BIT 25 (1985), 188–204

INVITED PAPER.

33 YEARS OF NUMERICAL INSTABILITY, PART I

GERMUND DAHLQUIST

*Department of Numerical Analysis and Computer Science, Royal Institute of Technology,
S-10044 Stockholm, Sweden*

Abstract.

BIT has played and plays a great role in the development of concepts concerning numerical (in)stability in initial value problems for *ODE*'s and related questions. This development is here seen through the looking-glass of the author, who experienced much of its pains and pleasures. The article is based on a talk given in 1981 at the Zürich symposium to commemorate the tenth anniversary of the death of the eminent Swiss numerical analyst, Heinz Rutishauser. The presentation is mainly chronological with a few digressions. Part I ends at the beginning of the stiff epoch.

Keywords: Numerical instability, history, differential equations, difference equations.

1. Freiburg im Breisgau, April 1951.

One day in the beginning of 1951 my employer, the Swedish Board for Computing Machinery capitulated to my requests to go to the GAMM meeting at Freiburg im Breisgau. So I went there with a ten-minute talk [8] in my baggage. I had just found the way to my room in a small Freiburg hotel, when the telephone rang, and I was told that a gentleman wanted so see me. Very strange! Fifteen minutes earlier I did not know the address myself, and I knew no other participants at this conference.

A serious man, a few years older than me, waited for me at the reception. He introduced himself as Heinz Rutishauser. He had read my abstract and feared that we had made the same discovery. I rehearsed my talk for him, which contained an example of some analysis I had done in order to choose a numerical method for some missile calculations on a relay computer in Stockholm, because I had had (justified) fears that the task might take a considerable time on that computer.

I reproduce below the essentials of my talk.

"This particular analysis is concerned with the application of the leapfrog method,

$$y_{n+1} = y_{n-1} + 2hf(y_n, t_n) \tag{1.1}$$

Received December 1984.

for the system,

$$(1.2)\qquad dy/dt = f(y,t), \qquad y(0) = c.$$

The local truncation error per unit of time is,

$$(1.3)\qquad p(t) \approx h^2 \dddot{y}/6.$$

I show, by a somewhat heuristic (though not very sloppy) argument, loc. cit., that the error may be decomposed according to the formula,

$$(1.4)\qquad y(t_n) - y_n \approx u(t_n) + av(t_n) + b(-1)^n w(t_n),$$

where u, v, w are solutions of the differential equations,

$$(1.5a)\qquad du/dt = J(t)u + p(t), \qquad u(0) = 0,$$

$$(1.5b)\qquad dv/dt = J(t)v,$$

$$(1.5c)\qquad dw/dt = -J(t)w,$$

where $J(t) = \partial f/\partial y$ is the Jacobian evaluated at $(t, y(t))$, and a and b are determined by the two initial conditions needed for the difference equation (1.1). For example, if $f(t,y) = -y$, we obtain for $t = t_n$ (with error-free initial data),

$$\exp(t) - y_n \approx -(h^2/6)y_0(t\exp(-t) + (-1)^n(h/2)\exp(t)).$$

It is pointed out that the oscillating term grows exponentially and might become the largest term: for $h = 0.1$ this happens for $t > 2$. It is also emphasized that the oscillatory component is due to the fact that the difference equation is of higher order than the differential equation.

For other (two-step) methods studied nothing is changed in (1.5a), (1.5b), except for the expression for the local truncation error, but (1.5c) reads,

$$dw/dt = cJ(t)w,$$

where c is a constant characteristic for the method (later called a growth parameter). For example, if a predicted value for y_{n+1} is iteratively improved until convergence by the use of (the fourth order accurate) Simpson's formula,

$$y_{n+1} - y_{n+1} = (h/3)(f(y_{n+1}) + 4f(y_n) + f(y_{n-1})), \quad (i = 1, 2, 3, \ldots)$$

(which is called Milne-Simpson's method below) then $c = \ldots$."

At this point, I had to turn the page, but before I had done so, Rutishauser filled in: $c = -1/3$. No doubt, we had made the same "discovery"! He told he had submitted a paper to a journal but had troubles with the refereeing process. Unfortunately this was 9 years before the birth of BIT. Carl-Erik Fröberg had only just conceived it.

Eventually Rutishauser submitted his manuscript to a new Swiss journal instead, where it appeared in 1952, [33]. His paper differs from mine in many respects. He obtains results for several methods and points out that when $\|h\partial f/\partial y\|$ is small enough, no growing oscillations will occur with the Runge-Kutta methods (which are one-step methods) and the Adams methods. He does not derive equations like (1.5c) but considers "frozen Jacobians", i.e. for every time τ, he considers the behavior for $t > \tau$ of the numerical scheme on a linear system with a constant matrix, equal to the Jacobian evaluated at the point $(\tau, y(\tau))$.

In the last sentence of my paper, I expressed my intention to publish the derivations and results more completely. The Zentralblatt reviewer remarked to this that he desired I would then mention the possibility of a rigorous error analysis along these lines. It took me some time to live up to that expectation (1958). I wanted to discuss these matters in the framework of a general theory for linear multistep methods. It was natural to treat some other aspects of that theory first ("zero-stability", see below), and in the general formulation it was not easy to obtain full rigor. The delay was also caused by my involvement in many other tasks at the Swedish Board for Computing Machinery and the International Meteorological Institute in Stockholm.

