CONSTRUCTIVE THEORY OF FUNCTIONS'87

SOFIA, 1988

ERROR ESTIMATES FOR THE NUMERICAL SOLUTION OF O.D.Es. BY A-METHODS USING THE τ-MODULUS

Ewald Quak

Let the function f be bounded and measurable on the interval $[a,b] \subset \mathbb{R}$ and let $k \in \mathbb{N}$.

The local modulus of smoothness of order k for the function f at the point $x \in [a,b]$ is defined as

$$\omega_k(f,x;\delta) := \sup \{ |\Delta_h^k f(t)| : t, t+kh \in [x - \tfrac{k\delta}{2}, x + \tfrac{k\delta}{2}] \},$$

where

$$\Delta_h^k f(t) := \begin{cases} \sum_{l=0}^{k} (-1)^{k-l} \binom{k}{l} f(t+lh) & \text{if } t, t+kh \in [a,b] \\ 0 & \text{else} \end{cases}.$$

The averaged modulus of smoothness or τ-modulus of order k of the function f is given as

$$\tau_k(f;\delta)_p := \| \omega_k(f,\cdot;\delta) \|_p \quad , \quad 1 \le p \le \infty.$$

The usefulness of giving estimates involving the τ-modulus has already been well established for a wide range of problems in approximation theory such as onesided trigonometrical approximation or approximation by positive linear operators, see [7] for a detailed treatment. The proof of the following basic properties can be found in [7].

i) $\tau_k(f;\delta)_p \leq \tau_k(f;\delta')_p$ for $\delta \leq \delta'$;

ii) $\tau_k(f+g;\delta)_p \leq \tau_k(f;\delta)_p + \tau_k(g;\delta)_p$;

iii) $\tau_k(f;\delta)_p \leq \delta\, \tau_{k-1}(f';k\delta/(k-1))_p$;

iv) $\tau_1(f;\delta)_p \leq \delta\, \| f' \|_p$;

v) $\tau_1(f;\delta)_1 \leq 2\delta \overset{b}{\underset{a}{V}} f$ if $\overset{b}{\underset{a}{V}} f < \infty$;

vi) $\tau_k(f;\lambda\delta)_p \leq (2\lambda+2)^{k+1}\, \tau_k(f;\delta)_p$.

Here $\overset{b}{\underset{a}{V}} f$ denotes the variation of a function f in the interval $[a,b]$ and it is assumed in iii) and iv) that f' exists on the interval $[a,b]$ and is bounded and measurable iii) or p-integrable iv).

Giving error estimates for the numerical solution of ordinary differential equations (O.D.Es.) via the τ-modulus is of particular interest as such estimates still hold under weaker assumptions compared with usual results (where often smoothness properties of high order derivatives of the solution are assumed).

In this paper the following initial value problem will be considered:

$$(1)\qquad \begin{aligned} &Y' = f(x,Y) \quad, x \in [x_0, x_0+D],\quad D>0, \\ &Y(x_0) = \eta_0, \end{aligned}$$

the function $f : [x_0, x_0+D] \times \mathbb{R} \to \mathbb{R}$ satisfying a Lipschitz condition

$$(2)\qquad |f(x,y)-f(x,z)| \leq K|y-z| \quad , K \text{ an absolute constant.}$$

Note that in the following Y_i will denote the exact solution at the point $x_i := x_0 + ih$, $h := \frac{D}{n}$, $i = 0,1,\ldots,n$ for $n \in \mathbb{N}$, whereas the numerical solution will be denoted by y_i .

A first result using the τ-modulus was given by Andreev in [3] for first and second order Runge-Kutta methods and for Adams k-step methods. A general result for linear k-step methods

$$(3) \quad L_{j+k-1}(y) = \sum_{i=0}^{k} a_i \, y_{j+i-1} - h \sum_{i=0}^{k} b_i f(x_{j+i-1}, y_{j+i-1}) = 0$$

$$a_k \neq 0 \qquad j=1,\dots,n-k+1$$

was established by Popov and Andreev.

Theorem A [6] Assume that $y_i = Y_i$ for $i=0,1,\dots,k-1$.

