CONFERENTIE VAN NUMERIEK WISKUNDIGEN

12, 13 en 14 oktober 1981

CONFERENTIEOORD WOUDSCHOTEN
ZEIST

Werkgemeenschap Numerieke Wiskunde

CONFERENTIE VAN NUMERIEK WISKUNDIGEN

12, 13 en 14 oktober 1981

CONFERENTIEOORD WOUDSCHOTEN
ZEIST

Werkgemeenschap Numerieke Wiskunde

ZESDE CONFERENTIE NUMERIEKE WISKUNDE

Uitgave verzorgd door het
MATHEMATISCH CENTRUM
AMSTERDAM

DOEL VAN DE CONFERENTIE

De Conferentie Numerieke Wiskunde wordt eenmaal per jaar gehouden onder auspiciën van de Werkgemeenschap Numerieke Wiskunde. Het doel van de conferentie is om kennis te nemen van recente ontwikkelingen binnen de numerieke wiskunde. Hiertoe worden jaarlijks twee thema's vastgesteld. Lezingen over deze thema's worden normaliter verzorgd door uitgenodigde buitenlandse deskundigen.

THEMA

De thema's zijn:

- Stabiliteitstheorie voor stijve beginwaardeproblemen.
- Aspecten van de eindige-elementenmethode.

ORGANISATIE

De organisatie is in handen van de voorbereidingscommissie bestaande uit de heren Wesseling (THD), van der Houwen (MC), Veltkamp (THE) en Verwer (MC), en van het Mathematisch Centrum.

UITGENODIGDE SPREKERS

R. JELTSCH, RWTH Aken, BRD
W. LINIGER, IBM Yorktown Heights, USA.
J. MEINGUET, Université Catholique de Louvain, België.
K.W. MORTON, University of Reading, UK.
P.A. RAVIART, CNRS - Université Pierre en Marie Curie, Paris, France.
G. WANNER, Université de Genève, Zwitserland.

Drie deelnemers hebben zich bereid verklaard in een korte bijdrage een schets van hun onderzoek op het gebied van één van de conferentiethema's te geven, t.w.
F. CROWET, Philips Research Laboratory, Brussel.
C. DIERIECK, Philips Research Laboratory, Brussel.
W.H. HUNDSDORFER, Rijks Universiteit Leiden.

PROGRAMMA

Maandag 12 oktober

10.00-11.15	aankomst en koffie	15.15-15.45	thee
11.15-12.30	opening, Wanner	15.45-16.45	Liniger
12.45	lunch	16.50-17.20	Crowet
14.15-15.15	Morton	17.25-17.55	Dierieck
		18.00	diner

dinsdag 13 oktober

8.00	ontbijt	12.45	lunch
9.00-10.00	Raviart	14.45-15.15	Hundsdorfer
10.00-10.30	koffie	15.15-15.45	thee
10.30-11.30	Jeltsch	15.45-16.45	Wanner
11.35-12.35	Meinguet	16.50-17.50	Morton
		18.00	diner

woensdag 14 oktober

8.00	ontbijt	12.45	lunch
9.00-10.00	Liniger	13.45-14.45	Meinguet
10.00-10.30	koffie	14.45	sluiting, thee
10.30-11.30	Raviart		vertrek
11.35-12.35	Jeltsch		

TITELS EN SAMENVATTINGEN VOORDRACHTEN

Maandag 12 oktober

11.15	opening	
	G. Wanner:	*Introduction to A-stability theory for stiff differential equations.*
14.15	K.W. Morton:	*Optimal finite-element methods for diffusion-convection problems.*
15.45	W. Liniger:	*Stability and contractivity theory for multistep and one-leg methods with uniform steps.*
16.20	F. Crowet:	*Streamfunction representation of an incompressible plane flow.*
17.25	C. Dierieck:	*The stream function formulation of the 2-d Stokes problem.*

Dinsdag 13 oktober

9.00	P.A. Raviart:	*Finite element approximation of nonsingular solutions of nonlinear problems.*
10.30	R. Jeltsch:	*A technique to compare stability regions of general linear methods with special applications to semi-discretizations of hyperbolic problems.*
11.35	J. Meinguet:	*A practical method for obtaining a priori error bounds in approximations of finite element type.*
14.45	W.H. Hundsdorfer:	*Nonlinear stability analysis for a simple Rosenbrock method.*
15.45	G. Wanner:	*Introduction to B-stability theory.*
16.50	K.W. Morton:	*Optimal finite-element methods for diffusion-convection problems.*

Woensdag 14 oktober

9.00	W. Liniger:	*The A-contractive methods: Stable methods with arbitrary variable integration steps for non-linear and variable coefficient problems.*
10.30	P.A. Raviart:	*Approximation of bifurcation points.*
11.35	R. Jeltsch:	*Dahlquist's second barrier cannot be broken.*
13.45	J. Meinguet:	*A practical method for obtaining a priori error bounds in approximations of finite element type.*

STREAMFUNCTION REPRESENTATION OF AN INCOMPRESSIBLE PLANE FLOW

F. Crowet, C. Dierieck, Philips Research Laboratory, Brussels, Belgium.

To describe 2-dimensional flows of incompressible fluids constrained to some cavity, it is convenient to express the velocity field in terms of a streamfunction. Indeed, since the fluid is supposed to be incompressible, the corresponding velocity field is divergence free. This difficult constraint can be automatically satisfied by introducing some streamfunction ψ whose derivatives are linked up to the components of the velocity by the relation:

$$v_1 = \partial_2 \psi \quad , \; v_2 = -\partial_1 \psi .$$

The object of our work is to examine the functional relation between vector-fields and streamfunctions.

The correspondence between a velocity field $\vec{v}$ and a streamfunction ψ is notoriously associated with a Poisson equation $(-\Delta\psi = \text{curl } \vec{v})$. However the boundary conditions are not explicitly stated. Indeed, one usually mentions boundary conditions of the Dirichlet type whenever the velocity field has a vanishing normal component at the boundary, i.e. the streamfunction should be constant on every connected component of the boundary, but these constants are unknown. In all other cases, exact boundary conditions are never specified.

Our original approach consists in having studied first the basic relation between equivalence classes of streamfunctions and velocity field, which was shown to be an isomorphism. As main result, we obtained a weak formulation of this isomorphism which enables the practical determination of a streamfunction associated with a velocity field. A decisive advantage of this variational approach is that one precisely knows for which velocity fields the streamfunction can be determined by a Dirichlet problem, a Neumann problem or a second order problem of a mixed type. Moreover if a Dirichlet problem is preferred to determine the streamfunction, then it is even possible to explicitly calculate the constants which the streamfunction should attain on each connected component of the streamfunction.

We should further point out that the variational formulation is well suited for numerical determination of some approximate of the streamfunction, using e.g. finite element methods.

THE STREAM FUNCTION FORMULATION OF THE 2-d STOKES PROBLEM

F. Crowet, C. Dierieck, Philips Research Laboratory, Brussels, Belgium.

Velocity fields describing 2-dimensional flows past several objects are determined by the Stokes problem. This paper deals with the problem of determining completely the associated stream function, either as solution of some variational problem or as solution of a biharmonic problem, including the exact boundary conditions.

NONLINEAR STABILITY ANALYSIS FOR A SIMPLE ROSENBROCK METHOD

W.H. Hundsdorfer, Rijks Universiteit Leiden.

Abstract. We consider the initial value problem

$$U'(t) = f(U(t)) \quad (t \geq 0) \ , \ U(0) = u_0$$

with $f: \mathbb{R}^s \to \mathbb{R}^s$ $(s \geq 1)$ and $u_0 \in \mathbb{R}^s$ known. It will be assumed that f satisfies

$$\langle f(\tilde{x})-f(x), \tilde{x}-x \rangle \leq \beta |\tilde{x}-x|^2 \quad (\text{for all } \tilde{x}, x \in \mathbb{R}^s)$$

where $\langle .,. \rangle$ denotes an inner product on $\mathbb{R}^s$, $|.|$ stands for the corresponding norm, and $\beta \in \mathbb{R}$ is a given number.

In recent years, Burrage and Butcher have studied the propagation of a perturbation in the initial value, when certain implicit Runge-Kutta methods are applied to this class of nonlinear differential equations.

In this talk such an analysis will be presented for the case that the following Rosenbrock method (M) is used.

$$(M) \qquad u_n = u_{n-1} + h(I-h\theta f'(u_{n-1}))^{-1} f(u_{n-1}) \quad (n \geq 1) \ .$$

Here $h > 0$ stands for the stepsize, $\theta > 0$ is a parameter and $u_n \approx U(nh)$ $(n \geq 1)$.

A TECHNIQUE TO COMPARE STABILITY REGIONS OF GENERAL LINEAR METHODS WITH SPECIAL APPLICATIONS TO SEMIDISCRETIZATIONS OF HYPERBOLIC PROBLEMS

Rolf Jeltsch and Olavi Nevanlinna

1. The comparison technique

Here we consider general "linear" methods, i. e. methods which, when applied to the test equation $y' = \lambda y$, $y(0) = 1$, $\lambda \in \mathbb{C}$, yield a characteristic polynomial

$$(1) \qquad \Phi(\zeta,\mu) = \sum_{i=0}^{k} \sum_{j=0}^{m} a_{ij}\, \mu^j\, \zeta^i \quad , \quad a_{ij} \in \mathbb{R} \quad , \quad \sum_{j=0}^{m} |a_{kj}| > 0 \quad .$$

The corresponding method will be referred to as a k-step, m-stage method. This class includes e. g. Runge-Kutta methods ($k = 1$) , linear multistep methods ($m = 1$) , linear k-step methods using m derivatives, predictor-corrector methods, cyclic methods. A method is represented by (1) and/or the algebraic function $P(\mu)$ defined by $\Phi(P(\mu),\mu) \equiv 0$. The stability region S consists of all μ for which the roots of $\Phi(\zeta,\mu)$ do not exceed 1 and those of modulus 1 are simple. For a convergent method $P(\mu)$ has exactly one branch $P_1(\mu)$ which approximates e^{μ} in the sense $P_1(\mu) \approx e^{\mu} - c_{p+1}\, \mu^{p+1}$ for $|\mu|$ small.
p is called the order and c_{p+1} the error constant of the method.
For some methods, e. g. Runge-Kutta methods, this order exceeds the usual order but for linear k-step multiderivative methods these notions are identical with the usual definitions.

The following theorem gives the possibility to compare stability regions and accuracy of two methods [17] , [23] .

Theorem 1 $\Phi(\zeta,\mu)$, $\Psi(\zeta,\mu)$ represent two methods. Let P, Q denote the algebraic functions, S_P , S_Q the stability regions and p_P , p_Q the error order. Assume that

a) $P(\mu)$ has at most s poles, which are not poles of $Q(\mu)$,

b) $q = \min \{p_P, p_Q\} \geq 2s + 1$,

c) Ψ is not a factor of Φ ,

d) $Q(\mu)$ has Property C .

Then $S_Q \not\subset \overline{S}_P$.
If b) is replaced by b') $p_P \geq p_Q \geq 2s - 1$, sign $c_Q = (-1)^s$ then one has either $S_Q \not\subset \overline{S}_P$ or $p_P = p_Q$ and $|c_P| \geq |c_Q|$. c_P , c_Q are the error constants. □

One can only compare a method with one which has Property C . The property basically insures, that outside the stability region S_Q exactly one branch, namely $Q_1(\mu)$, has a modulus exceeding one. Thus one can apply the order star technique in $\overline{\mathbb{C}} \setminus \overline{S}_Q$ using the quotient $P(\mu)/Q_1(\mu)$, [24] .
A test to determine whether a method has Property C or not is given. It is based on the oriented root locus $\Gamma = \{\mu \in \overline{\mathbb{C}} | \Psi(e^{it}, \mu) = 0 \text{ for } t \in [0,2\pi]\}$. Roughly speaking a method has Property C if and only if $\mathbb{C} \setminus \Gamma$ can be two colored, green and red say, such that red is always to the right and $(0,\varepsilon)$, $\varepsilon > 0$, is red. The proof makes use of an extension of Dahlquist's right hand rule: Assume some part γ of Γ is covered s-times in the same direction, when plotting the roots of $\Psi(e^{it},\mu)$ for t from 0 to 2π . Then the number of branches of $Q(\mu)$ of modulus exceeding 1 decreases by s , when moving from the right to left of γ . The red components of $\overline{\mathbb{C}} \setminus \Gamma$ will be the exterior of the stability region.

