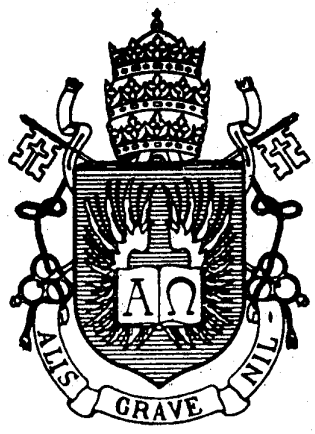

PUC

Series: Monografias em Ciência da Computação

Nº 12/83

NUMERICAL TREATMENT OF ORDINARY DIFFERENTIAL EQUATIONS : THE THEORY OF A-METHODS

by

Peter Albrecht

Departamento de Informática

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO
RUA MARQUÊS DE SÃO VICENTE, 225 – CEP-22453
RIO DE JANEIRO – BRASIL

PUC/RJ - DEPARTAMENTO DE INFORMÁTICA

Series: Monografias em Ciência da Computação
Nº 12/83

Series Editor: Antonio L. Furtado June 1983

NUMERICAL TREATMENT OF ORDINARY DIFFERENTIAL EQUATIONS: THE THEORY OF A-METHODS*

by

Peter Albrecht

Departamento de Informática

* Research partly sponsored by FINEP

ABSTRACT

Almost all commonly used methods for O.D.Es. and their most miscellaneous compositions are A-methods, i.e. they can be reduced to $z_0=\zeta$; $z_j=Az_{j-1}+h\phi(x_{j-1},z_j;h)$, $z_j \in \mathbb{R}^s$, $A \in \mathbb{R}(s,s)$, $j=1(1)m$. This paper presents a general theory for A-methods and discusses its practical consequences. An analysis of local discretization error (l.d.e.) accumulation results in a general order criterium and reveals which part of the l.d.e. effectively influences the global error. This facilitates the comparison of nethods and generalizes considerably the conpet of error constants. It is shown, as a consequence, that the global error cannot be safely controlled by the size of the l.d.e. and that the conventional error control may fail in important cases. Furthermore, Butcher's "effective order" methods, the conept of of Nordsieck forms, and Gear's interpretation of linear k-step schemes as relaxation methods are generalized.
The stability of step changing is shortly discussed.

KEYWORDS

Differential equations, order criterium, error control, error constant, stability, step changing.

RESUMO:

Quase todos os métodos comumente usados para E.D.Os. e suas mais variadas composições são métodos-A, isto é, podem ser reduzidos a $z_o = \zeta$; $z_j = Az_{j-1} + h\phi(x_{j-1}, z_{j-1}, z_j; h)$, $z_j \in \mathbb{R}^s$, $A \in \mathbb{R}(s,s)$, $j = 1(1)m$. Este trabalho apresenta uma teoria geral para métodos-A e discute suas consequências práticas. Uma análise da acumulação do erro de discretização local (l.d.e.) resulta em um critério de ordem geral que parte do l.d.e. realmente afeta o erro global. Isto facilita a comparação de métodos e generaliza consideravelmente o conceito de constante de erro. É mostrado, como uma consequência, que o erro global não pode ser seguramente controlado pelo módulo do l.d.e. e que o controle conveniconal de erro pode falhar em casos importantes. São generalizados os métodos "of effective order" de Butcher, o conceito das formas de Nordsieck e a interpretação de Gear dos métodos lineares de passo k como métodos de relaxação. A estabilidade da mudança de passo é brevemente discutida.

PALAVRAS CHAVE:

Equações diferenciais, critério de ordem, controle de erro , constante de erro, estabilidade, mudança de passo.

1. Introduction

This paper deals with methods for O.D.Es. that can be represented in the form $z_j = Az_{j-1} + h\phi(x_{j-1}, z_{j-1}, z_j; h), z_j \in \mathbb{R}^s, A \in \mathbb{R}(s,s), j = 1(1)m_h$; (1.1) they are called "A-methods" as they essentially depend upon the matrix A. Most of the currently used methods (and many others) can be reduced to this form in a variety of ways.

A special case of (1.1) has first been considered by Butcher [4], who showed that (1.1) is stable iff $\forall j \in \mathbb{N} : ||A^j|| \leq$ const. The resulting restrictions on the eigenvalues of A generalize Dahlquist's root condition.

