta x}\right) \qquad (60)$$

(c) Quadratic and Hermite Cubic finite-elements (see ref[12,14]).
The corresponding expressions obtained for $\hat{\mathcal{E}}(\omega)$ are listed in the following table, and are illustrated in Figure 4.

METHOD	$\hat{\mathcal{E}}(\omega)=\frac{c^*(\omega}{c}-1$
3-Point finite differences	$\frac{\sin(\omega\cdot\Delta x)}{\omega\cdot\Delta x}-1$
Linear finite elements	$\frac{3\cdot\sin(\omega\cdot\Delta x)}{(\omega\cdot\Delta x)(2+\cos(\omega\cdot\Delta x))}-1$
Box method	$\frac{\tan\left(\frac{\omega\cdot\Delta x}{2}\right)}{\left(\frac{\omega\cdot\Delta x}{2}\right)}-1$
5-Point finite differences	$\frac{4}{3}\frac{\sin(\omega\cdot\Delta x)}{\omega\cdot\Delta x}-\frac{1}{3}\frac{\sin(2\omega\cdot\Delta x)}{\omega\cdot\Delta x}-1$
Quadratic B-Splines	$\frac{3(10\cdot\sin(\omega\cdot\Delta x)+\sin(2\cdot\omega\cdot\Delta x))}{\omega\cdot\Delta x(33+26\cos(\omega\cdot\Delta x)+\cos(2\cdot\omega\cdot\Delta x))}-1$
Cubic Hermite finite elements	$\frac{105(B+\sqrt{C})}{435+138\cdot\cos(\omega\cdot\Delta x)+142\cos^2(\omega\cdot\Delta x)}-1$ $B=\left(29\cdot\frac{\sin(2\cdot\omega\cdot\Delta x)}{2\cdot\omega\cdot\Delta x}-64\cdot\frac{\sin(\omega\cdot\Delta x)}{\omega\cdot\Delta x}\right)/10$ $C=B^2+28\left(435+\cos(\omega\cdot\Delta x)+142\cos^2(\omega\cdot\Delta x)\right)$ $\cdot\left(11-12\cdot\cos(\omega\cdot\Delta x)\cdot\cos^2(\omega\cdot\Delta x)\right)/\left(1050\cdot(\omega\cdot\Delta x)^6\right)$
Quadratic finite elements	$\frac{\sqrt{D}-1}{3-\cos(\omega\cdot\Delta x)}\frac{\sin(\omega\cdot\Delta x)}{\omega\cdot\Delta x}-1$ $D=\left(4\frac{\sin(\omega\cdot\Delta x)}{\omega\cdot\Delta x}\right)^2+20\left(\frac{1-\cos(\omega\cdot\Delta x)}{(\omega\cdot\Delta x)^2}\right)\left(3-\cos(\omega\cdot\Delta x)\right)$

TABLE 3: Spectral error function for semi-discretizations of the equation $\frac{\partial u}{\partial t}+c\frac{\partial u}{\partial x}=0$

12. COMPARISON IN TERMS OF ACCURACY IN THE LARGE

As in the parabolic case, one may characterize accuracy in the large for the semi-discretizations of table 3 by the numbers $(\omega\cdot\Delta x)_{.01}$
The resulting comparison and ordering of the competing schemes analysed is shown in table 4

Method	*Equation Number in Text*	$(\omega\cdot\Delta x)_{.01}$
3-Point Finite Differences	(57)	.25
5-Point Finite Differences	(58)	.75
Linear Finite Elements	(59)	1.12
Quadratic B-Splines	(60)	1.7
Hermite Cubic Finite Elements		1.8
Quadratic Finite Elements		1.98

TABLE 4: Ordering of semi-discretizations of the equation $\partial u/\partial t+c\,\partial u/\partial x=0$ *in order of increasing "accuracy in the large"*

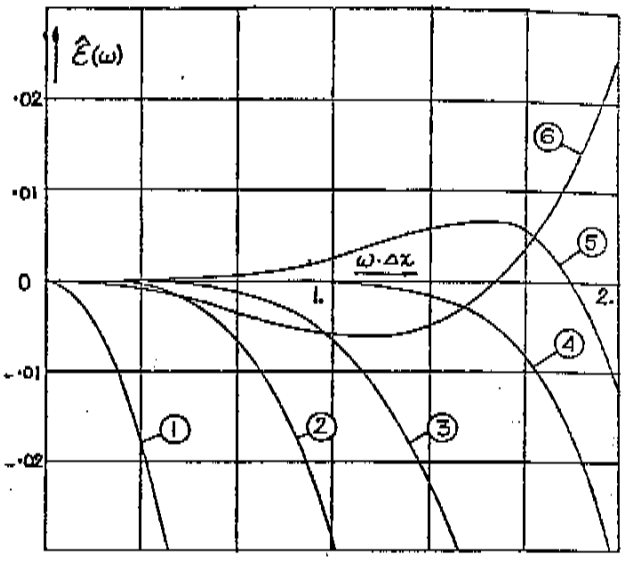

Figure 4: Spectral error-functions of different semi-discretisations for the hyperbolic equation $\frac{\partial u}{\partial t}+c\frac{\partial u}{\partial x}=$