The next to last sentence of my paper surprises me, when I read it again. "When Simpson's rule is used only once in each step, c depends on the predictor used". I never returned to this aspect, but Stetter [34] found a remarkable combination, where one can avoid the growing oscillations, at least for linear autonomous systems with real and negative eigenvalues.

2. The state of the art around 1951.

The phenomenon of increasing oscillations, which have nothing to do with the exact solution of the *ODE* system, later became known as weak (numerical) instability [12] or *weak stability* [19] or (to-day) weak zero-stability. Rutishauser used the word "instability", while I did not, in my 1951 paper.

I have never seen the term "numerical stability" in the pre-computer literature on numerical methods. People who did numerical work before the computer age often had a considerable craftmanship. I remember more than one of them uttering words like: "You will notice a lot of things when you put digits into the scheme that you did not expect, when you read the derivation of it". Phenomena of the kind discussed above were probably known to many of them, but they rarely wrote about them.

One exception is a paper in 1942 by Collatz and Zurmühl [4], see also Collatz [3], p. 86. Like Rutishauser and me, they consider two-step methods and mention "Aufrauhungserscheinung und Glättung", i.e. the results may become rough. This is first noticed in the third differences. There is not so much analysis of the phenomenon, but they suggest a smoothing procedure. When to apply it, and also partly how to apply it, is left to the judgement of the human computer.

An algorithmic version of related ideas was published in 1959 by Milne and Reynolds [29]. It seems likely that at that time similar devices existed in other programs. A new craftmanship had been created. Much wisdom was, and still is, buried in computer codes, in particular if they are written in low level languages.

Nevertheless, in the computer age there is undoubtedly much more open scientific discussion of aspects of computing, which were earlier, on a much smaller scale, considered as craftmanship. With the automatic computer one began to look at computing as a "process", a word used by Turing 1948 in an important paper on matrix computations [38]. Words like "noise" for rounding errors, numerical (in)stability and condition numbers belong to this new point of view, even though Turing did not use the word "stability". It is not so much the increase of the number of arithmetic operations that matters, but the disappearance of the human inspection of almost every arithmetic result, based on criteria, which were successively developed during the course of the work and partly forgotten when the work was finished.

Who invented the term "numerical instability" and when? I conjecture that it was first heard about in 1946, in the groups around von Neumann or Turing. I invite the reader to give counter-examples to this and other statements of this paper. I have tried but not succeeded entirely to get my beliefs confirmed. Wilkinson has confirmed that he used it several times in conversations with Turing about that time. Turing had remarked that Wilkinson seemed to mean different things every time he used that word. The Turing-Wilkinson conversation illuminates two things. First, there are indeed plenty of distinctions to be made. The most fundamental is the distinction between instability in the underlying mathematical problem and instability in an algorithm for the (exact or approximate) treatment of the problem. Second, we need both a more imprecise usage of the word and well-defined concepts for gaining more insight, in the form of theorems or otherwise. To-day, in connection with the numerical treatment of *ODE*'s we use prefixes to stability from a large subset of the alphabet, from A to Å, to cover different situations. (In the Scandinavian alphabet Å comes near the end.)

To my knowledge von Neumann first wrote about numerical stability in 1947 in his great work with Goldstine on rounding errors in matrix inversion [40]. The authors seem to hesitate about the use of the word "stability". For example, in Chapter 1, on p. 1027 they write "... the continuity of the result as a function of the parameters of the problem, or somewhat more loosely worded, of the mathematical stability of the problem". Then the expression "continuity or stability" is used a few times until on p. 1028 they use "stability" without a companion a couple of times. In the following chapters the word is not used.

Von Neumann's ideas on these things are exposed more explicitly in two papers published by him in 1950 with different coauthors, one on numerical weather prediction [1] and the other on hydrodynamical shocks [41]. There are also early papers by other authors, which are in part based on his ideas and Fourier technique, e.g. Crank and Nicolson 1947 [6], Eddy 1949 [15] and O'Brien, Hyman and Kaplan 1951 [30]. All these papers are concerned with finite difference methods for partial differential equations (*PDE*'s).

However, almost 20 years before all this, Courant, Friedrichs and Lewy published a now classical paper [5], where they study a few cases, when a partial difference equation (*PΔE*) converges formally to a *PDE*. In modern terminology the *PΔE* is consistent with a *PDE*. Among other things they studied a Cauchy problem for the *PΔE*, which is obtained, when the derivatives in the second order hyperbolic equation

$$\partial^2 u/\partial t^2 = c^2 \partial^2 u/\partial x^2 \tag{2.1}$$

are replaced by central difference quotients. They pointed out that in order that the solutions of the *PΔE* problem should converge to the solutions of the corresponding *PDE* problem, then one must choose $\Delta t \leqq \Delta x/c$. This condition is now often called a stability condition, but Courant et al. did not take this point of view. They studied convergence. In the hyperbolic case, the necessity of the condition was derived by the comparison of the domains of dependence for the *PΔE* and the *PDE*. If the condition is violated then they found that the solution of the *PΔE* has no chance to be influenced by an interval of data, which the solution of the *PDE* is known to depend on. Hence there cannot be convergence for Cauchy problems with arbitrary initial data.

As far as I know, before the computer age they did not write about the question "how is it possible that a consistent scheme for a well posed problem does not converge?". I may be wrong. It can also be another example showing that much wisdom does not find its way to the printer.