As $\rho(A_1) > 1$ and $\rho(A_2) > 1$ may imply $\rho(A_2A_1) \leq 1$ (ρ:spectral radius), an immediate consequence, not explored by Butcher, is the possibility of composing cyclically two (or more) unstable methods of type (1.1) such that the resulting composite method is stable. The Dahlquist barrier for the order of stable linear k-step methods, for instance, may thus be overcome.

Relevant examples - yet not derived from this argument - were the linear cyclic methods of Donelson and Hansen [7] which offered a second interesting aspect: their order of convergence was higher than the order of the linear k-step formulae that generated them. This raised the question, under which condition ("order condition") the local discretization errors with order q accumulate such that the resulting global error has order (q+1). Apparently closely related is the problem of global error estimation.

The answer given in Albrecht [1] led to a new approach to multistage methods, in particular to linear cyclic methods, and contained the results that the global error is proportional to

$$\psi_h^A[d] = \sup_{0\le j\le m_h} \| A^j d_0 + h \sum_{\ell=1}^{j} A^{j-\ell} d_\ell \|$$

d_j: local discretization errors,

and the "A-norm" ψ_h^A is stability functional (in the sense of Spijker) for all A-methods that satisfy the (generalized) root condition. Furthermore, ψ_h^A is, in a certain sense, the best of all meaningful stability functionals and permits two-sided bounds for the global discretization error.

Although originally designed for multistage methods, these results made the A-method concept of relevance also for classical (one-stage) methods.

In this paper, the theory of A-methods is extended, and its practical consequences are discussed. The aim is to develop the necessary tools for a theoretical appraisal of miscellaneous types of (composite) methods and for their practical application.

The earlier results of Albrecht [1] are resumed (in slightly generalized form) in paragraph 2 in order to make this paper self-sufficient. The order condition is generalized in paragraph 4, extending its application and providing a deeper understanding of its meaning. Also, a new approach to Butcher's "effective order" methods is given.

Frequently, several composite methods are available with the same order of consistency; in such cases, something like the error constant in linear k-step methods is needed to compare their efficiency. However, such a quantity should primarily not reflect the size of the local discretization error but its effective contribution to the global error. This idea leads, in paragraph 6, to the concept of "effective" local discretization errors and implies

a generalization of the classical idea of error constants to more general A-methods. It is interesting to note, that Henrici's definition of error constants, $C = c_{q+1}/(\beta_0+\beta_1+\ldots+\beta_k)$, is obtained when the new concept is applied to linear k-step methods.

The result that a method may converge with higher order than the order of its local discretization errors, already calls in question whether the global error can, in all cases, be monitored by the principal local discretization errors. In fact, as shown in paragraph 6, the conventional error control may fail. This result has implications to one of today's most popular set of methods, as is seen in paragraph7.

The user of a specific integration method is not always aware of the fact that his method becomes a composite process as soon as step changing is performed (which represents a major justification for the study of composite methods). An analysis of the effects of step changing demands a generalization of the concept of stability; a note on this problem is given in paragraph 8.

2. Basic Concepts and Theorems

2.1. We consider numerical methods for the initial value problem

$$Y' = f(x,Y) \ , \ Y(a) = \eta_0 \tag{2.1}$$

$\eta_0 \in \mathbb{R}^n$; $f:[a,b]\times\mathbb{R}^n \to \mathbb{R}^n$ Lipschitz continuous.

Y is assumed to be sufficiently smooth so that the derivatives encountered in the analysis exist and are bounded. In order to simplify notation, we restrict to the case n=1 (generalization to n>1 is done by applying the results to each component of (2.1); its formal representation seems not worth the notational inconveniences and the loss of transparency connected with it).

Let I_h and I_h^* be the grids

$$I_h := \left\{ x_j \,|\, x_0=a;\ x_j=x_{j-1}+h,\ j=1(1)m_h;\ m_h=\left[\frac{b-a}{h}\right] \right\} \tag{2.2.a}$$

$$I_h^* := \left\{ x_j^*=x_{rj+k-1},\ j=0(1)m_h^*;\ m_h^*=\left[\frac{m_h-k+1}{r}\right] \right\} \tag{2.2.b}$$

$h>0$ is the stepsize; k and r are integers.