A method represented by (1) is called explicit if $a_{kj} = 0$, $j = 1,2,\ldots,m$. Hence if $P(\mu)$ belongs to an explicit m-stage method, $P(\mu)$ has a pole of order m at infinity. For explicit methods one obtains therefore the comparison theorem [16] .

Theorem 2 $\Phi(\zeta,\mu)$, $\Psi(\zeta,\mu)$ represent two explicit m-stage, n-stage methods. Assumptions c) d) remain as before but a) b) have to be replaced by
$\hat{a}$) $m \geq n$ $\hat{b}$) $q \geq 1$.
Then

(2) $S_Q \not\subset \overline{S}_P$. □

For explicit m-stage methods m is a good measure for the work needed to perform one integration step. Thus, characteristics like S_P , p_P , c_P of a m-stage method described by $P(\mu)$ should be defined relative to m . To do this imagine a hypothetical one-stage method $\tilde{P}(\mu)$ which is used m times with stepsize $h = H/m$ where H is the stepsize used by the P method. Thus we want that $\tilde{P}^m(\mu/m)$ has the same characteristics as $P(\mu)$. One easily finds the scaled

characteristics

$$(3) \qquad \tilde{S}_P := \frac{1}{m} S_P \quad , \quad \tilde{p}_P := p_P \quad , \quad \tilde{c}_P := c_P \, m^{p_P} \quad .$$

Theorem 3 [16] Φ , Ψ as in Theorem 2 but assumption $\hat{a})$ is no longer needed. Then

$$(4) \qquad \frac{1}{n} S_Q =: \tilde{S}_Q \notin \bar{\tilde{S}}_P := \frac{1}{m} \bar{S}_P \quad \text{if} \quad Q_1^m \left(\frac{\mu}{m}\right) \neq P_1^n \left(\frac{\mu}{n}\right) \quad . \quad \square$$

Hence, if one has two different, convergent explicit methods characterized by $P(\mu)$ and $Q(\mu)$ both having Property C then $\tilde{S}_Q \notin \tilde{S}_P$ and $\tilde{S}_P \notin \tilde{S}_Q$. Thus there exist two initial value problems such that one method performs better on one problem while the other one is better on the other problem. Therefore methods have always been designed to solve special classes of problems. In [16] , [17] the size of stability regions for three different purposes is investigated:

1. For "general" nonlinear ordinary differential equations. Here one wants a disk $D_r := \{\mu \mid |\mu + r| \leq r\}$ in the stability region S with r as large as possible.
2. Problems which arise from semidiscretization of parabolic differential equation where one wants intervals $[-a,0]$ in S with a as large as possible.
3. Hyperbolic problems, where stability intervals $I_r = \{iy \mid y \in (-r,r)\}$ should be in S with r as large as possible.

The comparison technique is used in the following way:
One guesses the "optimal" method with Property C and proves with the comparison theorem that the guess was right. Since for methods with Property C one has $\partial S \sim$ imaginary axis for $|\mu|$ small this technique is not suited to answer questions on stability interval on the negative real axis. It is however well suited for analyzing stability of methods for solving general nonlinear or hyperbolic problems.

2. Application to semidiscretizations of hyperbolic problems

2.1 First Order Systems

Consider for example the initial value problem $u_t + u_x = 0$ in $(0,1) \times (0,T]$, $u(x,0) = u_0$ given where $u_0(\ )$, $u(\ ,t)$ are periodic functions of period 1. If one discretizes first the space variable x one is lead to a system of first order differential equations $v' = Av$, where A is a skew symmetric matrix, [1], [6], p. 205. Thus the eigenvalues of A are purly imaginary and thus one is interested in having a long interval I_r in S. Let us first consider explicit methods. Van der Houwen [22] has shown that for m-stage Runge-Kutta methods one has $r \leq 2[m/2]$ and that equality can be attained if m is odd. In [16] the following general result is given.

Theorem 4 $\Phi(\zeta,\mu)$ belongs to a convergent, explicit m-stage method. Then either $I_m \not\subset \overline{S}$ or $\overline{I}_m = \overline{S}$ and Φ has a factor $\Psi_m(\zeta,\mu)$ of the form

$$\Psi_m(\zeta,\mu) = \zeta^2 - 2i^m T_m(-i\mu/m)\,\zeta + (-1)^m \quad , \tag{5}$$

where T_m is the m-th degree Tchebycheff polynomial. □

This is proved by comparing Φ with Ψ_1 using Theorem 2. In the scaled sense Ψ_1, which represents the mid point rule $y_{n+1} - y_{n-1} = 2hy'_n$, has the longest stability interval, namely I_1. Unfortunately the order is only 2. For any $r < 1$, $k \in \{2,3,4\}$ there are explicit linear k-step methods with $I_r \subset S$ and $p = k$. However for $k \equiv 1 \mod 4$ an explicit linear k-step method with $p = k$ has no nontrivial interval of stability on the imaginary axis, [16].
If we want $I_\infty \subset S$ the method has to be implicit. In [14] it was shown for linear multistep methods that if $I_\infty \subset S$ then $p \leq 2$ and among all those methods with $I_\infty \subset S$ and $p = 2$ the trapezoidal rule has the smallest error constant. In [25] this result is extended.

Theorem 5 [25] Let Φ represent a m-stage method with $I_\infty \subset S$. Then $p \leq 2m$. Among all methods of order $p = 2m$ with $I_\infty \subset S$ the methods characterized by

$$\Pi_m(Q,\mu) = N_m(\mu) - QN_m(-\mu) \tag{6}$$

have the smallest error constant

$$c_{\Pi_m} = (-1)^m \frac{(m!)^2}{(2m)!\,(2m+1)!} \quad .$$

Here

$$N_m(\mu) = 1 + \frac{m}{2m}\mu + \frac{m(m-1)}{(2m)(2m-1)}\frac{\mu^2}{2!} + \ldots + \frac{m(m-1)\ldots 1}{(2m)\ldots(m+1)}\frac{\mu^m}{m!}$$

and $N_m(\mu)/N_m(-\mu)$ is the m-th diagonal Padé-approximation to $\exp(\mu)$. □

Using Theorem 1 and a further comparison theorem the following is shown.

Theorem 6 [17] For a m-stage method, $m = 1$ or 2 , with $p > 2m$ and $I_r \subset S$ one has

$$(7) \qquad r \leq r_{m,opt} = \begin{cases} \sqrt{3} & m = 1 \\ \sqrt{15} & m = 2 \end{cases} \quad .$$

If $p = 2m$ and $I_r \subset S$ with $r > r_{m,opt}$, then

$$|c_p| \geq \begin{cases} \frac{1}{12}\left(1 - \frac{3}{r^2}\right) & , \quad m = 1 \\ \frac{1}{720}\left(1 - \frac{15}{r^2}\right) & , \quad m = 2 \end{cases} \quad .$$

For $m = 2$ one never has equality in (7) . However if for $m = 1$ one has equality in (7) and Φ is irreducible then it must be Milne-Simpson where

$$\Phi(\zeta,\mu) = \zeta^2 - 1 + \mu(\zeta^2 + 4\zeta + 1)/3 \quad . \qquad \square$$

The comparison technique fails for $m > 2$ for there are too many branch points. The result for $m = 1$ has also been proved in [5] using a totally different approach. Milne-Simpson has only order 4 . However for any $r < r_{1,opt} = \sqrt{3}$ there existes a linear 4-step method of order 6 such that $I_r \subset S$, see also [6] .

2.2 Second Order System

Consider for example the initial value problem $u_{tt} = u_{xx}$ in $[0,1] \times (0,T]$, $u(x,0)$, $u_t(x,0)$, $x \in [0,1]$ and $u(0,t)$, $u(1,t)$, $t \in [0,T]$ given.

If one discretizes first the space variable one is lead to a system of second order differential equations

(8) $\quad v'' = Av \quad ,$

where A is a symmetric negative definite matrix, [8] , [20] p. 260.
In the following a stability analysis of numerical methods for $y'' = g(t,y)$ is presented, [11] , [17] . To keep the notation consistent with the previous sections we consider the test equation $y'' = \lambda^2 y$, $\lambda \in I_\infty = i\mathbb{R}$. Again we consider "linear" methods, i. e. methods which, when applied to the test equation yield a characteristic polynomial

$$(9a) \qquad \Psi(\zeta,\mu) = \sum_{i=0}^{k} \sum_{j=0}^{2m} a_{ij}\, \mu^j\, \zeta^i \quad , \quad a_{ij} \in \mathbb{R} \quad , \quad \sum_{j=0}^{2m} |a_{kj}| > 0 \quad ,$$

$$k > 1 \quad , \quad \mu := h\lambda$$

with

$$(9b) \qquad \Psi(\zeta,-\mu) = \Psi(\zeta,\mu) \quad , \quad \text{i. e.} \quad a_{ij} = 0 \quad \text{for} \quad j \quad \text{odd.}$$

Again we call the corresponding method a k-step, m-stage method. This class of methods includes e. g. linear k-step methods ($m = 1$), [12] p. 289 , [15] linear k-step 2m-derivative methods $\sum_{i=0}^{k} \sum_{j=0}^{2m} a_{ij}\, h^j\, y_{n+i}^{(j)} = 0$, Nyström-, hybrid methods [11] .
We consider convergent methods only, i. e. the roots of $\Psi(\ ,0)$ do not exceed 1 in modulus and the ones of modulus 1 are at most double roots. The error order and error constant c_p is defined in exactly the same way as in the first section. Observe that for linear k-step methods $c_p = C/2$ where C is the error constant introduced in Henrici [12] . The exact solutions of the test equation are bounded for $\lambda \in I_\infty \setminus \{0\}$. One is therefore interested in having a long interval $I_r \setminus \{0\}$ on the imaginary axis inside the stability region. S .
In order to investigate whether one should solve (8) directly or transform it to a system of first order we introduce the following map Σ .
If $\Phi(\zeta,\mu)$ has form (1) and characterizes a m-stage method for solving $y' = f(t,y)$ then

$$(10) \qquad \Sigma\, \Phi := \Phi(\zeta,\mu)\, \Phi(\zeta,-\mu)$$

characterizes a m-stage method for solving $y'' = g(t,y)$.

Observe that using our definitions the error order and error constant are invariant under the map Σ . Moreover if I_r is the longest stability interval of Φ then $I_r \subset \overline{S}_{\Sigma\Phi}$ and r cannot be enlarged. In addition for multistep m-derivative methods the computational work to solve (8) by $\Sigma\Phi$ is the same as solving the corresponding first order system by Φ . Let us first consider explicit methods, i. e. $a_{kj} = 0$, $j = 1,2,\ldots,2m$. Theorem 4 gives immediately the result.

Theorem 7 $\Psi(\zeta,\mu)$ given by (9) belongs to a convergent, explicit m-stage method. Then either $I_{2m} \not\subset \overline{S}$ or $I_{2m} = \overline{S}$ and Ψ has a factor Ψ_{2m} given by (5) . □

Hence if we scale by m as before then Ψ_2 , which represents Störmers formula $y_{n+1} - 2y_n + y_{n-1} = h^2 y_n''$, has the longest scaled stability interval. Observe that $\Sigma\Psi_1\big|_{(\sqrt{\zeta},\, \mu/2)} = \Psi_2(\zeta,\mu)$. If we want $I_\infty \subset \overline{S}$ then the method has to be implicit. In [4] it was shown for linear multistep methods that if $I_\infty \subset \overline{S}$ then $p \leq 2$ and among all those methods with $I_\infty \subset \overline{S}$ and $p = 2$ the method

$$(11) \qquad y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{4}\,(y''_{n+1} + 2y''_n + y''_{n-1})$$

has the smallest error constant. In general one finds using order star theory the following result.