In 1936, however, Collatz [2] discussed some good and bad examples of finite difference approximation from the point of view of error propagation (in the l_∞-norm). In 1947, von Neumann and Goldstine, loc. cit. p. 1028, gave the fóllowing interpretation: "That the stability of the strict problem need not imply that of an arbitrarily close approximant was made particularly clear by some important results of R. Courant, K. Friedrichs and H. Lewy". Anyway, an answer to the question above was given in *the equivalence theorem of Lax*, presented in 1953 at a seminar at New York University (where Courant was the director of the Institute of Mathematical Sciences), [31], p. 39. I quote from [31], p. 44.

"Given a properly posed initial-value problem and a finite difference approximation to it that satisfies the consistency condition, stability is the necessary and sufficient condition for convergence". Here "stability" is defined as the uniform boundedness of an infinite set of difference operators

$$(C(\Delta t))^n, \qquad 0 < \Delta t < \delta, \qquad 0 \leqq n\Delta t < T. \tag{2.2}$$

Implicitly, the uniformity is also required with respect to the spatial increments Δx_i, $i = 1, 2, \ldots, d$, where d is the number of space dimensions, for relations $\Delta x_i = g_i(\Delta t)$ are assumed.

The result and the proof of Courant, Friedrichs and Lewy fascinated me. I became curious to see what happens, if the initial values are analytic. In this case the initial data are determined for all x by the values on an arbitrarily short interval, and the domain of dependence argument is not applicable. I found a much more liberal convergence condition, see [10], than Courant et al., but this is not of practical interest. I quote from [10], p. 100. "Since most functions of Applied Mathematics are at least piece-wise analytic, one might expect that the theorems just obtained would be more relevant to numerical practice than the negative results of Courant et al. This is, however, not the case (if the difference equation is used for *recursive* numerical computation...). The reason is the presence of round-off errors, which behave very much like non-analyticities. In terms of Fourier analysis, the round-off introduces 'wave components' ... which have a very rapid growth. Thus that circumstance which caused trouble for the proof of convergence under more general assumptions, gives rise to so-called numerical instability in a computation with finite differences." (I think I meant finite word length). Then I illustrated the error growth by a perturbation scheme, which I had learned from Collatz, and pointed out that "from the point of view of discrete Fourier analysis, as used by ... von Neumann, O'Brien et al., Hyman and others ... the convergence ... when $f(x)$ is an analytic non-periodic function ... appears as the result of a cancellation between different diverging components". Although [10] is outside the mainstream, both in the study of *PΔE*'s for numerical use and in my career, I can still understand why it was fun to make this odd study, and I certainly learned a lot from it.

Let me make a digression here. In [5], the lack of convergence (and hence the instability) was explained by a "physical" argument concerning domains of dependence for a *PDE* and a consistent *PΔE*. There are similar applications of physical ideas in the more recent research on numerical instability, see Trefethen [37] and the literature quoted by him. Trefethen shows how the concept of group velocity can be applied to numerical phenomena in the solution of *PΔE*'s. There may be refraction effects at lines, where the grid size is changed, because in *PΔE*-approximations to (2.1) the wave speed depends on the grid size and on the wave number. There may be reflection effects at artificial boundaries, which sometimes have to be introduced because the computer cannot handle an infinite domain. Furthermore, instability may arise, when a *PΔE* requires more boundary conditions than the *PDE*. Methods with this feature may have other advantages (high order of consistency), but they cannot be accepted unless the "unphysical" boundary conditions are handled in a stable way, see again [37]. The situation is not exactly the same in the example of Section 1, where the leap-frog method demands extra initial conditions instead of boundary conditions, but there are connections.

The earliest development of numerical stability theory was connected with *PDE*'s, while I shall here mainly discuss *ODE*'s. There are, however, important relations between the problems. *ODE*'s may, of course, be considered as a simple case of *PDE*'s, where one does not have the extra difficulty of approximating unbounded differential operators. On the other hand, many numerical schemes for *PDE*'s can be interpreted or derived by a two stage process, sometimes called the *method of lines*. In the first stage, the *PDE* is replaced by an approximately equivalent huge system of *ODE*'s, derived by a discretization in the space variables only, by means of finite differences or some Galerkin method. The dimension of the system as well as its Lipschitz constant grows without bounds, as the accuracy of the spatial discretization increases. In the second stage a general numerical method for *ODE*'s is applied to this system.

Ideas from the numerical solution of *ODE*'s are therefore important for *PDE*'s. Conversely, ideas from *PDE*'s and their physical background may be useful for the design and analysis of numerical methods for *ODE*'s, since they indicate difficulties which *perhaps* a general purpose program for *ODE*'s should be able to resolve. There is, however, an important special difficulty with the method of lines, which is to be discussed at the end of this section.

Crank and Nicolson 1947 [6] seem to have been the first to point out that it is advantageous to use an *implicit method* in order to obtain good stability properties. They consider heat conduction problems and suggest the trapezoidal method for the time integration, and compare it with the leap-frog method. They study the error propagation first by a perturbation scheme and then by a discrete Fourier analysis scheme, suggested to them by von Neumann (via Hartree). While leap-frog runs wild already for arbitrarily small values of $\Delta t/\Delta x$, no restriction on this ratio is needed for the trapezoidal method.

Laasonen [24] finds that the use of an implicit method is advantageous in the study of existence and uniqueness questions for parabolic problems, for essentially the same reason.