Let $L_{j+k-1}(P) = 0$ if P is an algebraic polynomial of degree m.

Let the roots μ_l of the polynomial $\sum_{i=0}^{k} a_i x^i$ satisfy the condition

$$(4) \quad |\mu_l| \leq 1 \text{ for } l=1,\dots,k \text{ and } |\mu_l| < 1 \text{ for multiple roots.}$$

Then for the solution Y of (1) and the values $\{y_i\}_{i=0}^{n}$ obtained from (3) the following estimate holds

$$\max_{0 \leq i \leq n} |Y_i - y_i| \leq C \, \tau_m (Y' ; h)_1 \, .$$

The main drawback in the use of linear k-step methods lies in the existence of the so-called Dahlquist barrier: any explicit (i.e. $b_k=0$) linear k-step method satisfying (4) has at most order of convergence $m=k$, any implicit ($b_k \neq 0$) method at most $m=k+2$ for k even and $m=k+1$ for k odd (see [5], p. 231 f.)

One way to overcome this problem is to apply cyclic methods using not only one but several linear methods (in a fixed order). The aim of this paper is to show that for a wide range of so-called A-methods analogous estimates to the one in Theorem A can be obtained. In the following the basic concepts of A-methods will be illustrated, for a detailed treatment concerning the history, basic definitions and theorems see the survey paper of Albrecht [2].

Consider for example the linear cyclic 3-step method of Donelson and Hansen [4]

$$33y_{3j} + 24y_{3j-1} - 57y_{3j-2} = h(10f_{3j} + 57f_{3j-1} + 24f_{3j-2} - f_{3j-3})$$

$$(5)\quad 720y_{3j+1} - 1347y_{3j} - 456y_{3j-1} + 1083y_{3j-2}$$

$$= h(251f_{3j+1} + 456f_{3j} - 1347f_{3j-1} - 350f_{3j-2})$$

$$33y_{3j+2} + 24y_{3j+1} - 57y_{3j} = h(10f_{3j+2} + 57f_{3j+1} + 24f_{3j} - f_{3j-1})$$

$$y_0 = \eta_0,\ y_1 = \eta_1,\ y_2 = \eta_2 \quad \text{(starting values)}.$$

Regarding the vector $z_j := (y_{3j}, y_{3j+1}, y_{3j+2})^T$ as an approximation for the "discretization vector" $Z_j := (Y_{3j}, Y_{3j+1}, Y_{3j+2})^T$, method (5) can be written as

$$(6)\qquad z_j = Az_{j-1} + h\phi(x^*_{j-1}, z_{j-1}, z_j; h) \quad , \quad z_0 = \zeta(h) \quad ,$$

where $\zeta = (\eta_0, \eta_1, \eta_2)^T$ are the starting values, A a real 3x3-matrix given as $A = L^{-1}U$

$$L = \begin{pmatrix} 33 & 0 & 0 \\ -1347 & 720 & 0 \\ -57 & 24 & 33 \end{pmatrix} \quad , \quad U = \begin{pmatrix} 0 & 57 & -24 \\ 0 & -1083 & 456 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$\phi(x^*_{j-1}, z_{j-1}, z_j; h) = L^{-1} \begin{pmatrix} 10f_{3j} + 57f_{3j-1} + 24f_{3j-2} - f_{3j-3} \\ 251f_{3j+1} + 456f_{3j} - 1347f_{3j-1} - 350f_{3j-2} \\ 10f_{3j+2} + 57f_{3j+1} + 24f_{3j} - f_{3j-1} \end{pmatrix}$$

$$x^*_{j-1} = x_{3j+2} \quad , j=1,\dots,[(n-2)/3] \ .$$

Using vectors z_j representing numerical approximations for "discretization vectors" such as $(Y_{rj}, Y_{rj+1}, \dots, Y_{rj+k-2}, Y_{rj+k-1})^T$ (as above), $(Y_{j+k-1}, Y_{j+k-2}, \dots, Y_j, h\,Y'_{j+k-1})^T$ or $(Y_{j+k-1}, h\,Y'_{j+k-1}, \dots, h\,Y'_{j+1}, h\,Y'_j)^T$ and others a wide range of numerical methods can be expressed in a way similar to (6).