Theorem 8 [11] . Let Ψ given by (9) represent a m-stage method with $I_\infty \subset \overline{S}$. Then $p \leq 2m$. Among all methods of order $p = 2m$ with $I_\infty \subset \overline{S}$, the methods characterized by $\Sigma\Pi_m$ have the smallest error constant. □

Clearly (11) is represented by $\Sigma\Pi_1$ where Π_1 characterizes the trapezoidal rule. (11) has long been known, see e. g. [20] p. 263. In order to prove a result corresponding to Theorem 6 one has to extend the comparison theorems and make use of the fact (9b) .

Theorem 9 [17] For a linear multistep method ($m = 1$) with $p > 2$ and $I_r \setminus \{0\} \subset S$ one has $r \leq \sqrt{6}$.
Equality occurs only for Cowell's method given by

(12) $\quad y_{n+1} - 2y_n + y_{n-1} = h^2(y''_{n+1} + 10y''_n + y''_{n-1})/12$

or trivial modifications of it.

If $p = 2$ and $I_r \setminus \{0\} \subset S$ with $r > \sqrt{6}$, then

(13) $\quad |c_p| \geq \frac{1}{12}\,(1 - \frac{6}{r^2})$. □

In (13) equality occurs for a certain method given in [17] and, when applied to the wave equation is listed in [1] p. 200 . Method (12) has order 4 only. However for $r < \sqrt{6}$ there exist linear 4-step methods of order $p = 6$ with $S = I_r \setminus \{0\}$, [6] , [15] , [19] .

DAHLQUIST'S SECOND BARRIER CANNOT BE BROKEN

In this paper we consider linear multistep methods

(1) $\quad \sum_{i=0}^{k} \alpha_i\, y_{n+i} = h \sum_{i=0}^{k} \beta_i\, f(t_{n+i}, y_{n+i})$, $\alpha_i, \beta_i \in \mathbb{R}$, $h > 0$,

for solving initial value problems $y' = f(t,y)$, $y(0)$ given. Here h is the stepsize and y_n an approximation to the exact solution $y(t_n)$, where $t_n = nh$. The characteristic polynomial, see [18] , is $\Phi(\zeta,\mu) = \rho(\zeta) - \mu\sigma(\zeta)$, where

(2) $\quad \rho(\zeta) = \sum_{i=0}^{k} \alpha_i\, \zeta^i$, $\sigma(\zeta) = \sum_{i=0}^{k} \beta_i\, \zeta^i$.

We use the normalization $\sigma(1) = 1$. The stability region S is defined as in [18] . The error order p and the error constant c_{p+1} given in [18] describe the local error

(3) $\quad [L_h y](t) := \sum_{i=0}^{k} (\alpha_i y(t + ih) - \beta_i y'(t + ih))$

for small h ; for one has

(4) $\quad [L_h y](t) := c_{p+1}\, h^{p+1} y^{(p+1)}(t) + c_{p+2} h^{p+2} y^{(p+2)}(t) + O(h^{p+3})$,

$\quad c_{p+1} \neq 0$,

for sufficiently smooth $y(t)$.
When solving stiff initial value problems where one has large differences in time constants one is interested in methods where the stability region S contains as much of the left half plane $\mathbb{C}^- = \{\mu | \text{Re } \mu \leq 0\}$ as possible. If $\mathbb{C}^- \subset S$ then a method is called A-stable. Dahlquist [3] has shown, that the order p of an A-stable linear multistep method cannot exceed 2 and among the ones with $p = 2$ the trapezoidal rule has the smallest error constant.
There are two ways to "break" this second Dahlquist barrier.
Either one can change, enlarge the class of methods unter consideration or one can replace this strong concept of A-stability by a somewhat weaker stability condition. In this presentation we shall consider the second approach and replace A-stability by A(α)-stability [26] , A_0-stability [2] and stiff stability [7] , respectively. For all these stability concepts one has found high order almost A-stable methods [2] , [10] , [13] , [17] , [21] . In [17] a general principal how such methods can be constructed is given in the proof of the following

Theorem 1 Let ρ , σ represent a linear k-step method of order p with $D_R := \{\mu | \ |\mu + R| \leq R\} \subset S$ for some $R > 0$. Then to any $K \in \mathbb{N}$, $r < R$ and any closed set $\Omega \subset \text{Int } S \subset \overline{\mathbb{C}}$ there exists a linear k+K-step method $\tilde{\rho}$, $\tilde{\sigma}$ of order $p + K$ with $D_r \subset \tilde{S}$ and $\Omega \subset \text{Int } \tilde{S}$. Moreover if ρ , σ is an explicit method then $\tilde{\rho}$, $\tilde{\sigma}$ can be chosen to be explicit too. □

Applying this theorem to the θ method with $\theta > 1/2$ and $1/(2\theta - 1) < \varepsilon$ yields the

Corollary [17]. Let $k \in \mathbb{N}$, $R > 0$ and $\varepsilon > 0$. Then there exist linear multistep methods with order $p = k$, $D_R \subset S$ and $\overline{\mathbb{C}} \setminus S \subset \{\mu | \ |\mu| < \varepsilon\}$. □
This Corollary contains as special examples the results in [10] and [13] .
The proofs are constructive in the sense that the methods are explicitly given. However from the construction of these methods one observes that one cannot hope to really get more accurate methods. For example in [13] to any $D > 0$ stiffly stable methods with $\{\mu \in \mathbb{C} | \text{Re } \mu \leq -D\} \subset S$, and $p = k$ have been constructed. But since $c_{k+1} = O(D^{-k})$ as $D \to 0$ these methods are not accurate. This may indicate that in addition to a high error order one should request that the

error constant is kept small too. This can indeed be done when $p = k = 3$. Genin [9] has given a two parameter family of stiffly stable methods where

$$D = \frac{1}{-36c_4\,(b_2 - 144c_4^2)} \quad . \tag{5}$$

Here b_2 is a positive parameter and the error constant c_4 is a negative parameter. Thus for any $c_4 < 0$ and $D > 0$ there exists a linear 3-step method of order 3 with the prescribed error constant c_4 and stiff stability parameter D. Hence this third order method can be made almost A-stable and the error constant can be made as small as one wishes. These methods are however bad since one can show $c_{p+2} = c_5 = O(D^{-1})$ as D tend to 0. These results indicate that it is not enough to describe the accuracy of a method by just considering the order and the error constant. Moreover these classical notions of order and error constant measure only the first term of the asymptotic expansion of the error as h tends to 0. Hence these are concepts for small values of h. If one applies methods with "large" stability regions to the above mentioned stiff systems stability allows to use a stepsize h for which the asymptotic error expansion is no longer adequate. In this case the concepts of error order and error constant are no longer appropriate. Therefore in [17] the accuracy for fixed positive h is measured by the L_1-norm of the Peano-kernel of the local error (3). For $y(t) \in C^q$, $q \le p+1$ one has by the Peano-kernel theorem

$$[L_h y](t) = h^q \int_{-k}^{0} K_q(s)\, y^{(q)}(t - hs)\, ds \tag{6}$$

where the Peano-kernel is given by

$$K_q(s) = L_1 \frac{s_+^{q-1}}{(q-1)!} \quad ,$$

$$s_+ = \begin{cases} s & \text{for } s > 0 \\ 0 & \text{for } s \le 0 \end{cases} \quad .$$

Let M be the smallest constant such that <u>for all</u> $y \in C^q$ one has

$$|[L_h y](t)| \le h^q M \sup_{s \in [t, t+kh]} |y^{(q)}(s)| \quad . \tag{7}$$

By (6) we see that

$$M = \int_{-k}^{0} K_q(s)\, ds =: \|K_q\|_1 \quad .$$

The estimate (7) for the local error is the best possible which is valid for all functions in C^q. Hence it will be realistic for some functions, but rather pessimistic for very smooth ones. Observe that the estimate (7) consists of three factors. M depends on the method only. In particular it is independent of the stepsize h. The last factor measures the smoothness of the solution of the underlying problem and is almost independent on h and the method. Observe that

$$c_{p+1} = \int_{-k}^{0} K_{p+1}(s)ds$$

and thus

$$\|K_{p+1}\|_1 = |c_{p+1}|$$

if and only if $K_{p+1}(s)$ does not change sign. In [17] the following lower bounds for $\|K_q\|_1$ are proved.

Theorem 2 Let $k \in \mathbb{N}$. There exists a constant C_k such that the following holds. If $D_R \subset S$ and $q \leq p + 1$ then

$$\|K_q\|_1 \geq C_k \left(\frac{R}{3}\right)^{p-2} .$$

C_k can be specified more precisely if $q = p + 1$ for one has

Theorem 3 Let $k \in \mathbb{N}$ and $k + 1 \geq p \geq \max\{k,2\}$. Assume that

$$D_R \subset S \qquad \text{if } p = k + 1$$

$$D_R \cup \{\infty\} \subset S \qquad \text{if } p = k .$$

Then

$$-c_{p+1} \geq \frac{1}{12} \left(\frac{R}{6}\right)^{p-2} . \tag{8}$$

Observe that $c_{p+1} > 0$ if $p \leq k$. These results show clearly that A-stability cannot be approximated with "accurate" methods of order $p > 2$. Applying (8) to Genin's example we see that even so one can make D in (5) as small as one wishes for any fixed $c_4 < 0$, the radius R will remain bounded by $-72c_4$.

In Jeltsch, Nevanlinna [17] similar results for explicit methods are given. In addition results of the same flavor are given where the size of the Peano-kernel is related to stability intervals on the real axis.

The following result which is proved using the comparison technique indicates again that there are no accurate high order almost A-stable linear multistep methods. Let $H_\delta := \{\mu \mid |\mu| \leq \delta, \operatorname{Re}\mu \leq 0\}$.

Theorem 4 [17] . There is a constant C such that for a linear multistep method with $p \geq 3$ and $D_R \cup H_\delta \subset S$ one has

$$\delta^2 R < C \quad . \qquad \square$$

References

[1] W. F. Ames, Numerical Methods for Partial Differential Equations, Academic Press, New York, San Francisco, 1977

[2] C. W. Cryer, A new class of highly-stable methods: A_0-stable methods, BIT 13,(1973),153 - 159

[3] G. Dahlquist, A special stability problem for linear multistep methods, BIT 3, (1963), 27 - 43

[4] G. Dahlquist, On accuracy and unconditional stability of linear multistep methods for second order differential equations, BIT 18, (1978), 133 - 136

[5] K. Dekker, Stability of linear multistep methods on the imaginary axis, to appear

[6] V. A. Dougalis, Multistep-Galerkin Methods for Hyperbolic Problems, Math. Comp. 33, (1979), 563 - 584

[7] C. W. Gear, The automatic integration of stiff ordinary differential equations, Information Processing 68, A. J. H. Morell, ed., North Holland, Amsterdam, 187 - 193, 1969

[8] E. Gekeler, Galerkin-Runge-Kutta Methods and Hyperbolic Initial Boundary Value Problems, Computing 18, (1977), 79 - 88

[9] Y. Genin, A new approach to the synthesis of stiffly stable linear multistep formulas, IEEE Transactions on Circuit Theory, vol. CT-20, 352 - 360, 1973

[10] R. D. Grigorieff, J. Schroll, Über A(α)-stabile Verfahren hoher Konsistenzordnung, Computing 20, (1978), 343 - 350

[11] E. Hairer, Unconditionally stable methods for second order differential equations, Numer. Math. 32, (1979), 373 - 379

[12] P. Henrici, Discrete variable methods in ordinary differential equations, New York, Wiley, 1962

[13] R. Jeltsch, Stiff stability and its relation to A_0- and A(0)-stability, SIAM J. Numer. Anal. 13, (1976), 8 - 17

[14] R. Jeltsch, Stability on the imaginary axis and A-stability of linear multistep methods, BIT 18, (1978), 170 - 174

[15] R. Jeltsch, Complete Characterization of Multistep Methods with an Interval of Periodicity for Solving y" = f(x,y), Math. Comp. 32, (1978), 1108 - 1114

[16] R. Jeltsch, O. Nevanlinna, Stability of explicit time discretizations for solving initial value problems, Numer. Math. 37, (1981), 61 - 91

[17] R. Jeltsch, O. Nevanlinna, Stability and accuracy of time discretizations for initial value problems, to appear in Helsinki University of Technology, Espoo, Math. Inst. Report - Mat 1981 and Numer. Math.