In 1949 Fox and Goodwin [17] emphasize that the trapezoidal method is useful also for *ODE*'s, when some components die out more quickly than others. In a short section called "building up errors" they demonstrate this on a particular linear example, by comparing the closed form solutions of the *ODE* and the *OΔE* for one particular step size. They are fairly brief there, since the central theme of their paper is not this, but to propose their new difference correction technique. In modern terminology their example is a "moderately stiff" *ODE* system.

In the same year (1949) Loud [27] studies the behavior of some methods on a linear system $dy/dt = Ay$, where A is a constant diagonalizable matrix. By a similarity transformation, the study is reduced to the scalar equations

$$(2.3) \qquad dy/dt = \lambda y, \qquad \lambda \in \mathrm{spectr}(A), \qquad h\lambda = q \in \bar{C}.$$

The solution of the corresponding difference equation is of the form,

$$(2.4) \qquad y_n = b_1 m_1^n + \ldots + b_k m_k^n, \qquad m_i = m_i(q).$$

Here $m_1(q)$ is an approximation to $\exp q$ The power series expansion of this root is used in an estimate of the global error. The other m_i tend to zero with q, for all methods but one. The methods studied include the classical Runge-Kutta method and some multistep methods from a textbook. The exception is Milne-Simpson's method, about which he comments "that there is a second term, which does not tend to zero, and high powers of it may well become large. For this reason this method is not suitable for long-run automatic computation". He does not discuss this quantitatively, e.g. the coefficient $-1/3$ which Rutishauser and I derived is not mentioned. Nor does he point out that the second term is disturbing only if $\mathrm{Re}\, q < 0$. Nevertheless, the techniques he uses are nowadays standard, and his opinions seem to be generally accepted today by the producers of general purpose software.

I like to conclude this section by three comments about the opinions and techniques.

I was for several years less categorical in my verdict of Milne-Simpson's method which is a 4th order accurate method and has a small error constant (1/180): "..., it is to be expected that 'some weakly unstable methods' are favorable if $\partial f/\partial y \geqq 0$, because of their small truncation error ... If $\partial f/\partial y < 0$ then the ... weak instability . . may make them inferior to methods with a lower 'order of consistency' in integrations over a long range. In such cases Runge-Kutta's and Adams' methods are safer.", [12], p. 52. On p. 72 I report the results of an integration of a Bessel differential equation from $t = 2$ to $t = 10$ with $h = 0.1$. The error at the endpoint is $2.2E-6$. By a 6th order accurate weakly (un)stable 4-step method the error was less than $4E-9$. I think it is still an open question how many branches we shall split "the general purpose" into. There may still be a market for the weakly (un)stable methods. In the fifties my neighborhood thought that I overemphasized the weak (in)stability. Now most people think that my verdict was not negative enough.

It is customary to apply the study of the test equation (2.3) also to non-linear systems, where λ runs through the spectra of the Jacobians for all t. Some questions about this generalizations will be discussed in Part II.

The reduction of the study of a linear system $dy/dt = Ay$ to the scalar equations (2.3) makes sense only if the transformation T, which diagonalizes A, has a condition number $\|T\| \cdot \|T^{-1}\|$ of moderate size. This is important, for example, in discussions of the method of lines mentioned above. There one has to consider a sequence of systems,

$$dy/dt = A_m y, \qquad m = 1, 2, 3, \ldots$$

The order of the system and the accuracy of the spatial discretization increase with m. It is not sufficient to consider the spectra of the matrices A_m, for the condition number of the diagonalising transformation may grow rapidly with m. The study of the test equation (2.3) may be misleading concerning restrictions on the choice of time step. The corresponding difficulty is well handled in the theory of *PΔE*'s, where the discretization in space and time are simultaneously studied, but it is sometimes overlooked in the discussion of the method of lines. In 1959 Kreiss made a profound study of the uniform boundedness of $\exp(A_m t)$ over all positive t and all matrices in a family of matrices of *fixed* order, but in this case the order is unbounded. It is sometimes advantageous to study these matters in terms of norms instead of spectra (energy method or contractivity analysis, more about this in Part II). Similarity transformations, other than diagonalization, may also be useful.

3. The period 1952–1963.

I do not think that Rutishauser and I were aware of Loud's paper when we met in 1951. We were, however, aware of an article by Todd [36], where he exemplifies a stronger type of instability, which is not removed when the step size tends to zero. I decided to look into these phenomena for a class of methods for the integration of first order systems, now called linear multistep methods, where y_n is intended to be an estimate of $y(t_n)$, $t_n = n \cdot h$, $n = 0, 1, 2, \ldots$:

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = h \sum_{j=0}^{k} \beta_j f(y_{n+j}, t_{n+j}). \tag{3.1}$$

This class contains many of the best known methods, such as the Adams methods, the backward differentiation methods, Euler's method, the leap-frog method and Milne-Simpson's method.