1: *3-point finite differences.*
2: *5-point finite differences.*
3: *Linear finite elements.*
4: *Quadratic B-Splines.*
5: *Quadratic finite elements.*
6: *Cubic Hermite finite elements.*

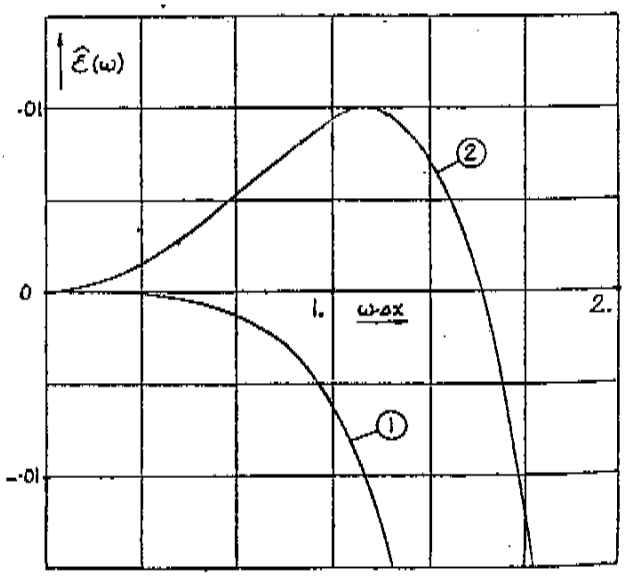

Figure 5: Spectral error functions of the linear finite element (1) and the "best" approximation (2) for the hyperbolic equation $\partial u/\partial t+c\,\partial u/\partial x=0$

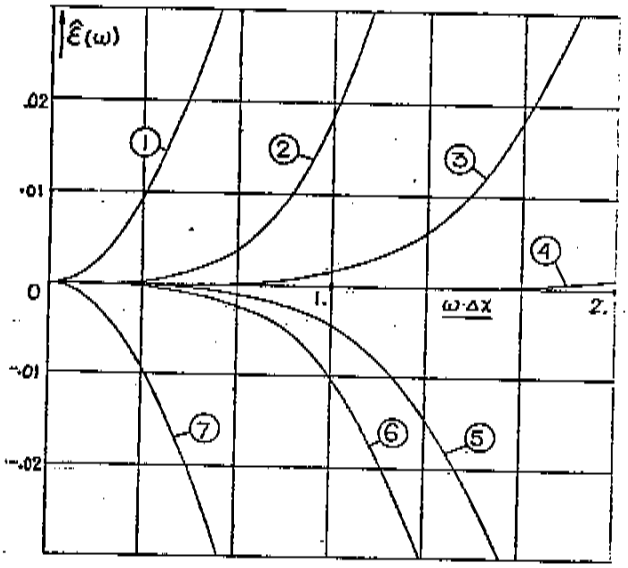

Figure 2: Spectral error functions of different semi-discretizations for the parabolic equation $\frac{\partial u}{\partial t} = \sigma \frac{\partial^2 u}{\partial x^2}$

1: Linear finite elements.

2: Quadratic finite elements.

3: Quadratic B-Splines.

4: Cubic Hermite finite elements.

5: Stormer-Numerov approximation.

6: 5-point finite differences.

7: 3-point finite differences.

6. RELATION TO THE APPROXIMATION OF AN ELLIPTIC PROBLEM

The preceding analysis applies with minor modification to the elliptic problem

$$f(x) = \frac{d^2u}{dx^2} \tag{40}$$

approximated with the same finite difference and finite element techniques by

$$A_1 \cdot f_n = A_2 \cdot u_n \tag{41}$$

where A_1 and A_2 have the same definition as before. (See e.g. Birkhoff and Gulati, ref. [1] for a recent discussion of this problem.)

The parallel with the initial value parabolic problem discussed here is that (41) is used from left to right (given f, find u) whereas (8) is used from right to left (given u, find $\partial u/\partial t$). With f taken sinusoidal as $e^{i\omega x}$, the exact solution of (40) is

$$u(x) = \frac{e^{i\omega x}}{\hat{X}(\omega)} \tag{42}$$

with

$$\hat{X}(\omega) = -\omega^2 \tag{43}$$

By contrast, the approximate solution given by (41) for the same $f(x)$ is

$$u_n = \frac{e^{i\omega x_n}}{\hat{A}(\omega)} \tag{44}$$

With a minor modification of definition, the spectral truncation error is the same as (20).