[18] R. Jeltsch, O. Nevanlinna, A technique to compare stability regions of general linear multistep methods with special applications to semidiscretizations of hyperbolic problems, Zeist, 12 - 14 Oct. 1981

[19] J. D. Lambert, I. A. Watson, Symmetric multistep methods for periodic initial value problems, J. Inst. Math. Appl. v 18, (1976), 189 - 202

[20] R. D. Richtmyer, K. W. Morton, Difference Methods for Initial Value Problems, J. Wiley, London, Sidney, 2nd Ed., 1967

[21] H. J. Stetter, Improved absolute stability of predictor-corrector schemes, Computing 3, (1968), 286 - 296

[22] P. J. van der Houwen, Construction of integration formulas for initial value problems, Amsterdam, North-Holland, 1977

[23] G. Wanner, The comparison method for the study of order and stability properties, in Proceedings of the informal symposium on numerical methods for initial value problems, ed. by G. Dahlquist et. al., Royal Inst. of Tech., Stockholm, Dep. of Numerical Analysis, TRITA-NA-7923, 24, 29 - 32, 1979

[24] G. Wanner, E. Hairer, S. P. Nørsett, Order stars and stability theorems, BIT 18, (1978), 475 - 489

[25] G. Wanner, E. Hairer, S. P. Nørsett, When I-stability implies A-stability, BIT 18, (1978), 503

[26] O. B. Widlund, A note on unconditionally stable linear multistep methods, BIT 7, (1967), 65 - 70

Rolf Jeltsch, Institut für Geometrie und Praktische Mathematik,
RWTH Aachen, D-5100 Aachen, Fed. Rep. Germany

Olavi Nevanlinna, Institut of Mathematics, Helsinky University of Technology,
SF 02150 Espoo 15, Finland

W. Liniger, IBM Yorktown Heights, USA

1. Stability and Contractivity Theory for Multistep and One-leg Methods with Uniform Steps.

During the past few years, considerable progress was made in the stability analysis of numerical methods for ordinary differential equations. Stability and contractivity results were obtained for linear multistep and one-leg methods as applied to certain classes of linear and nonlinear equations. Families of novel integration formulas (representing proper subsets of, say, the A_0-stable or A-stable formulas) were identified by these theories which possess much stronger stability properties then many of the popular ones (say, the backward differentiation formulas). These developments will be surveyed for the case of uniform grids.

2. The A-Contractive Methods: Stable Methods with Arbitrary Variable Integration Steps for Nonlinear and Variable Coefficient Problems.

1. The linear multistep formula

$$\sum_{j=0}^{k} \alpha_j x_{n-k+j} - h \sum_{j=0}^{k} \beta_j \dot{x}_{n-k+j} = 0 \tag{1}$$

is said to be contractive [1] at a location q in the complex plane if for all $(k+1)$-tupels $\{y_0, y_1, \ldots, y_k\}$ satisfying $\sum_{j=0}^{k} (\alpha_j - q\beta_j) y_j = 0$ we have $\| Y_k \| \leq \| Y_{k-1} \|$ in some given norm $\| \cdot \|$, where $Y_m = (y_{m-k+1}, \ldots, y_m)$, $m = k-1, k$. If not stated otherwise, the results given in this paper are valid in the ℓ_∞-norm. If (1) is applied to $\dot{x} = \lambda x$, λ constant, complex, then contractivity at $q = h\lambda$ implies $\| X_n \| \leq \| X_{n-1} \|$, $n = k, k+1, \ldots$, where $X_n = (x_{n-k+1}, x_{n-k+2}, \ldots, x_n)$, and thus contractivity at q implies stability at q. The formula (1) is said to be A-contractive [1] if it is contractive at all q, Re $q \leq 0$. A-contractivity implies A-stability and thus A-contractive formulas are of order of accuracy $p \leq 2$.

In [1] we identified the one-parameter family of all A-contractive $p = k = 2$-formulas with uniform steps. This result was generalized to variable steps in [2] as follows: For steps (h_{n-1}, h_n) of arbitrary ratio $0 < r_n = h_n / h_{n-1} < +\infty$, we derived the one-parameter family of all $p = k = 2$-formulas which are both A-contractive with respect to the ℓ_∞-norm and contractive for all dissipative non-linear systems in inner product norms (G-stable). Here we give, for any $k \geq 1$ and for arbitrary non-uniform grids, the $(k-1)$-parameter family of all A-contractive k-step second-order formulas. These formulas have very strong stability properties. For example, stability can be guaranteed for *arbitrary* step sequences if at each step *any one* of the A-contractive formulas is implemented as a one-leg method and applied to the variable coefficient test problem $\dot{x} = \lambda(t)x$ with *any* $\lambda(t)$, *Re* $\lambda(t) \leq 0$. By contrast, the variable-step version

of the two-step backward differentiation formula, which for uniform steps is A-stable but which is not A-contractive in the ℓ_∞-norm, can be destabilized by certain combinations of variable steps and variable coefficients [2, 3].

2. For variable steps $h_n = t_n - t_{n-1}, > 0$, $n = 1, 2, \ldots$, the multistep formula can be written as

$$\sum_{j=0}^{k} \alpha_{j,n} x_{n-k+j} - h_n \sum_{j=0}^{k} \beta_{j,n} \dot{x}_{n-k+j} = 0. \tag{2}$$

Let

$$\Theta_{j,n} := (t_{n-k+j} - t_{n-k})/h_n,\ j = 0, \ldots, k. \tag{3}$$

We assume that $\Theta_{j,n} > \Theta_{\ell,n}$ whenever $j > \ell$ to assure that the grid $\{t_n\}$ is monotone increasing. For uniform steps, $\Theta_{j,n} = j$. The definition (3) implies that $\Theta_{0,n} = 0$ and $\Theta_{k,n} = \Theta_{k-1,n} + 1$ for all n and the $k-1$ independent quantities $\Theta_{1,n}, \ldots, \Theta_{k-1,n}$ can be used to describe the variability of the grid. If we define the m^{th} weighted moments of the $\alpha_{j,n}$ and $\beta_{j,n}$ by

$$A_{m,n} := \sum_{j=0}^{k} \Theta_{j,n}^m \alpha_{j,n},\ B_{m,n} := \sum_{j=0}^{k} \Theta_{j,n}^m \beta_{j,n} \tag{4}$$

and normalize the formula (2) by

$$B_{0,n} = \sum_{j=0}^{k} \beta_{j,n} = 1, \tag{5}$$

then (2) is consistent iff $A_{0,n} = 0$ and $A_{1,n} = B_{0,n} = 1$, i.e. iff

$$\alpha_{k,n} = 1 - \sum_{j=0}^{k-2} (\Theta_{k-1,n} - \Theta_{j,n}) u_{j,n},$$

$$\alpha_{k-1,n} = -1 + \sum_{j=0}^{k-2} (\Theta_{k,n} - \Theta_{j,n}) u_{j,n}, \tag{6}$$

where

$$u_{j,n} := -\alpha_{j,n},\ j = 0, \ldots, k-2 \tag{7}$$

for $k \geq 2$. Note that $\alpha_{k,n} = -\alpha_{k-1,n} + \sum_{j=0}^{k-2} u_{j,n}$. We now state the result of this paper; the proof will be given elsewhere.

Theorem:

For any $k \geq 1$, the set of all A-contractive k-step second-order formulas for arbitrary variable steps is the $(k-1)$-parameter family defined by (6),

$$\beta_{j,n} = (\Theta_{j,n} - \frac{1}{2} A_{2,n}) \alpha_{j,n},\ j = 0, 1, \ldots, k, \tag{8}$$

and (for $k \geq 2$) by (7) and

$$u_{j,n} \geq 0,\ j = 0, \ldots, k-2,$$

$$\sum_{j=0}^{k-2} (\Theta_{k,n} - \Theta_{j,n}) u_{j,n} \leq 1, \tag{9}$$

where $\Theta_{j,n}$ and $A_{2,n}$ are defined by (3) and (4), respectively.

□

Note that a) for $k = 1$, the only A-contractive (A-stable) formula with $p = 2$ is the Trapezoidal Rule; b) relations (6) - (9) imply (5) and the second-order accuracy

condition $A_{2,n} = 2B_{1,n}$; c) the condition (9) is necessary and sufficient for contractivity at $q = 0$; d) for $k \geq 2$, the A-contractive second-order k-step formulas are in one-to-one correspondence with the points of the $(k-1)$-dimensional simplex S defined by (9). The vertices of S, i.e. the intercepts $\{u_{j,n} = (\Theta_{k,n}-\Theta_{j,n})^{-1}$, $u_{\ell,n} = 0$ for $\ell \neq k$, $j = 0, 1, \ldots, k-2\}$ of the $u_{j,n}$-axes with the plane of equality (9b), and the origin, represent the $(k-j)$-step Trapezoidal Rules

$$\frac{1}{\Theta_{k,n}-\Theta_{j,n}}(x_n-x_{n-k+j})-\frac{h_n}{2}(\dot{x}_n + \dot{x}_{n-k+j}) = 0,\ j = 0, \ldots, k-1, \qquad (10)$$

respectively, normalized by (5). The edge $\{u_{j,n} = 0,\ 0 \leq j \leq k-3;\ O \leq u_{k-2,n} \leq (\Theta_{k,n}-\Theta_{k-2,n})^{-1}\}$ of S represents the A-contractive $p = k = 2$-formulas derived earlier.

References

1. O. Nevanlinna and W. Liniger, "Contractive Methods for Stiff Ordinary Differential Equations. Part I," BIT 18 (1978) 457-474.

2. G. Dahlquist, W. Liniger and O. Nevanlinna, "Stability of Two-step Methods for Variable Integration Steps," IBM Research Report RC 8494, September 1980; submitted to SIAM J. Numer. Anal.

3. F. Odeh and W. Liniger, "On A-stability of Second-order Two-step Methods for Uniform and Variable Steps," Proc. IEEE Intl. Conf. Circuits and Computers 1980 (N. B. Rabbat, Ed.) Vol. 1, pp. 123-126.

A PRACTICAL METHOD FOR OBTAINING A PRIORI ERROR BOUNDS IN APPROXIMATIONS OF FINITE ELEMENT TYPE - J. Meinguet, Univ. Catholique de Louvain, Belgium.

Error *estimates* (as opposed to *bounds*) often involve unknown constants, in which case they are naturally ill-suited to *quantitative* studies. The present lectures are based essentially on some fairly recent research work whose primary purpose was precisely to cover this shortcoming of such estimates by devising a *method of practical value* for obtaining *realistic a priori upper bounds of approximation errors* in a *wide variety of situations*. As explained in detail in [13, 14, 15, 22], such a method can be elaborated from a *structural analysis of error coefficients*, which evolves quite naturally in the setting of operator theory in normed linear spaces while referring to such classical tools as the *Peano kernel theorem* and its *qualitative* generalization known as the *Bramble-Hilbert lemma*. As exemplified in the above references and also in [1, 11], that method can yield at *reasonable cost* explicit upper bounds for certain *generic constants* which tend to pervade the modern literature on error estimation (typically in spline analysis, in connection with the rate of convergence of the finite element method, see e.g. [6] and [7], Theorems 2, 4, 5, 6).