Since we consider arbitrary systems of *ODE*'s, the solution $y(t)$ is an arbitrary (sufficiently regular) function. Introduce the generating polynomials ϱ and σ, and the operator L_h

$$\varrho(\zeta) = \sum_{j=0}^{k} \alpha_j \zeta^j, \qquad \sigma(\zeta) = \sum_{j=0}^{k} \beta_j \zeta^j, \tag{3.2}$$

$$L_h y(t) = \sum \alpha_j y(t+jh) - h \sum \beta_j y'(t+jh). \tag{3.3}$$

$L_h y(t)$ will be called the local truncation error. We say that the *order of consistency* is p, if p is the largest integer such that $L_h P(t)$ vanishes identically for any pth degree polynomial. This requirement leads to $p+1$ homogeneous linear equations for the $2k+2$ coefficients of the method. By the expansion of $L_h y(t)$ into powers of h, it is then seen that for an arbitrary function y,

$$L_h y(t) \simeq ch^{p+1} y^{(p+1)}(t), \qquad (h \to 0), \tag{3.4}$$

where c and p are independent of the function y. Hence, for $y(t) = \exp t$ we obtain

$$\varrho(\exp h) - h\sigma(\exp h) \sim ch^{p+1}.$$

Set $\exp h = \zeta$. Then $h = \log \zeta \simeq \zeta - 1$, and

$$\varrho(\zeta)/\sigma(\zeta) - \log \zeta \sim (c/\sigma(1)) \cdot (\zeta - 1)^{p+1}, \qquad \zeta \to 1. \tag{3.5}$$

An attractive feature of k-step methods is that the amount of work per step is comparatively small. If the method is explicit, i.e. if $\beta_k = 0$, then only one evaluation of the function f is needed at each step. Some people are still of the opinion that k-step methods are basically unsound, for the following "philosophical" reason: "The future of the solution of (1.2) is uniquely determined by its value at one point. A sound approximate method should not be different from the underlying problem in such a fundamental respect." There is a point in this remark. Due to this fundamental difference, one has to consider the question of numerical stability, also for arbitrarily small values of $h\|f'(y)\|$. Since a fairly large number of multistep methods satisfy most reasonable requirements of stability, I do not see any "philosophical reasons" for discarding the whole class.

We follow Loud's example and consider the linear test problem,

$$dy/dt = \lambda y, \qquad h\lambda = q = \text{complex constant}.$$

When it makes sense we also include the the limiting case $q = \infty$.

The set of complex numbers q, for which all solutions of the difference equation obtained when a numerical method is applied to this linear test problem, is called the *stability region S* (or the region of absolute stability) of the method. A method is called *zero-stable* iff $0 \in S$. These definitions apply to any numerical method for the solution of initial value problems for *ODE*'s.

The concept of stability region did not become a hit until the beginning of the 60's, but as early as 1954 Gray [18] published diagrams related to stability regions in this sense. They contain, however, much more information than the plots of stability regions usually seen today and were perhaps too complicted to become popular.

We now return to the linear multistep methods and the linear test problem. The difference equation (3.1) then becomes linear, and the general solution is given by (2.4), where the m_i are roots of the characteristic equation,

$$\varrho(\zeta) - q\sigma(\zeta) = 0, \tag{3.6}$$

provided that the roots are simple. (See e.g. [19], p. 214 concerning the case of multiple roots). The method is zero-stable iff the characteristic equation satisfies the following root condition: no root should lie outside the unit circle, and the roots on the unit circle should be simple.

In 1956 I published a proof of the following *convergence theorem* [11]: Consider all *ODE* problems of the form (1.2), which have a unique solution on a finite interval, such that f satisfies a Lipschitz condition in a vicinity of that solution. Then, for a consistent method zero-stability is necessary and sufficient for the uniform convergence of the solution of *OΔE* (3.1) to the solution of the *ODE* (1.2), when h and the errors of the initial values tend to zero.

As in many other places I here modernized the terminology. The parallel to the equivalence theorem of Lax then becomes clearer. When I wrote [11], I was not yet familiar with the work of Lax, which was not published until 1955 [25], although it had been presented at a seminar in 1953. Had I known it, it is likely that I had changed my terminology. The terminology I use today is to a large extent due to Henrici [19], who did very much for the further development and the publicity of the theory of multistep methods.

Let us look more at the parallel development in *PDE*'s. The convergence theorem for linear multistep methods showed, roughly speaking, that if a method handles the trivial equation $dy/dt = 0$ with arbitrary initial values well, then it will also handle the equation $dy/dt = f(t,y)$ well, if f is Lipschitz-bounded. In 1962 Kreiss obtained an analogous result for *PΔE*'s, which we quote from [32], p. 58.

"If the difference system

$$u_{n+1} = C(\Delta t)u_n$$

is stable, and $Q(\Delta t)$ is a bounded family of operators, then the difference system

$$u_{n+1} = (C(\Delta t) + \Delta t Q(\Delta t))u_n$$

is also stable." On this occasion it is appropriate to mention that this was published in BIT [23]. The article also contained *the fundamental matrix theorem of Kreiss* on necessary and sufficient conditions for the uniform power-boundedness of a family of matrices (of fixed order). This discrete version of the theorem in [22] is a powerful tool to the stability analysis for difference methods for partial differential equations, see e.g. [32], Ch. 4.

Another aspect of linear multistep methods intrigued me, however, more than the convergence theorem. By (3.5) the order of consistency is related to the asymptotic behavior of $\varrho(\zeta)/\sigma(\zeta)-\log\zeta$ as $\zeta \to 1$, while the zero-stability is related to the location of the zeros of $\varrho(\zeta)$. This reminded me of analytic number theory, where number-theoretical results are deduced by the application of complex analysis to certain generating functions of various arithmetical functions. A central problem is to deduce knowledge about the zeros of Riemann's zeta-function from asymptotic properties of the function. I had been an enthusiastic student of this beautiful theory, but after one publication [9] and great hesitation I had decided to abandon it, since all the remaining interesting problems seemed too difficult for me. So I now thought: perhaps numerical analysis is not only fresh but also fun! When I played around a little and found that the "most consistent" methods for $k = 3$ and $k = 4$ were not zero-stable, then I became a life-time addict to Numerical Analysis.