$|*|$ denotes the maximum norm in $\mathbb{R}^s$, and $\mathbb{G}_h$ the space of grid functions $g_h: I_h^* \to \mathbb{R}^s$ with the norm $||g||_h := \sup_{x\in I_h^*} |g_h(x)|$.

We shall consider sets of grid functions g_h, $h\in H = (0,h_0]$, in the resp. set of spaces $\mathbb{G}_h$ with associated norms $||*||_h$. If $\sup_{h\in H} ||g||_h \le \text{const.}$, $g_h \in \mathbb{G}_h$ is uniformly bounded in H.

2.2. We characterize the solution Y of (2.1) at $x_j^* \in I_h^*$ (and its neighbourhood) by a vector $Z_j \in \mathbb{R}^s$. $Z_h \in \mathbb{G}_h$ with values Z_j will be called a discretization of Y. The components of the discretization vectors Z_j usually define an interpolation polynomial of Y in the neighbourhood of x_j^*.

Approximations to Z_h are denoted by $z_h \in \mathbb{G}_h$, its values by z_j. The choice of Z_h (in particular of its dimension s) is arbitrary to a large extent, and mainly depends on the type of method to be used and the algebraic operations to be performed on it.

The following discretization vectors will be of interest here:

$$(Y_{j+k-1},\ Y_{j+k-2},\ \cdots,\ Y_j,\ hF_{j+k-1})^T \tag{2.3.a}$$

$$(Y_{j+k-1},\ hF_{j+k-1},\cdots,\ hF_{j+1},\ hF_j)^T \tag{2.3.b}$$

$$(Y_{rj+k-1},\ hY'_{rj+k-1},\cdots,\frac{1}{k!}h^k\ Y^{(k)}_{rj+k-1})^T \quad ;r,k\text{: integers} \tag{2.3.c}$$

$$(Y_{rj},\ Y_{rj+1},\ \cdots,\ Y_{rj+k-2},\ Y_{rj+k-1})^T \tag{2.3.d}$$

$$(Y_{j-1+a_2},\ Y_{j-1+a_3},\ \cdots,\ Y_{j-1+a_s},\ Y_j)^T \tag{2.3.e}$$

where $F_j := f(x_j,Y_j)$ and $Y_{j+a} := Y(x_j+ah)$.

2.3. Approximations z_j of Z_j are obtained from

$$z_0=\zeta(h);\ z_j= Az_{j-1} + h\phi(x^*_{j-1},\ z_{j-1},z_j;h),\ j=1(1)m^*_h \tag{2.4}$$

where $\zeta,z_j\in \mathbb{R}^s$; $A\in \mathbb{R}(s,s)$; $\phi:I^*_h\times \mathbb{R}^s\times \mathbb{R}^s\times (0,h_o] \to \mathbb{R}^s$,

h_o such that (2.4) has a unique solution for all $h\in(0,h_o]=:H$, and

$$|\phi(x,u_1,v_1;h)- \phi(x,u_2,v_2;h)| \le K_1|u_1-u_2|+ K_2|v_1-v_2| \tag{2.5}$$

$\forall(x\in I^*_h;\ h\in H;\ u_1,u_2,v_1,v_2\in \mathbb{R}^s)$; $K_1>0$, $K_2>0$ constants.

2.4. Definition. Algorithms which can be reduced to form (2.4) will be called A-methods (resp. B- or C-methods if the method is characterized by the matrix B or C). Such a reduction is not unique and depends of the choice of Z_h (see examples in §3).

Definition. The matrix A with eigenvalues μ_i satisfies the root condition if $|\mu_i| \le 1$, i=1(1)s, and if eigenvalues with modulus 1 have linear elementary divisors. It satisfies the

weak root condition if, in addition, $\exists \mu_k \neq 1 : |\mu_k| = 1$, and the strong root condition if $(\forall \mu_i \neq 1)(|\mu_i| < 1)$.

Obs.: (a) A satisfies the root condition iff for any matrix norm:
$\exists D > 0 : (\forall j \in \mathbb{N})(||A^j|| \leq D)$.

(b) In earlier papers of the author [1][2], A-methods with A satisfying the root condition were called "$\hat{A}$-methods".

(c) ϕ is not assumed continuous w.r.t. its first argument. The aim is to facilitate the analysis of step changing procedures (see § 8) and the consideration of non-cyclic combinations of integration formulae (e.g. Butcher's "effective order methods", see sec. 4.8.).