We will review and summarize here some of the significant results obtained so far in matter of *quantitative estimation* of interpolation and (more generally) approximation errors. Two wide classes of (multivariate) applications need to be considered in the first instance, namely : (a) *pointwise or uniform approximation in* $C^m(\bar{\Omega})$ and (b) *pointwise and mean-square approximation in* $H^m(\Omega)$. Our approach basically amounts to manipulating (most carefully !) suitable *explicit representations* of any given function f as a polynomial Pf (where P denotes a linear projector) plus a remainder term (expressed in integral form, in terms of generalized partial derivatives of f), in order to get *key estimates* of the general form

$$|f-Pf|_I \leq C|f|_{II},$$

where $|.|_I$ and $|.|_{II}$ denote seminorms involving appropriately related subsets of generalized partial derivatives of f and C, the so-called *error coefficient*, is a numerical constant depending on Ω. Needless to

say, such inequalities are formally related to *variations of the well known Bramble-Hilbert lemma* (see e.g. [4,5]) and could be proved quite similarly, that is in a strictly nonconstructive way, by making use of Morrey-like results (see [23], pp. 85-86) and generalizations thereof. Our proof, which is somewhat reminiscent of Sobolev's approach to imbedding theorems (see e.g. [28], pp. 50-56), is completely different; essentially constructive throughout, it can yield directly (i.e., without passing through the analysis of "change of scale" effects) realistic upper bounds for the best possible error coefficient C_o; of course, these results depend on the specific representation formula at hand : the (truncated) *Taylor series* (see e.g. [13, 14, 22]), the *Kowalewski-Ciarlet-Wagschal formula* (see e.g. [1,8]), the *averaged Taylor series* [14, 15] and more generally the *Sobolev representation formula* (see [28] and [9,10]), etc. It should be noted that a systematic, constructive approach to the important topic of (multivariate) *representation formulas* can be found in the recent papers [16,17,18,19,20,21] whose main concern is different from the present one, however; it is shown there, in particular, how easily *convolutions* (or *Fourier transforms*) can be used, together with *basic integral inequalities* (see e.g. [3] and [12]) to find in a unified way "appropriate" expressions of distributions (or functions) in terms of a prescribed subset of generalized partial derivatives; such representation formulas can provide appropriate substitutes for the truncated Taylor series and for its integral remainder term, such as requested for extending the Peano kernel theorem to the representation of errors in intrinsically multivariate situations.

We will also mention here the somewhat surprising fact that *optimal error coefficients* C_o can often be characterized (see e.g. [20], where this interesting possibility is illustrated in connection with two famous integral inequalities in Sobolev spaces due to Friedrichs and to Poincaré). It occasionally happens that such *characterizations* can be treated further, to yield bounds (see e.g. [2], [25], [26] and specially [27]) but there are apparently simple situations (see e.g. [24]) for which such characterizations are uninteresting. As regards *lower bounds* for C_o, they can usually be obtained more economically *than upper bounds* (see e.g. [13] for the class (a), [24] for the class (b), and [29] for more general techniques).

By way of *concrete applications* of the above material, we will consider two specific, non-trivial examples, namely *multivariate numerical integration with nonnegative weights* and *bivariate Lagrange interpolation over a triangulated domain.*

References

[1] Arcangéli, R. et Gout, J.L. : *Sur l'évaluation de l'erreur d'interpolation de Lagrange dans un ouvert de* $\mathbb{R}^n$, R.A.I.R.O. Analyse numérique *10*, 5-27(1976).

[2] Bandle, C. : *Isoperimetric Inequalities and Applications.* London : Pitman Publishing 1980.

[3] Besov, O.V., Il'in, V.P. and Nikol'skii, S.M. : *Integral Representations of Functions and Imbedding Theorems*, Vol. 1. Washington, D.C. : V.H. Winston and Sons, 1978.

[4] Bramble, J.H. and Hilbert, S.R. : *Estimation of Linear Functionals on Sobolev Spaces with Application to Fourier Tranforms and Spline Interpolation*, SIAM J. Numer. Anal. *7*, 112-124(1970).

[5] Bramble, J.H. and Hilbert, S.R. : *Bounds for a Class of Linear Functionals with Applications to Hermite Interpolation*, Numer. Math. *16*, 362-369(1971).

[6] Ciarlet, P.G. : *The Finite Element Method for Elliptic Problems.* Amsterdam : North-Holland Publishing Company 1978.

[7] Ciarlet, P.G. and Raviart, P.A. : *General Lagrange and Hermite Interpolation in* $\mathbb{R}^n$ *with Applications to Finite Element Methods*, Arch. Rational Mech. Anal. *46*, 177-199(1972).

[8] Ciarlet, P.G. and Wagschal, C. : *Multipoint Taylor Formulas and Applications to the Finite Element Method*, Num. Math. *17*, 84-100(1971).

[9] Dupont, T. and Scott, R. : *Constructive Polynomial Approximation in Sobolev Spaces.* Recent Advances in Numerical Analysis (C. de Boor and G. Golub eds.), 31-44. New York : Academic Press 1978.

[10] Dupont, T. and Scott, R. : *Polynomial Approximation of Functions in Sobolev Spaces.* Report N° 79-13, ICASE, NASA Langley Research Center, Hampton (1979).

[11] Gout, J.L. : *Estimation de l'erreur d'interpolation d'Hermite dans* $\mathbb{R}^n$, Num. Math. *28*, 407-429(1977).

[12] Hardy, G.H., Littlewood, J.E. and Pólya, G. : *Inequalities.* London : Cambridge University Press 1934.

[13] Meinguet, J. : *Realistic Estimates for Generic Constants in Multivariate Pointwise Approximation.* Topics in Numerical Analysis II (J.J.H. Miller ed.), 89-107. London : Academic Press 1975.

[14] Meinguet, J. : *Structure et estimations de coefficients d'erreurs,* R.A.I.R.O. Analyse Numérique *11*, 355-368(1977).

[15] Meinguet, J. : *A Practical Method for Estimating Approximation Errors in Sobolev Spaces.* Multivariate Approximation (D.C. Handscomb ed.), 169-187. London : Academic Press 1978.

[16] Meinguet, J. : *An Intrinsic Approach to Multivariate Spline Interpolation at Arbitrary Points.* Polynomial and Spline Approximation (B.N. Sahney ed.), 163-190. Dordrecht : D. Reidel Publishing Company 1979.

[17] Meinguet, J. : *Multivariate Interpolation at Arbitrary Points Made Simple,* Journal of Applied Mathematics and Physics (ZAMP) *30*, 292-304(1979).

[18] Meinguet, J. : *A Convolution Approach to Multivariate Representation Formulas.* Multivariate Approximation Theory (W. Schempp and K. Zeller eds.), 198-210. Basel : Birkhäuser Verlag 1979.

[19] Meinguet, J. : *Basic Mathematical Aspects of Surface Spline Interpolation.* Numerische Integration (G. Hämmerlin ed.), 211-220. Basel : Birkhäuser Verlag 1979.

[20] Meinguet, J. : *From Dirac Distributions to Multivariate Representation Formulas.* Approximation Theory and Applications (Z. Ziegler ed.), 225-248. New York : Academic Press 1981.

[21] Meinguet, J. : *On Surface Spline Interpolation : Basic Theory and Computational Aspects.* Report N° 81-05, Seminar für Angewandte Mathematik ETH, Zürich (1981).

[22] Meinguet, J. and Descloux, J. : *An Operator-Theoretical Approach to Error Estimation,* Numer. Math. *27,* 307-326(1977).

[23] Morrey, C.B. : *Multiple Integrals in the Calculus of Variations.* New York : Springer-Verlag 1966.

[24] Natterer, F. : *Berechenbare Fehlerschranken für die Methode der Finiten Elemente.* International Series of Numerical Mathematics *28,* 109- 121. Basel : Birkhäuser Verlag 1975.

[25] Payne, L.E. : *Isoperimetric Inequalities and their Applications,* SIAM Review *9,* 453-488(1967).

[26] Pólya, G. and Szegö, G. : *Isoperimetric Inequalities in Mathematical Physics.* Princeton : Princeton University Press 1951.

[27] Sigillito, V.G. : *Explicit "a priori" Inequalities with Applications to Boundary Value Problems.* London : Pitman Publishing 1977.

[28] Sobolev, S.L. : *Applications of Functional Analysis in Mathematical Physics.* Providence : American Mathematical Society 1963.

[29] Weinberger, H.F. : *Variational Methods for Eigenvalue Approximation.* Philadelphia : Society for Industrial and Applied Mathematics 1974.

OPTIMAL FINITE-ELEMENT METHODS FOR DIFFUSION-CONVECTION PROBLEMS

K.W. Morton (University of Reading, England)

Abstract

The success of finite element methods with elliptic problems $Lu = f$ stems largely from their optimality properties in the self-adjoint case. Of all members of a trial space S^h, the approximation obtained through the Galerkin formulation is the closest in the energy norm to the true solution. Several valuable consequences follow from this fact: the coarsest possible mesh may be used with confidence; superconvergence phenomena allow more accurate and detailed information about the solution to be recovered; and localised mesh refinement can be given a sound basis.

However lack of self-adjointness in the operator L erodes these properties and the Galerkin approximation eventually becomes a very poor one. An important class of practical problems of this type are the diffusion-convection problems governed by

$$\underline{\nabla} \cdot (a\underline{\nabla}u - \underline{b}u) + f = 0,$$

where a is a diffusion coefficient and $\underline{b}$ a given convection velocity field. Sharp boundary layers may occur when b/a is large and result is violently oscillatory approximations.

These problems and developments in finite element methods to solve them successfully will be the subject of the two lectures. Similar difficulties arising from oscillatory solutions were encountered with centred difference approximations, and upwind schemes of various kinds introduced to overcome them. In the first lecture we will consider some of this background and the development of similarly designated upwind finite element schemes. These are based on the use of the Petrov-Galerkin method in which a space of test function T^h, distinct from the trial space S^h, is introduced. Another approach, involving the intro-

duction of an artificial diffusion tensor, will also be described. In both cases the methods can be related to exponential fitting methods used with finite differences and also to seeking optimality with respect to the diffusion operator norm.

In the second lecture we will concentrate on an alternative approach based on approximately symmetrizing the problem. This introduces an energy norm of the form

$$\int \{\rho a^2 |\underline{\nabla} v|^2 + [\rho |\underline{b}|^2 + \underline{\nabla} \cdot (\rho a \underline{b})] v^2\} d\Omega,$$

where ρ is a weighting function, with respect to which a near optimal approximation is constructed. For problems in one dimension rigorous error estimates can be derived to demonstrate the near optimality. In two dimensions a formulation will be presented in which solution of a convection problem for the flux $\underline{v} = \underline{b}u - a\underline{\nabla}u$ is alternated with a self-adjoint problem for u.

References

ALLEN, D. & SOUTHWELL, R. 1955 Relaxation methods applied to determine the motion, in two dimensions, of a viscous fluid past a fixed cylinder. Quart. J. Mech. and Appl. Math. VIII, 129-145.

BABUSKA, I. & AZIZ, A.K. 1972 Survey lectures on the mathematical foundations of the finite element method: The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations (ed. A.K. Aziz), New York, Academic Press, 3-363.

BABUSKA, I. & RHEINBOLDT, W.C. 1978 Error estimates for adaptive finite element computation. SIAM J. Numer. Anal. 15, 736-754.

BARRETT, J.W. & MORTON, K.W. 1980 Optimal finite element solutions to diffusion-convection problems in one dimension. Int. J. Num. Meth. Engng. 15, 1457-1474.

BARRETT, J.W. & MORTON, K.W. 1981 Optimal Petrov-Galerkin methods through approximate symmetrization. Submitted to IMA J. Num. Anal.

BARRETT, J.W. & MORTON, K.W. 1981 Optimal finite element approximation for diffusion-convection problems. Proc. MAFELAP 1981 Conf. (ed. J.R. Whiteman).

BRAMBLE, J.H. & SCHATZ, A.H. 1977 Higher order local accuracy by averaging in the finite element method. Math. Comp. 31, 94-111.

BROOKS, A. & HUGHES, T.J.R. 1980 Streamline-upwind/Petrov-Galerkin methods for advection dominated flows. Proc. 3rd Int. Conf. Finite Element Methods in Fluid Flow, Banff, Canada.

CHRISTIE, I., GRIFFITHS, D.F., MITCHELL, A.R. & ZIENKIEWICZ, O.C. 1978 Finite element methods for second order differential equations with significant first derivatives. Int. J. Num. Meth. Engng. 10, 1389-1396.

DOOLAN, E.P., MILLER, J.J.H. & SCHILDERS, W.H.A. 1980 Uniform Numerical Methods for Problems with Initial and Boundary Layers. Dublin: Boole Press.