After some time I was able to express *the conflict between consistency and stability* for linear multistep methods in the following way [11]: although the $2k+2$ coefficients of a linear k-step method can be chosen so that $p = 2k$, zero-stability implies that $p \leqq 2\lfloor k/2 \rfloor + 2$. For an explicit zero-stable method, $p \leqq k$.

Later [12] I found that if $p = k+2$, then the method is weakly stable in the sense that the difference equation obtained, when the method is applied to (2.3), may have an exponential growth of the type illustrated above for the leap-frog and the Milne-Simpson methods, even if $\operatorname{Re} q < 0$.

The proof of the convergence theorem mentioned above also provided a bound for the global error, of the following form, where M is the Lipschitz constant L multiplied by a factor that depends on the method, and l' is a bound for the sum of the norms of the initial errors and all the local errors until $t = t_n$.

$$\|y_n - y(t_n)\| \leqq l'K \exp(Mt_n),$$

which was characterized as "in general rather poor". The contribution of the local truncation errors to l' does not exceed

$$\sum_{\nu=1}^{n} ch^{p+1}\|y_{\nu-1}^{(p+1)}\| \leqq ct_n h^p \max\|y^{(p+1)}\|$$

which shows that for a *zero-stable method the order of consistency is also the order of accuracy.*

Error bounds of this type existed in the pre-computer literature. Collatz [2] quotes some of them. Sometimes the bound is expressed as follows, where l is an upper bound for the local truncation error.

$$\|y_n - y(t_n)\| \leqq l(\exp(Mt_n)-1)/(Mh). \tag{3.7}$$

Such bounds can easily become ridiculous over-estimates. For example, let $M = L$, $t = 10$, $f(y) = Ay$, where A is diagonalizable with all eigenvalues in the interval $(-5, 0)$. Then $\exp(Mt) > \exp 50 > 10^{21}$. If Euler's method is used with $h < 0.3$ (say) then the actual error will not exceed a constant of moderate size times h.

One way to improve this is by replacing the Lipschitz constant by an upper bound of the logarithmic norm of the Jacobian. Let $\|A\|$ be the operator norm of the matrix A induced by the vector norm $\|\cdot\|$,

$$\|A\| = \sup_x \|Ax\|/\|x\|.$$

The *logarithmic norm* of the matrix B is then defined as follows,

$$\mu(B) = \lim_{h\to 0+} \frac{\|I+hB\|-1}{h} \tag{3.8}$$

This concept was introduced in 1958 independently by Lozinskii [28] and Dahlquist [11]. Lozinskii applied it to obtain realistic error bounds for Adams' methods, while I applied it to weakly stable multistep methods. For related ideas, see also [16] and [39].

The improvement will be briefly illustrated by an application to Euler's method. I here partly follow Henrici [20], p. 107. Euler's method reads

$$y_{n+1} - y_n = hf(y_n).$$

Let the local truncation error at $t = t_n$ be $l\theta_n$, where θ_n is a vector whose norm does not exceed 1. Then

$$y(t_{n+1}) - y(t_n) = hf(y(t_n)) + l\theta_n.$$

Set

$$e_n = y_n - y(t_n), \qquad J_n = \int_0^1 f'(y(t+ve_n))dv.$$

Even if J_n depends on e_n, it follows that

$$e_{n+1} - e_n = hJ_n e_n - l\theta_n \quad \text{and} \quad \|e_{n+1}\| \leqq \|I + hJ_n\|\|e_n\| + l. \tag{3.9}$$

Let $1 + h\mu(J, h)$ be an upper bound of $\|I + hJ_n\|$. When $h \to 0$, $\mu(J, h)$ tends to an upper bound $\hat{\mu}$ of the logarithmic norm of the Jacobian. For example, if $\|\cdot\|$ is an inner-product norm, it can be shown that

$$(3.10) \qquad \|I+hJ_n\| \leqq \exp(h\mu(J_n)+\|hJ_n\|^2/2).$$

Then (3.9) yields, by induction, *a bound for* $\|e_n\|$ *of the same structure as* (3.7), *where M is replaced by* $\mu(J,h)$, which is never larger than the Lipschitz constant and *may even be negative*. Note, for example, that if $B = zI$, then $\mu(B) = \operatorname{Re} z$, while $\|B\| = |z|$. The efficiency of the estimate depends on the choice of norm, see e.g. Ström [35].

If $\mu(J,h) < 0$ we obtain a uniform bound, i.e. for all n,

$$(3.11) \qquad \|y_n - y(t_n)\| \leqq l/|h\mu(J,h)| \leqq (h/2)\max_{t<t_n}\|\ddot{y}\|/|\mu(J,h)|.$$

For inner-product norms this bound by (3.10), is valid when

$$\mu(J_n) < 0, \; \|hJ_n\|^2 < |2h\mu(J_n)|$$

for all n. This condition is sharp in the sense that for the linear test problem, i.e. if $hJ = q$, this is exactly the inequality which defines the stability region for Euler's method. *Similar bounds* and conditions can be derived for any consistent linear multistep method that is *strongly zero-stable* (i.e. all zeros of $\varrho(\zeta)$ except 1, are strictly inside the unit circle) and for which $\beta_k/\alpha_k > 0$, see [42], Lemma 2.3 and Theorem 3.3.