7. A PROPERTY OF THE ERROR FUNCTION NEAR $\omega\cdot\Delta x = 0$

An important property of the error function $\hat{\mathcal{E}}(\omega)$ which is satisfied by all the algorithms considered here is as follows: *If the semi-discretization (8) is of order of accuracy** p, then its spectral error function (20) and its (p-1) first derivatives with respect to $(\omega\cdot\Delta x)$ are zero at the origin. Example: The three point central-difference approximation (30) has an order of accuracy 2. The corresponding spectral error function

$$\hat{\mathcal{E}}(\omega) = \frac{2\,(\cos(\omega\cdot\Delta x)-1)}{-(\omega\cdot\Delta x)^2} - 1$$

$$= -\frac{2\cdot(\omega\cdot\Delta x)^2}{4!} + \cdots\cdots \tag{45}$$

verifies:

$$\hat{\mathcal{E}}(0) = 0$$

$$\frac{d\hat{\mathcal{E}}(0)}{d(\omega\cdot\Delta x)} = 0 \tag{46}$$

8. ACCURACY IN THE LARGE

A zero-error algorithm would have the zero-line as its spectral error function $\hat{\mathcal{E}}(\omega)$ at least up to the folding frequency $\omega\cdot\Delta x = \pi$). The accuracy of non-perfect algorithms may thus be compared in a global sense by comparing how well their respective spectral error functions approximate that line. We know, as a corollary of the property stated in the preceding section, that an "optimal order" algorithm is one that approximates $\hat{\mathcal{E}}(\omega) = 0$ as well as possible near $\omega\cdot\Delta x = 0$, or in the "maximally flat" sense (borrowing this term from filter theory).

But solutions of actual problems are superpositions of sinusoidal components of different frequencies which all contribute to the global error.

In order to have a measure which describes the quality of the approximation not only near $(\omega\cdot\Delta x)=0$ but beyond, we define, somewhat loosely and with due apology, by γ-accuracy of the semi-discretization (8) the length of the segment of the $(\omega\cdot\Delta x)$ axis between zero (where $\hat{\mathcal{E}}(\omega)$ is always zero) and the point $(\omega\cdot\Delta x)_\gamma$ beyond which $|\hat{\mathcal{E}}(\omega)|$ becomes larger than the positive number γ.

For all practical purposes, we shall use $\gamma = .01$ (or one percent) in the remainder. Taking the one-percent point is somewhat arbitrary and other values would produce different results. It is however believed that the one-percent value is a reasonable choice for practical applications, and that other choices not drastically different would result in a similar ordering of the algorithms in terms of their relative merits.

In the following table, the algorithms of Table I are listed in order of increasing .01-accuracy.

---

* See ref. [16] for a detailed proof.

** The order of accuracy is defined as the order in Δx of the first term which does not cancel out in Taylor series developemnts of left and right hand sides of (8) in powers of Δx when an exact solution is substituted for u (with A_1 normalized)

Method	Equation Number In Text	$(\omega\cdot\Delta x)_{.01}$
3-Point Finite Differences	(30)	.35
Linear Finite Elements (B-Spline)	(32)	.35
Quadratic Finite Elements		.85
5-Point Finite Differences	(31)	1.00
Stormer-Numerov	(34)	1.23
Quadratic B-Splines	(33)	1.48
Hermite Cubic Finite Elements		3.15

TABLE 2: *Ordering of semi-discretisations of the equation* $\partial u/\partial t = \sigma\,\partial^2 u/\partial x^2$ *in order of increasing "accuracy in the large"*

9. DISCUSSION

A comparison of the several three point discretizations is interesting: Somewhat unexpectedly, the linear finite-element algorithm (32) has about the same .01-accuracy as the simple 3 point central difference formula (30). The Stormer-Numerov algorithm (36) which has maximum order of accuracy is, as expected, reasonably better than either of the preceding two.