DOUGLAS, J., Jr. & DUPONT, T. 1973 Some superconvergence results for Galerkin methods for the approximate solution of two-point boundary problems. Proc. Conf. on Topics in Numerical Analysis (ed. J.J.H. Miller), Dublin: Royal Irish Academy, 89-92.

DOUGLAS, J., Jr. & DUPONT, T. 1973 Superconvergence for Galerkin methods for the two-point boundary problem via local projections. Num. Math. 21, 270-278.

GRESHO, P.M. & LEE, R.L. 1979 Don't suppress the wiggles - they're telling you something: in AMD Vol. 34, Finite Element Methods for Convection Dominated Flow (ed. T.J.R. Hughes), ASME, New York.

GUYMON, G.L., SCOTT, V.H. & HERRMANN, L.R. 1970 A general numerical solution of the two-dimensional diffusion-convection equation by the finite element method. Water Resources 6, 1611-1617.

HEINRICH, J.C., HUYAKORN, P.S., MITCHELL, A.R. & ZIENKIEWICZ, O.C. 1977 An upwind finite element scheme for two-dimensional convective transport equations. Int. J. Num. Meth. Engng. 11, 131-143.

HEINRICH, J.C. & ZIENKIEWICZ, O.C. 1979 The finite element method and 'upwinding' techniques in the numerical solution of convection dominated flow problems: in Finite Element Methods for Convection Dominated Flows (ed. T.J.R. Hughes) AMD Vol. 34, Am. Soc. Mech. Eng. (ASME), 105-136.

HEMKER, P.W. 1977 A numerical study of stiff two-point boundary problems. Thesis, Amsterdam: Math. Cent.

HUGHES, T.J.R. & BROOKS, A.N. 1979 A multidimensional upwind scheme with no crosswind diffusion: in Finite Element Methods for Convection Dominated Flows (ed. T.J.R. Hughes) AMD Vol. 34, Am. Soc. Mech. Eng. (ASME), 19-35.

IL'IN, A.M. 1969 Differencing scheme for a differential equation with a small parameter affecting the highest derivative. Math. Notes Acad. Sci. U.S.S.R. 6, 596-602.

JOHNSON, C. & NÄVERT, U. 1980 Analysis of some finite element methods for advection-diffusion problems. Res. Rep. 80.01R, Dept. Comp. Sci., Chalmers Univ. Göteborg.

LESAINT, P. & ZLÁMAL, M. 1979 Superconvergence of the gradient of finite element solutions. R.I.A.R.O. Analyse numérique 13 No. 2, 139-166.

MORTON, K.W. & BARRETT, J.W. 1980 Optimal finite element methods for diffusion-convection problems. Proc. Conf. Boundary and Interior Layers: Computational and Asymptotic Methods (ed. J.J.H. Miller) Dublin: Boole Press, 134-148.

FINITE ELEMENT APPROXIMATION OF NONLINEAR PROBLEMS

P.A. Raviart, CNRS-Université Pierre en Marie Curie, Parijs.

A large class of boundary-value problems for nonlinear elliptic partial differential equations may be put into the following abstract framework. We are given three Banach spaces V, W and X such that

$$W \subset X \quad \text{with compact imbedding.}$$

We introduce a linear operator $T \in \mathcal{L}(X;V)$ and C^p mapping $(p \geq 2)$ $G : (\lambda,u) \in R \times V \to G(\lambda,u) \in X$ such that

$$D_u G(\lambda,u) \in \mathcal{L}(V;W)$$

We set :

$$F(\lambda,u) = u + TG(\lambda,u).$$

Thus F is a C^p mapping from $R \times V$ into V such that $D_u F(\lambda,u) = I + T\, D_u G(\lambda,u)$ is a compact perturbation of the identity.

Now, we consider the nonlinear problem : Given λ in R, find $u \in V$ solution of

$$(1) \qquad F(\lambda,u) = 0.$$

In order to be more specific, let us show for instance that the Navier-Stokes equations for an incompressible viscous flow fit into the above framework

$$(2) \quad \begin{cases} -\frac{1}{R}\Delta \underline{u} + (\underline{u}\cdot\nabla)\underline{u} + \nabla p = f \quad \text{in} \quad \Omega\,, \\ \nabla\cdot\underline{u} = 0 \quad \text{in} \quad \Omega, \\ \underline{u} = \underline{0} \quad \text{on} \quad \Gamma\,, \end{cases}$$

where Ω is a bounded domain of $R^N (N = 2,3)$ with boundary Γ and R stands for the Reynolds number. We set in that case :

$$V = H_o^1(\Omega)^N \times L_o^2(\Omega) \quad , \quad X = H^{-1}(\Omega)^N \ ,$$

where

$$H_o^1(\Omega) = \{ \varphi \in L^2(\Omega) \ ; \ \frac{\partial \varphi}{\partial x_i} \in L^2(\Omega) \quad , \quad 1 \leqslant i \leqslant N \ , \ \varphi_{|\Gamma} = 0 \ \},$$

$$H^{-1}(\Omega) = \text{dual space of } H_o^1(\Omega) \ ,$$

$$L_o^2(\Omega) = \{\varphi \in L^2(\Omega) \ , \ \int_\Omega \varphi \, dx = 0 \ \} \ .$$

Next, we define the linear operator T as follows : with any $\underline{g} \in H^{-1}(\Omega)^N$, we associate the solution $(\underline{v},q) = T\underline{g} \in H_o^1(\Omega)^N \times L_o^2(\Omega)$ of the Stokes problem :

$$(3) \quad \begin{cases} -\Delta \underline{v} + \nabla q = g \quad \text{in} \quad \Omega, \\ \nabla \cdot \underline{v} = 0 \quad \text{in} \quad \Omega, \\ \underline{v} = \underline{0} \ \text{on} \ \Gamma \ . \end{cases}$$

Finally, G is the C^∞ mapping

$$(\lambda, u = (\underline{u},p)) \in R \times (H_o^1(\Omega)^N \times L_o^2(\Omega)) \to G(\lambda,u) = \lambda\{(\underline{u}\cdot\nabla)\underline{u} - f\} \in H^{-1}(\Omega)^N.$$

Note that for $v = (\underline{v},q) \in V$, we have

$$D_u G(\lambda,u) = \lambda\{(\underline{u}\cdot\nabla)\underline{v} + (\underline{v}\cdot\nabla)\underline{u}\}$$

so that $D_u G(\lambda,u) \in \mathcal{L}(V,W)$ where $W = L^{4/3}(\Omega)^N$ for instance.

Now, the pair $(\underline{u},p) \in V$ is a solution of the Navier-Stokes equations (2) if and only if $u = (\underline{u},Rp)$ is a solution of (1) where $\lambda = R$.

Note that many other examples of practical importance also fit into this framenwork such as the nonlinear diffusion equations, the Von Karman equations for a plate ...

Let us now describe a general method of approximation of Problem (1). For each value of a discretization parameter $h > 0$, we are given a finite dimensional subspace V_h of the space V and an operator $T_h \in \mathcal{L}(X;V_h)$. We define the C^p mapping $F_h : (\lambda,u) \in R \times V \to F_h(\lambda,u) \in V_h$ by

$$F_h(\lambda,u) = u + T_h G(\lambda,u).$$

Thus, the approximate problem consists in finding $u_h \in V_h$ solution of

$$F_h(\lambda,u_h) = 0. \tag{4}$$

If we go back to the example of the Navier-Stokes equations, we thus associate with each finite element approximation of the Stokes problem (3) a finite element method for the Navier-Stokes equations.

Now, the natural question arises to deduce the existence, uniqueness and convergence of the solutions u_h of Problem (4) from the approximation properties of the operator T_h. We then have to distinguish several cases.

In the first lecture, we shall consider the case of a branch $\{(\lambda,u(\lambda)) ; \lambda \in \Lambda\}$ of <u>nonsingular</u> solutions of (1) where Λ is a compact interval of R. This means that :

$$\text{(5)} \quad \begin{cases} \lambda \to u(\lambda) \text{ is a continuous function from } \Lambda \text{ into } V ; \\ F(\lambda,u(\lambda)) = 0 \qquad \forall\, \lambda \in \Lambda ; \\ D_u F(\lambda,u(\lambda)) \text{ is an isomorphism of } V \text{ for all } \lambda \in \Lambda. \end{cases}$$

Using the implicit function theorem, we find that $\lambda \to u(\lambda)$ is C^p function from Λ into V.

THEOREM. Assume that the p-th derivative D^pG is bounded on all bounded subsets of $\Lambda \times V$ and $\{(\lambda,u(\lambda)) ; \lambda \in \Lambda\}$ is a branch of nonsingular solution of (1). Assume in addition that

$$(6) \qquad \lim_{h\to 0} \|(T_h - T)f\|_V = 0 \qquad \forall f \in X.$$

Then, there exists a neighborhood $\mathcal{O}$ of the origin in V and for $h \leq h_o$ small enough, a unique C^p function $\lambda \in \Lambda \to u_h(\lambda) \in V_h$ such that for all $\lambda \in \Lambda$

$$F_h(\lambda,u_h(\lambda)) = 0 \quad , \quad u_h(\lambda) - u(\lambda) \in \mathcal{O}$$

Furthermore, we have for all integer m with $0 \leq m \leq p-1$

$$\left\|\frac{d^m}{d\lambda^m}(u_h(\lambda)-u(\lambda))\right\|_V \leq C_m \sum_{\ell=0}^{m} \left\|(T_h-T)\frac{d^\ell}{d\lambda^\ell} G(\lambda,u(\lambda))\right\|_V$$

where C_m is a constant independant of h and λ.

This result enables us to prove that in a number of finite element approximation of nonlinear problems, we obtain the same orders of convergence as in the linear case.

In the second lecture, we shall be concerned with the approximation of the solution of (1) in the neighborhood of a simple singular-point (λ_o,u_o) of F, i.e. a point $(\lambda_o,u_o) \in R \times V$ which satisfies the following properties :

$F(\lambda_o,u_o) = 0$;

$D_uF(\lambda_o,u_o) = I + TD_u G(\lambda_o,u_o)$ is singular and -1 is an eigenvalue of the compact operator $TD_u G(\lambda_o,u_o)$ with algebraic multiplicity 1.

We shall considere in particular the case of a simple bifurcation point for which the solutions of (1) consist exactly of two branches which intersect transversally at (λ_o,u_o). We shall discuss the form of the solutions of the approximate problem in the neighborhood of (λ_o,u_o) and give practical sufficient conditions ensuring that Problem (4) does have a bifurcation point near (λ_o,u_o).

REFERENCES

For standard mathematical results on the finite element method, see :

P.G. CIARLET, The finite element method for elliptic problems, North-Holland (1978) .

For the finite element approximation of the Stokes and Navier-Stokes equations, see :

V. GIRAULT & P.A. RAVIART, Finite element approximation of the Navier-Stokes equations, Lecture notes in Mathematics # 749, Springer Verlag (1979).

Concerning the approximation of non-linear problems, see :

F. BREZZI, J. RAPPAZ, P.A. RAVIART, Finite dimensional approximation of non linear problems.

Part I : Branches of nonsingular solutions, Num. Math. 36, 1-25 (1980)

Part II : Limit points, Num. Math. 37, 1-28 (1981)

Part III : Simple bifurcation points (to appear very soon in Num. Math.)

See also :

H.B. KELLER, Approximation methods for nonlinear problems with applications to two-point boundary value problems, Math. Comp., 29, 464-474 (1975)

and

H. FUJII & M. YAMAGUTI, Structure of singularities and its numerical realization in nonlinear elasticity, J. Math. Kyoto Univ. 20, 489-590 (1980).

INTRODUCTION TO A-STABILITY THEORY FOR STIFF DIFFERENTIAL EQUATIONS

G. Wanner, Université de Genève, Zwitserland

First Lecture.

1. The Beginnings.

"Stiff" differential equations are ordinary differential equations

$$(1.1) \qquad y' = f(y(t)), \qquad y(t_o)=y_o$$

which produce very bad numerical results for very many numerical methods. Such equationsarise very frequently in the applications, e.g. in network analysis, chemical kinetics, control theory, initial value problems in parabolic and hyperbolic equations after finite element or finite difference discretizations in space dimensions etc. An example of such a bad behavior is presented in Fig. 1.