A particularly powerful result, essentially due to Desoer and Haneda [14] can be obtained for the implicit (or backward) Euler method,

$$y_{n+1} - y_n = hf(y_{n+1}).$$

As before, let $\hat{\mu}$ be an upper bound of the logarithmic norm of the Jacobian. If $\hat{\mu}h < 1$ the following bound holds in any norm,

$$(3.12) \qquad \|e_{n+1}\| \leqq (1-h\hat{\mu})^{-1}\|e_n\| + l\|(I-hJ_n)^{-1}\theta_n\|.$$

As above, $l\theta_n$ is the local truncation error, $\|\theta_n\| \leqq 1$.

A bound for $\|e_n\|$ is easily obtained from this. If $\hat{\mu} < 0$ it becomes particularly simple. Then the bound

$$(3.13) \qquad \|e_n\| \leqq l''/|h\hat{\mu}|,$$

holds for any n, if it is true for $n = 0$. Here l'' is an upper bound for the second term on the right hand side of (3.12). Note that *if* $\hat{\mu} < 0$, *these bounds hold for any step size h*! The results are also easy to generalize to variable step size.

Error bounds, valid without any stability restriction for the step size, exist also for some other methods, though not for arbitrary norms. A necessary condition on the method for the possibility of such bounds is obtained from the linear test problem. A somewhat weaker result for the trapezoidal method was obtained by me in 1963 and published in BIT [13], but now we have entered the stiff epoch, which belongs to Part II.

I regard bounds like (3.7) with their inadequate use of the Lipschitz constant L (or the related paremeter M) as typical for some pre-computer age theory of numerical methods. The bounds do not distinguish between well- and ill-conditioned initial value *ODE* problems. They are also unable to distinguish between methods, which can handle problems which are well-conditioned even if Lt is large ("stiff" or "moderately stiff" problems), with reasonably large step sizes (more about them in Part II), and methods (e.g. leap-frog), which cannot do so.

There is a parallel in numerical linear algebra. In 1943 Hotelling had observed that errors in the input to a step of the Doolittle variant of elimination can be amplified by a factor 4 in the step, see [21], p. 7. He concluded that for a linear system with p unknowns an estimated limit of the error amplification is 4^{p-1}. "The rapidity with which this increases with p is a caution against relying on the results of the Doolittle method or other similar elimination methods with any moderate number of decimal places when the number of equations and unknowns is at all large." About 1950 this misconception seemed to be widely spread. We have, after the work of Wilkinson and others, a better framework of concepts for discussing such matters, so that statements with more nuances can be made.

There were not many full time numerical analysts between Chebyshev (say) and the computer age. The development of numerical methods was to a large extent in the hands of scientists, engineers and mathematicians, some of whom were very prominent in their special fields, and also imaginative in the design of methods, but they rarely had the patience to develop the appropriate distinctions. For example, Hotelling was a prominent statistician, and [21] contains ideas which were new and interesting at the time.

There has been a great progress in numerical analysis during the last decades. Yet I do not think that its language is fully developed even today. There are gaps between the mathematician's and the numerical analyst's attitude to asymptotic formulas and to words like "bounded", "sufficiently small", "convergence" etc. Take the example mentioned above. If Mt is bounded then $\exp(Mt)$ is so too, in the language of Pure Mathematics. When it comes to numerical work, then 50 is not very large, but it does not sound right to use the word "bounded" in connection with exp(50).

Bounds analogous to (3.7) are useful for proving that the solution of an *ODE* depends continuously on initial values and parameters, and this is at times valuable knowledge also for a numerical analyst. For numerical estimation the bounds are useless. For some time people seemed to believe that there must be such a gap between rigorous bounds and actual errors. While the pure mathematician discusses "well posed" and "ill posed" problems as binary alternatives, the numerical analyst has a continuous scale of more or less "well- or ill-conditioned" problems. I believe that we need more transformations of this kind of the pure mathematician's terminology.