The linear k-step method (3) can be represented - using the "discretization vector"

$(Y_{j+k-1}, Y_{j+k-2}, \dots, Y_j, h\, Y'_{j+k-1})^T$:

$$A = \begin{pmatrix} -a_{k-1} & -a_{k-2} & \dots & -a_1 & -a_0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}$$

$$\phi(x^*_{j-1}, z_{j-1}, z_j; h) = \Big(\sum_{i=0}^{k} b_i\, f_{j+i-1}, 0, \dots, 0,\ f_{j+k-1} \Big)^T .$$

Even Runge-Kutta methods can be expressed in such a way using a special discretization vector (see [2]).

In general, approximations z_j of a discretization vector Z_j are are obtained from

(7) $\quad z_0 = \zeta(h) \quad z_j = A z_{j-1} + h\phi(x^*_{j-1}, z_{j-1}, z_j; h)$,

where $\zeta, z_j \in \mathbb{R}^s$, $A \in \mathbb{R}(s,s)$, $\phi : I^*_h \times \mathbb{R}^s \times \mathbb{R}^s \times (0, h_0] \to \mathbb{R}^s$ ($I^*_h := \{x^*_j = x_{rj+k-1},\ j=0,\dots,[(n-k+1)/r]\}$ $k, r \in \mathbb{N}$ being the grid points belonging to the chosen discretization vector), h_0 such that (7) has a unique solution for all $h \in (0, h_0]$ and

(8) $\quad |\phi(x, u_1, v_1; h) - \phi(x, u_2, v_2; h)| \le K_1 |u_1 - u_2| + K_2 |v_1 - v_2|$,

$K_1, K_2 > 0$ constants.

Now it is important to note that a linear k-step method as a component of an A-method (7), (8) may be unstable (i.e. condition (4) is not fulfilled). Instead, only the matrix A as a whole has to satisfy a "matrix root condition" for its eigenvalues μ_l

(9) $\quad |\mu_l| \le 1 \quad$ for $l = 1, \dots, s$

and eigenvalues of modulus 1 have only linear elementary divisors.

Note that (9) is satisfied iff for any matrix norm $\|A^j\| \leq \bar{K}$ for a constant $\bar{K} > 0$ and all $j \in \mathbb{N}$.

Now the following estimate using the τ-modulus holds:

Theorem Let Y be the exact solution of the problem (1), (2). Let $\{y_i\}_0^n$ be the values obtained from the components of an A-method (7), (8), (9), each component being a linear multistep method $L^{(j)}$ such that $L^{(j)}(P) = 0$ for every polynomial P of degree m_j. Let $m := \min_j m_j$.

Then $\max_{0 \leq i \leq n} |Y_i - y_i| \leq C\,(\tau_m(Y';h)_1 + |d_0|)$,

$d_0 \in \mathbb{R}^s$ being the vector of the starting errors of the method ($|\cdot|$ denoting the maximum norm in $\mathbb{R}^s$).

Proof: The discretization vector Z_j of the exact solution Y is regarded as the result of a perturbed A-method

$$z_j = Az_{j-1} + h\phi(x^*_{j-1}, z_{j-1}, z_j; h)$$

$$Z_j = AZ_{j-1} + h\phi(x^*_{j-1}, Z_{j-1}, Z_j; h) + hd_j \quad ,$$

i.e. the $d_j \in \mathbb{R}^s$ are the local discretization errors of the A-method.

By a general result of Albrecht (Theorem 2.10 in [2], see also [1] for the detailed proof), the following estimate holds

$$|Z_j - z_j| \leq C^* \sup_{0 \leq j \leq n^*} \left| A^j d_0 + h \sum_{l=1}^{j} A^{j-l} d_l \right|$$

where n^* denotes the number of steps to be made ($n^* = O(n)$) and C^* is a constant independent of the stepsize h as will be all other constants occurring in the following.