This phenomenon has been discovered around 1951 independently by G. Dahlquist /1/, H.Rutishauser and Curtiss-Hirschfelder. In order to explain it, we analyze the effect of the method (the "explicit mid-point rule", or "leap-frog method")

$$(1.2) \qquad y_{n+1} = y_{n-1} + 2h\, f(y_n)$$

to small perturbations of the numerical solution. Let therefore

$$(1.2') \qquad z_{n+1} = z_{n-1} + 2h\, f(z_n)$$

be a second sequence of (perturbed) computed solutions (Fig.2). We are interested in the development of the errors $w_n=y_n-z_n$. If we substract (1.2) and (1.2') we obtain

$$(1.3) \qquad w_{n+1} = w_{n-1} + 2h\, f'(y_n).w_n + \ldots \quad .$$

Denoting now $hf'(y_n)=z$, this relation is approximated by

$$(1.4) \qquad w_{n+1} = w_{n-1} + 2z.w_n \ .$$

Difference equations of this type are solved by putting $w_n=R^n$ which leads to

$$(1.5) \qquad R^2 = 1 + 2z.R \ .$$

Our method will produce stable behavior, if both roots of this

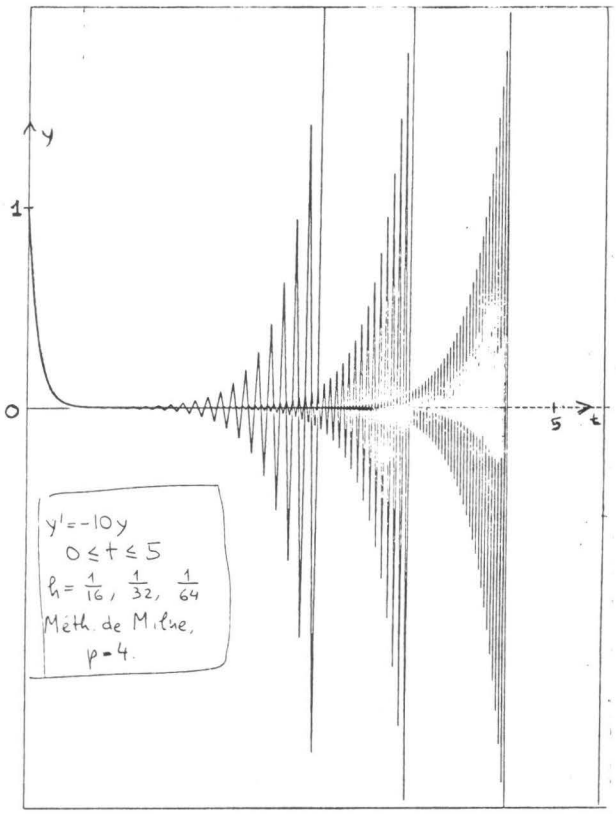

Fig. 1.

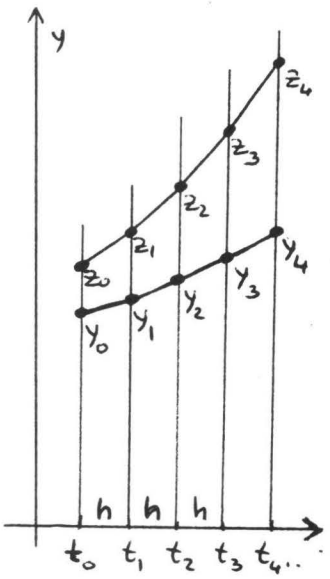

Fig. 2.

"characteristic equation" (1.5) are smaller than one in modulus. We call the set of all admissible z-values

$$S = \{ z \mid \text{all roots of the charact. equation are} < 1 \}$$

the <u>stability domain</u> of the method (Dahlquist Nov. 1963).

As can be seen from (1.5) (the product of the two roots is -1 and their sum is 2z), the stability domain for this method is a small interval on the imaginary axis, sketched in Fig. 3. Since for the example of Fig. 1 the values of $z=hf'(y_n)$ are -10/16, -10/32, and -10/64, ly outside of S, we have explained the bad behavior of the method.

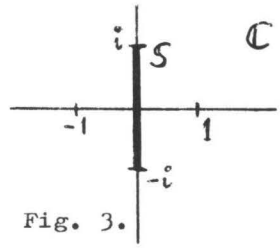

Fig. 3.

All this analysis is, of course, also valid in the case of a <u>system</u> of equations. In this case, y_n, $f(y_n)$, z_n, w_n represent vectors and $f'(y_n)$ in (1.3) is the Jacobian matrix of all partial derivatives. It is then transformed to diagonal form and we are led to (1.4) where $z=h\lambda$ and λ represents the <u>eigenvalues</u> of $f'(y_n)$.

Let us now examine a simple explicit Runge Kutta method such as

$$(1.6)\qquad \begin{cases} g_n = y_n + \frac{h}{2} f(y_n) \\ y_{n+1} = y_n + h\, f(g_n) \end{cases} .$$

The same analysis as above gives us in this case the following:

$$(1.7)\qquad w_{n+1} = R(z)w_n \qquad \text{where} \qquad R(z) = 1 + z + \frac{z^2}{2}.$$

The stability domain for this method

$$S = \left\{ z \;\middle|\; |R(z)| \leq 1 \right\}$$

is plotted in Fig. 3a.
We see that this method produces stable results for the problem of Fig. 1, however only if $h < \frac{2}{10}$.
On the other hand, if we apply

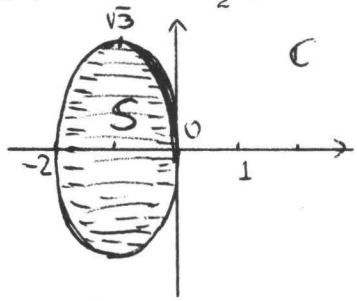

Fig. 3a.

method (1.6) to an oscillatory equation of the type $y'' + k^2 y = 0$, we are led to a linear system with eigenvalues $\pm ik$, so that <u>here</u> method (1.6) is always unstable and method (1.2) is stable for $h < 1/k$.

Let us finally look at the BDFormula of order 2 ("method of Gear")

$$(1.10)\qquad y_{n+2} - \frac{4}{3} y_{n+1} + \frac{1}{3} y_n = \frac{2}{3} h\, f(y_{n+2}).$$

This is an implicit equation for the computation of y_{n+2} and has the characteristic equation

(1.11) $\qquad (1 - \frac{2}{3} z)R^2 - \frac{4}{3} R + \frac{1}{3} = 0 .$

The modules of the two roots $R_1(z)$ and $R_2(z)$ of this equation are plotted in Fig. 4. The stability domain of this method, i.e., the set of all z for which both roots ly below sea level $|R|=1$, is shown in Fig. 5. It can be seen, that there is no more problem of stability, as long as z remains in the negative half plane. Methods, which satisfy this property are called A-stable. (Dahlquist 1963).

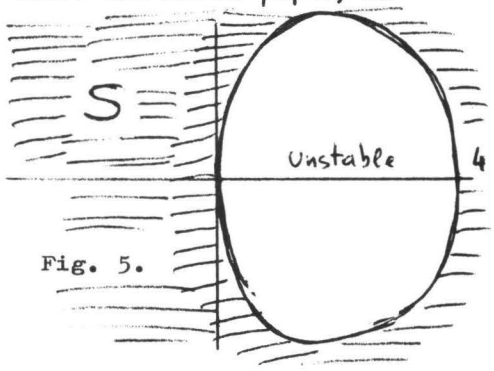

Fig. 5.

2. The Order Star Theory

We now look again to Fig. 4. We see that the two roots possess a branching point at z=-1/2. If this point is surrounded, the two roots are interchanged. We thus consider naturally R(z) as beeing defined on a Riemann surface M having the same property. Further, if z approaches the point 3/2, the leading term $(1 - \frac{2}{3} z)$ in (1.11) becomes zero, so one of the two roots tends to infinity, while the remaining root converges to the root of the remaining equation $-\frac{4}{3} R + \frac{1}{3} = 0$. So the function R(z) possesses on M exactly one pole. We now consider the set

$$(2.1) \qquad A = \left\{ z \in M \;\middle|\; \left|\frac{R(z)}{e^z}\right| > 1 \right\} = \left\{ z \in M \;\middle|\; |S(z)| > 1 \right\}$$

where $S(z) = R(z)/e^z$. This set is called order star and is shaded in Fig. 4.

A consequence of the fact that the method (1.10) is of order 2, i.e., that the local error behaves like $C.h^3$, is that $R_1(z)$ approaches the exact solution of y'=zy also of order 2 for $z \to 0$:

$$(2.2) \qquad e^z - R_1(z) = C z^3 + O(z^4) \qquad \text{for } z \to 0.$$

Lemma 1. Whenever for $z \to 0$

$$(2.3) \qquad e^z - R_1(z) = C z^{p+1} + \dots ,$$

the set A is "starlike" in the neighbourhood of z=0 with p+1 fingers. A consists of regularly distributed p+1 sectors of angle $\pi/(p+1)$.

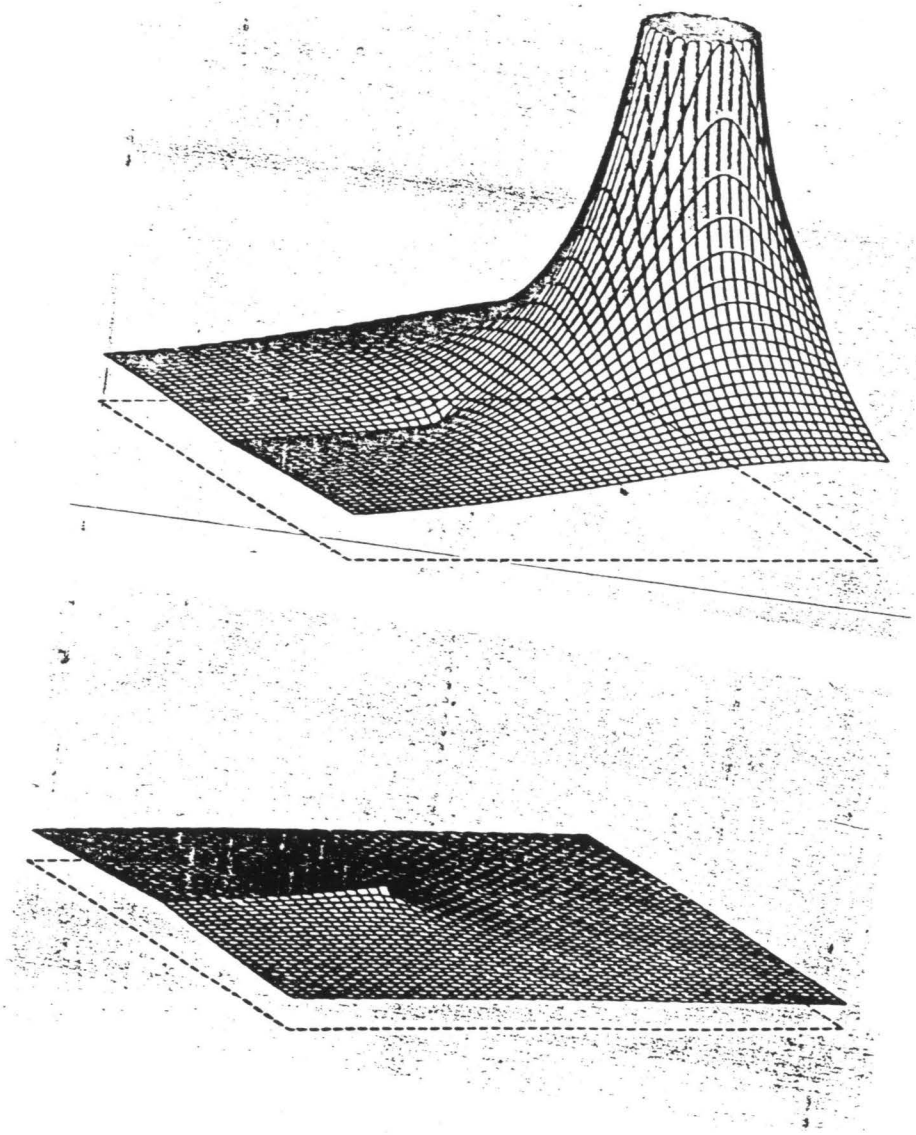

Fig. 4. Stability function and order star for second order backward difference formula.