REFERENCES

1. J. Charney, R. Fjørtoft and J. von Neumann, *Numerical integration of the barotropic vorticity equation*, Tellus 2 (1950).
2. L. Collatz, *Über das Differenzenverfahren bei Anfangswertproblemen partieller Differentialgleichungen*, ZAMM 16 (1936), 239–247.
3. L. Collatz, *Numerische Behandlung von Differentialgleichungen*, 2. Aufl., Springer Verlag, Berlin/Göttingen/Heidelberg (1955).
4. L. Collatz and R. Zurmühl, *Beiträge zu den Interpolationsverfahren der numerischen Integration von Differentialgleichungen erster und zweiter Ordnung*, ZAMM 22 (1942), 42–55.
5. R. Courant, K. O. Friedrichs and H. Lewy, *Über die partiellen Differenzengleichungen der mathematischen Physik*, Math. Ann. 100 (1928), 32–74.
6. J. Crank and P. Nicolson, *A practical method for numerical integration of solutions of partial differential equations of heat-conduction type*, Proc. Cambridge Phil. Soc. 43 (1947), 50–67.
7. C. F. Curtiss and J. O. Hirschfelder, *Integration of stiff equations*, Proc. Nat. Acad. Sci. 38 (1952), 235–243.
8. G. Dahlquist, *Fehlerabschätzungen bei Differenzenmethoden zur numerischen Integration gewöhnlicher Differentialgleichungen*, ZAMM 31 (1951), 239–240.
9. G. Dahlquist, *On the analytic continuation of Eulerian products*, Ark. Mat. 1 (1951), 533–554.
10. G. Dahlquist, *Convergence and stability for a hyperbolic difference equation with analytic initial-values*, Math. Scand. 2 (1954), 91–102.
11. G. Dahlquist, *Convergence and stability in the numerical integration of ordinary differential equations*, Math. Scand. 4 (1956), 33–53.
12. G. Dahlquist, *Stability and error bounds in the numerical integration of ordinary differential equations*, Almqvist & Wiksell, Uppsala (1958), 86 pp. (Also distributed as Trans. Roy. Inst. Techn., Stockholm, Nr. 130, 1959.)
13. G. Dahlquist, *A special stability problem for linear multistep methods*, BIT 3 (1963), 27–43.
14. C. A. Desoer and H. Haneda, *The measure of a matrix as a tool to analyze computer algorithms for circuit analysis*, IEEE Trans. CT-19 (1972), 480–486.
15. R. P. Eddy, *Stability in the numerical solution of initial value problems in partial differential equations*, Naval Ordn. Lab. Memo. 10232, (1949).
16. H. Eltermann, *Fehlerabschätzungen bei näherungsweiser Lösung von Systemen von Differentialgleichungen erster Ordnung*, Math. Z. 62 (1955), 469–501.
17. L. Fox and E. T. Goodwin, *Some new methods for the numerical integration of ordinary differential equations*, Proc. Cambridge Phil. Soc. 45 (1949), 373–388.
18. H. J. Gray, Jr., *Numerical methods in digital real-time simulation*, Quart. Appl. Math. 12 (1954), 133–140.
19. P. Henrici, *Discrete Variable Methods in Ordinary Differential Equations*, J. Wiley & Sons, (1962).
20. P. Henrici, *Problems of stability and error propagation in the numerical integration of ordinary differential equations*, Proc. Int. Congr. Math. 1962, Almqvist & Wiksell, Uppsala, 102–113.
21. H. Hotelling, *Some new methods in matrix calculation*, Ann. Math. Stat. 14 (1943), 1–34.
22. H. O. Kreiss, *Über Matrizen die beschränkte Halbgruppen erzeugen*, Math. Scand. 7 (1959), 71–80.
23. H. O. Kreiss, *Über die Stabilitätsdefinition für Differenzengleichungen die partielle Differentialgleichungen approximieren*, BIT 2 (1962), 153–181.
24. P. Laasonen, *Über eine Methode zur Lösung der Wärmeleitungsgleichung*, Acta Math. 81, (1949), 309 ff.

25. P. D. Lax and R. D. Richtmyer, *Survey of the stability of linear finite difference equations*, Comm. Pure Appl. Math. 10 (1956), 267–293.
26. W. Liniger, *Zur Stabilität der numerischen Integrationsmethoden für Differentialgleichungen*, Faculté des Sciences de l'Université de Lausanne, (1957).
27. W. S. Loud, *On the long-run error in the numerical solution of certain differential equations*, J. Math. Phys. 28 (1949), 45–49.
28. S. M. Lozinskii, *Error estimate for numerical integration of ordinary differential equations I*, Izv. Vyss. Uçebn. Zaved. Matematika (1958) 52–90, errata (1959), 222. (In Russian).
29. W. E. Milne and R. R. Reynolds, *Stability of a numerical solution of differential equations* II, J. Assoc. Comp. Mach. 7 (1960), 46–56.
30. G. G. O'Brien, M. A. Hyman and S. Kaplan, *A study of the numerical solution of partial differential equations*, J, Math. Physics, 29, (1951), 223–251.
31. R. D. Richtmyer, *Difference Methods for Initial-value Problems*, Interscience, New York (1957).
32. R. D. Richtmyer and W. K. Morton, *Difference Methods for Initial-value Problems*, Interscience, New York (1967).
33. H. Rutishauser, *Über die Instabilität von Methoden zur Integration gewöhnlicher Differentialgleichungen*, Z. angew. Math. Physik 3 (1952), 65–74.
34. H. Stetter, *Stabilizing predictors for weakly unstable correctors*, Math. Comp. 19 (1955), 84–89.
35. T. Ström, *On logarithmic norms*, SIAM J. Num. Anal. 12 (1975), 741–753.
36. J. Todd, *Notes on numerical analysis* I, *Solution of differential equations by recurrence relations*, Math. Tables Aids Comput. 4 (1950), 39–44.
37. L. N. Trefethen, *Group velocity in finite difference schemes*, SIAM Rev. 24 (1982), 113–136.
38. A. M. Turing, *Rounding-off errors in matrix processors*, Quart. J. Mech., 1 (1948), 287–308.
39. W. Uhlmann, *Fehlerabschätzungen bei Anfangswertaufgaben gewöhnlicher Differentialgleichungen 1. Ordnung*, ZAMM 37 (1957), 88–99.
40. J. von Neumann and H. H. Goldstine, *Numerical inverting of matrices of high order*, BAMS 53 (1947), 1021–1099.
41. J. von Neumann and R. D. Richtmyer, *A method for numerical calculations of hydrodynamical shocks*, J. Appl. Phys. 21 (1950), 232 ff.
42. G. Söderlind and G. Dahlquist, *Error propagation in stiff differential systems of singular perturbation type*, Report TRITA-NA-8108 (1981), Royal Inst. Techn., S-10044 Stockholm.