10. A "BEST" THREE-POINT DISCRETIZATION

The three-point semi-discretization

$$\frac{1}{10.36}\left[\frac{du_{n-1}}{dt} + 8.36\frac{du_n}{dt} + \frac{du_{n+1}}{dt}\right] = \sigma\left[\frac{u_{n-1} - 2u_n + u_{n+1}}{\Delta x^2}\right] \tag{47}$$

obtained entirely empirically by seeking the coefficients of A_1 which maximize $(\omega\cdot\Delta x)_{.01}$ gives:

$$(\omega\cdot\Delta x)_{.01} = 1.84 \tag{48}$$

which is about 50 percent more than the .01-accuracy of the Stormer-Numerov method. While the latter approximates the line $\hat{\mathcal{E}}(\omega) = 0$ optimally in the Mac Laurin or "maximally flat" sense, the "best" method (47) approximates this line optimally in the Tchebychev sense, and results in a better distribution of errors. The two $\hat{\mathcal{E}}(\omega)$ curves are compared below:

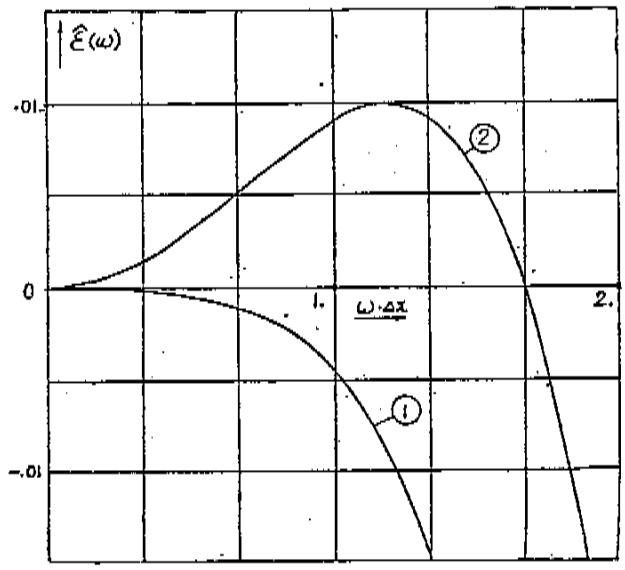

Figure 3: Spectral error functions of the Stormer-Numerov approximation (1) and the "best" approximation (2) for the parabolic equation $\partial u/\partial t = \sigma\,\partial^2 u/\partial x^2$.

11. HYPERBOLIC CASE

We consider the simple hyperbolic equation

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0 \tag{50}$$

for which

$$\hat{X}(\omega) = -ic\omega \tag{51}$$

It has the exact sinusoidal solution

$$u(x,t) = e^{i\omega(x-ct)} \tag{52}$$

The parameter c has the physical interpretation of a velocity (the solution is displaced without distortion at the velocity c in the x direction).
Solutions of the semi-discrete approximation of (8) may be written, separating real and imaginary parts of $\hat{A}(\omega)$, as:

$$u_n(t) = e^{Re(\hat{A}(\omega))\cdot t}\cdot e^{i\omega(x_n - c^*(\omega)t)} \tag{53}$$

where

$$c^*(\omega) = -\frac{Im(\hat{A}(\omega))}{\omega} \tag{54}$$

(53) is the expression of sinusoidal solution with phase velocity $c^*(\omega)$ and with (non-conservative) amplitude $\exp(Re(\hat{A}(\omega))t)$. For all symmetric approximations, however, $Re(\hat{A}(\omega)) = 0$, and

$$u_n(t) = e^{i\omega(x_n - c^*(\omega)t)} \tag{55}$$

Then,

$$\hat{\mathcal{E}}(\omega) = \frac{\hat{A}(\omega)}{\hat{X}(\omega)} - 1 = \frac{c^*(\omega)}{c} - 1 \tag{56}$$

becomes simply the relative error on the phase velocity introduced by the semi-discretization. (Again, we limit ourselves to these symmetric cases here). Within a ratio of 2π, the definition (56) of the error is identical to that used for hyperbolic equation by Kreiss and Oliger [4] and by Swartz and Wendroff [13]. The 2π factor stems from the fact that we use velocity error, rather than the phase-error-per-period which they use.

Comparison of Several Schemes

We compare several semi-discretizations of the hyperbolic equation (50). These are

(a) Standard 3 point and 5 point central difference approximations

$$\frac{du_n}{dt} = -c\left(\frac{-u_{n-1} + u_{n+1}}{2\cdot\Delta x}\right) \tag{57}$$

and

$$\frac{du_n}{dt} = -c\left(\frac{u_{n-2} - 8u_{n-1} + 8u_{n+1} - u_{n+2}}{12\cdot\Delta x}\right) \tag{58}$$

(b) Galerkin approximation with linear and quadratic B-splines

$$\frac{1}{6}\left(\frac{du_{n-1}}{dt} + 4\frac{du_n}{dt} + \frac{du_{n+1}}{dt}\right) = -c\left(\frac{-u_{n-1} + u_{n+1}}{2\cdot\Delta x}\right) \tag{59}$$