Condition (9.) implies

$$|A^j d_0| \leq \| A^j \| \; |d_0| \leq \tilde{K} |d_0| \quad \text{and}$$

$$|h \sum_{l=0}^{j} A^{j-l} d_l| \leq \tilde{K} \, h \sum_{l=1}^{n} |d_l| \quad .$$

Andreev's proof for Adams methods in [3] can now be transferred to show that each component multistep method is a bounded linear functional such that

$$|L_l^{(j)}(g)| \leq \| L_l^{(j)} \| \; \tilde{C} \; \omega_{m_j}(g, x^*_{l-1}; C_j h)$$

for every function g that is bounded on the interval $[x_{l-1}-m_jC_jh/2, x_{l-1}+m_jC_jh/2]$ and the norm $\| L_l^{(j)} \|$ is bounded by a suitable constant, independent of h. Combining the component results gives

$$|d_l| \leq C \; \omega_m(Y', x^*_{l-1}; \hat{C} \, h)$$

and altogether

$$|h \sum_{l=1}^{j} A^{j-l} d_l| \leq C \sum_{l=1}^{n^*} \int_{x^*_{l-1}}^{x^*_l} \omega_m(Y', x^*_{l-1}; \hat{C} \, h) dx$$

$$\leq C \sum_{l=1}^{n^*} \int_{x^*_{l-1}}^{x^*_l} \omega_m(Y', x; \bar{C} \, h) dx \overset{vi)}{\leq} C \; \tau_m(Y'; h)_1$$

□

Remarks.

1. If the starting values in Theorem A had not been exact, the term $|d_0|$ would also have occurred in the statement of Theorem A.
2. Presuming that the starting values are calculated by a method that also allows a suitable τ-modulus estimate, the result of the theorem would be

$$\max_{0\leq i\leq n} |Y_i - y_i| \leq C\ \tau_m(Y';h)_1 \quad .$$

Example.

To allow a further insight of the concepts presented here, consider again Donelson and Hansen's method (5).

For each component holds $m_j = 5$. Being implicit 3-step methods, the components do not satisfy condition (4) because of the Dahlquist barrier. Treating (5) in the A-method form (6) implies $m = 5$ and $\mu_1 = 1$, $\mu_2 = \mu_3 = 0$ for the eigenvalues of A, i.e. condition (9) holds.

Taking into account Remark 2, the result of the theorem for the method (5), (6) becomes

$$\max_{0\leq i\leq n} |Y_i - y_i| \leq C\ \tau_5(Y';h)_1 \quad .$$

Therefore the properties iii), v), vi) of the τ-modulus allow the statement

$$\max_{0\leq i\leq n} |Y_i - y_i| = O\ (h^5)$$

for a function Y such that $\bigvee_{x_0}^{x_0+D} Y^{(5)} < \infty$, not only for a function $Y \in C^6\ [x_0, x_0+D]$ as the classical results suggest, thus illustrating the significance of a τ-modulus estimate for numerical methods as already mentioned before.

REFERENCES

1. P. Albrecht. Die numerische Behandlung gewöhnlicher Differentialgleichungen. Hanser Verlag, München 1979.
2. P. Albrecht. Numerical Treatment of O.D.Es.: The Theory of A-Methods. Numer. Math., 47, 1985, 59-87.
3. A.S. Andreev. Error estimates for numerical solutions of Cauchy-problem. SERDICA, 10, 1984, 206-215.
4. J. Donelson, E. Hansen. Cyclic composite multistep predictor-corrector methods. SIAM J. Numer. Anal., 8, 1971, 137-157.
5. P. Henrici. Discrete variable methods in ordinary differential equations. Wiley, New York 1962.
6. V.A. Popov, A.S. Andreev. On the error estimation in numerical methods. Computational Mathematics, Banach Center Publications, 13, Warsaw 1984, 647-658.
7. B. Sendov, V.A. Popov. Averaged moduli of smoothness. Bulgarian Mathematical Monograpns, 4, Bulgarian Academy of Sciences, Sofia 1983 (in Bulgarian).

Universität Dortmund

Fachbereich Mathematik

Postfach 500 500

D-4600 Dortmund 50

Bundesrepublik Deutschland