Example: For p=3, the order star in the neighbourhood of the origin is sketched in Fig. 5.

Fig. 5.

The proof of Lemma 1 is by simple complex analysis. One divides (2.3) by e^z and obtains for small z that

$$z \in A \iff \mathrm{Re}(Cz^{p+1}) < 0.$$

Lemma 2. Let $F \subset A$ be such that $\partial F \subset \partial A$, with F bounded. Then:
F contains m poles of R $\Leftrightarrow$ ∂F contains m points with $R(z)=e^z$.
The same is true if we replace "$A \mapsto \complement A$", and "poles $\mapsto$ zeros".

The proof of this is a consequence of the fact that, along the boundary of A, the arg(S) always turns in the negative sense. The total number of rotations is therefore equal to the points where $\arg(S)=0$, i.e., where $R(z)=e^z$.

Lemma 3. The method is A-stable if and only if
i) all poles of R(z) are in the positive half plane, and
ii) the order star A has no intersection with the imaginary axis.

Proof. This is a consequence of $|e^{iy}|=1$, so that on the imaginary axis the factor e^z in the definition of A can be left away, and the maximum modulus principle.

Theorem ("the second Dahlquist barrier"). If a linear multistep method is A-stable, then its order cannot exceed 2. If p=2, the error constant cannot be smaller than 1/12, which is the error constant of the trapezoidal rule.

Proof. This famous theorem of Dahlquist (1963) is simply obtained by applying Lemma 2 to the set $F = A \cap \mathbb{C}^+$. If p is greater that 2, we need two fingers in $\mathbb{C}^+$, therefore the origin is twice on the boundary of F, we therefore need two poles of R, which is not possible for linear multistep methods. As to the second assertion, one has to replace in the definition of A the exponential function by $(1 + \frac{z}{2})/(1 - \frac{z}{2})$. More about this idea will be presented in the talk of R. Jeltsch.

Theorem 2. Padé Approximations to e^z with q poles and r zeros produce A-stable methods if and only if

$$q \geq r \geq q-2 .$$

This theorem can be seen by looking at the order stars for these approximations, which due to the high order of approximation at the origin, must necessarily have the structure as shown in Fig. 6.

More similar results of this type are easily obtainable by similar reasoning and will be presented in the lecture or can be found in the literature.

References.

1. G. Dahlquist, ZAMM Sonderheft, GAMM-Tagung 1951, Freiburg i. Br.

2. G.Dahlquist, A special stability problem for linear multistep methods, BIT 3 (1963), 27-43.

3. G.Wanner, E.Hairer and S.P.Nørsett, Order stars and Stability theorems, BIT 18 (1978), 475-489, 501.

4. S.P.Nørsett and G.Wanner, The real pole sandwich for rational approximations and oscillation equations, BIT 19 (1979), 79-94.

5. R.Jeltsch and O.Nevanlinna, Stability of explicit time discretizations for solving initial value problems, Num.Math 1981

6. A.Iserles, Order stars and a saturation theorem for first oder hyperbolics, IMA Journal of Num.Anal, to appear 1982.

Introduction to B-stability Theory

Second lecture in Zeist.

For a long time, people have been unsatisfied by the A-stability analysis, due to the neglections made between equations (1.3) and (1.4) above and the hypothesis that $f'(y_n)$ remains constant for the whole problem, until a new break-through has been made by G. Dahlquist with his G-stability theory for one-leg methods /1/.

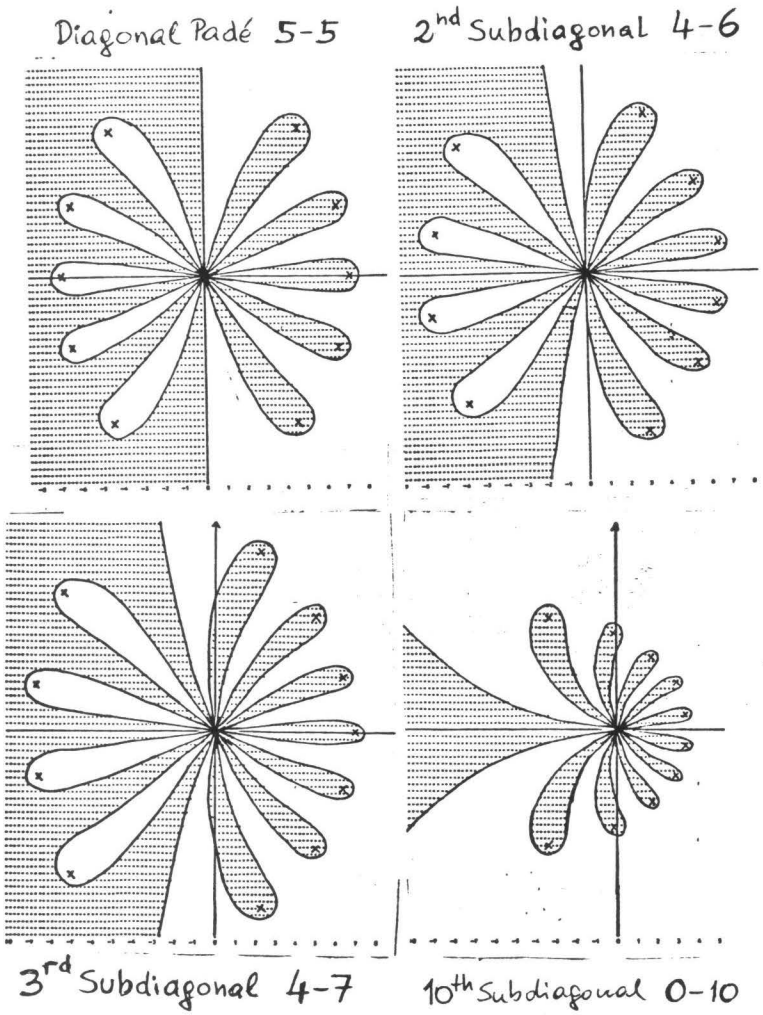

Fig. 6.

Order stars for Padé approximations of the exponential function.

For an arbitrarly nonlinear differential equation (1.1), two theoretical solutions $y(t)$ and $z(t)$ approach in the euclidean norm if and only if

$$\text{Re} \langle f(y)-f(z),\ y-z \rangle \leq 0. \tag{3.1}$$

We call a Runge-Kutta method B-stable, if this condition also implies contractivity

$$\| y_{n+1} - z_{n+1} \| \leq \| y_n - z_n \|$$

for two different numerical solutions /2/.
A very satisfactory theory of this type of stability is possible for implicit Runge-Kutta methods /3/, and all B-stable Runge-Kutta methods can be given explicitely /4,5/.

This theory has also an surprising application to A-stability, since with its motivation all A-stable rational approximations can be obtained /6/.

A big conjecture, stating that "B-stability is equivalent to A-stability", in the sense that every A-stable algebraic approximation can be realized as a B-stable (or G-stable) method, is only partially proved (Dahlquist 1978, Hairer-Wanner /4/, /6/).

References.

1. G.Dahlquist, Error analysis for a class of methods for stiff nonlinear initial value problems, Springer Lecture Notes in Math Nr 506, 60-74.
2. J.C.Butcher, A stability property of implicit Runge-Kutta methods, BIT 15 (1975), 358-361
3. K.Burrage and J.C.Butcher, Stability criteria for implicit Runge-Kutta methods, SIAM J. Numer. Anal 16 (1979), 46-57.
4. E.Hairer and G.Wanner, Algebraically stable and implementable Runge-Kutta methods of high order, SIAM J. Numer anal, to appear 1981
5. G.Sottas and G.Wanner, The number of positive weights of a quadrature formula, to appear 1982
6. E.Hairer, Constructive characterization of A-stable approximations to exp(z), to appear 1982

DEELNEMERS

1.	Asselt, E.J. van	Mathematisch Centrum	Amsterdam
2.	Axelsson, A.O.H.	Katholieke Universiteit	Nijmegen
3.	Bakker, M.	Mathematisch Centrum	Amsterdam
4.	Boerstoel, J.W.	Nat. Lucht-en Ruimtevaart lab.	Amsterdam
5.	Bollen, J.A.M.	Technische Hogeschool	Enschede
6.	Braam, H.	ECN	Petten
7.	Crowet, F.	Philips Research Lab.	Brussel
8.	Cuppen, J.J.M.	Universiteit van Amsterdam	Amsterdam
9.	Dekker, Th.J.	Universiteit van Amsterdam	Amsterdam
10.	Dierieck, C.	Philips Research Lab.	Brussel
11.	Doolan, E.	Kon. Shell. Lab.	Rijswijk
12.	Exter Blokland, A.W. den	KNMI	De Bilt
13.	Geurts, A.J.	Technische Hogeschool	Eindhoven
14.	Ginneken, C.J.J.M.v.	Technische Hogeschool	Eindhoven
15.	Griend, J.A. van de	Rijks Universiteit	Leiden
16.	Groen, P. de	Vrije Universiteit	Brussel
17.	Groot, J. de	Philips	Eindhoven
18.	Hemker, P.W.	Mathematisch Centrum	Amsterdam
19.	Heijer, C. den	Philips	Eindhoven
20.	Hout, R. van der	AKZO	Arnhem
21.	Houwen, P.J. van der	Mathematisch Centrum	Amsterdam
22.	Hundsdorfer, W.H.	Rijks Universiteit	Leiden
23.	Jansen, J.K.M.	Technische Hogeschool	Eindhoven
24.	Jarausch, H.	RWTH	Aken
25.	Jeltsch, R.	RWTH	Aken
26.	Kok, J.	Mathematisch Centrum	Amsterdam
27.	Laan-de Klerk, P.	Techn. Hogeschool Twente	Enschede
28.	Liniger, W.	IBM, Yorktown Heights	USA
29.	Loon, P.M. van	Technische Hogeschool	Eindhoven
30.	Meinguet, J.	Université de Catholique de Louvain,	België
31.	Morton, K.W.	University of Reading	UK
32.	Pfluger, P.	Universiteit van Amsterdam	Amsterdam
33.	Piers, W.J.	Nat. Lucht- en Ruimtevaart lab.	Amsterdam

34.	Raviart, P.A.	CNRS Université Pierre en Marie Curie	Parijs
35.	Riele, H.J.J. te	Mathematisch Centrum	Amsterdam
36.	Schilders, W.H.A.	Philips	Eindhoven
37.	Sluis, A. van der	Rijks Universiteit	Utrecht
38.	Sommeijer, B.P.	Mathematisch Centrum	Amsterdam
39.	Spijker, M.N.	Rijks Universiteit	Leiden
40.	Stelling, G.S.	Rijkswaterstaat	Rijswijk
41.	Traas, C.R.	Technische Hogeschool Twente	Enschede
42.	Veldhuizen, M. van	Vrije Universiteit	Amsterdam
43.	Veldman, A.E.P.	Nat. Lucht-en Ruimtevaart lab.	Amsterdam
44.	Veltkamp, G.W.	Technische Hogeschool	Eindhoven
45.	Verwer, J.G.	Mathematisch Centrum	Amsterdam
46.	Vries, H.B. de	Mathematisch Centrum	Amsterdam
47.	Wanner, G.	Université de Genève	Zwitserland
48.	Weber, C.	Philips	Eindhoven
49.	Wesseling, P.	Technische Hogeschool	Delft
50.	Wetterling, W.W.E.	Technische Hogeschool Twente	Enschede
51.	Wilde, M. de	ECN	Petten
52.	Wolkenfelt, P.H.M.	Technische Hogeschool	Eindhoven
53.	Wuytack, L.	Universiteit Antwerpen	Antwerpen
54.	Zeeuw, P.M. de	Mathematisch Centrum	Amsterdam