A-methods were considered independently in Albrecht [3], Pierce [13], and Skeel [16]; Butcher [4] presented special cases already in 1966 for which he defined stability and consistency.

2.5. Definition. (2.4) is a k-step method if it needs k starting values; it is a r-stage method if (s-r) of its components are identities (see examples in §3).

It is seen later that the step number k loses some of its significance by the fact that "equivalent" methods may have different step numbers (see §7).

2.6. Definition. z_h converges (for $h \to 0$) with order $p \in \mathbb{N}$ to Z_h iff

$$\exists c > 0 : (\forall h \in H)(||Z - z||_h \leq c h^p).$$

$E_j^Z := (Z_j - z_j)$ are called global discretization errors of the A-method (2.4) with respect to the discretization Z_h.

2.7. Definition. Let $d_h^Z \in \mathcal{G}_h$ be defined by

$$d_0^Z := (Z_0 - \zeta(h)) \tag{2.6}$$

$$hd_j^Z := Z_j - AZ_{j-1} - h\phi(x_j^*, Z_{j-1}, Z_j; h), \quad j=1(1)m_h^*.$$

The d_j^Z, $j=0(1)m_h^*$, are called local discretization errors[1] of the A-method (2.4) with respect to the discretization Z_h.

d_0^Z, in particular, is the error of the starting values.

The largest number $q \in \mathbb{N}$ for which

$$\exists c>0: \ (\forall\, h \in H)\ (||d^Z||_h \leq ch^q)$$

is called order of consistency of the method.

In practice, the l.d.e.s cannot be calculated exactly for unknown Y; however, their lowest order term (in h) can be estimated. This term will be called principal local discretization error; it is decisive for step and order control in modern codes. One aim of this paper is to inquire whether error control based upon the principal l.d.e. is always justified (see §6).

2.8. Frequently, the components of d_j^Z have different orders (for $h \to 0$). The order of consistency as defined above is based on the lowest order component. Cooper [6,pg.66] proposes the alternative of restricting the definition to a few selected components. The same effect can be achieved by proper change of the discretization. The d_j^Z may have different orders (see sec. 3.6.); due to the above definition, the order of consistency is then equal to the lowest order of all d_j^Z, $j=0(1)m_h^*$.

---

[1] labelled l.d.e.s; the superscripts are omitted when no confusion can arise.

2.9. Let δ_j, $j=0(1)m_h^*$, be "perturbations" of the A-method (2.4) and w_j the corresponding solution,

i.e. $w_0 = \zeta + \delta_0$; $w_j = Aw_{j-1} + h\phi(x_{j-1}^*, w_{j-1}, w_j; h) + h\delta_j$. (2.7)

Note that $\delta_j = d_j^Z$ if $w_j = z_j$.

We define a functional $\psi_h^A[\delta] := \sup_{0\le j\le m_h^*} |A^j\delta_0 + h\sum_{\ell=1}^{j} A^{j-\ell}\delta_\ell|$ (2.8)

ψ_h^A is a norm; it generalizes the Spijker-norm [18,p.81] and is relevant for the following central result on A-methods which has been proved earlier [1], [2, p.42].

2.10. Theorem. For any $z_h, w_h, \delta_h \in \mathbb{G}_h$ that satisfy (2.4) and (2.7)

$$\exists c_1>0, c_2>0: (\forall\, h\in(0,h_1])(c_1\psi_h^A[\delta] \le ||w-z||_h \le c_2\psi_h^A[\delta]) \quad (2.9)$$

if and only if $\exists D>0: (\forall\, j\in\mathbb{N})(||A^j||\le D)$ and $h_1DK_2<1$, $h_1\le h_0$.

As a direct consequence we obtain:

2.10.1. Corollary. All A-methods are stable w.r.t. the functional[2] ψ_h^A if and only if A satisfies the root condition.

For $\delta_j = d_j^Z$, $\|A^j\| \le D$ and $h_1DK_2 < 1$, $h_1 \le h_0$, we obtain from theorem 2.10.

$$\exists c_1, c_2>0: (\forall\, h\in(0,h_1])(c_1\psi_h^A[d^Z] \le \|z-z\|_h \le c_2\psi_h^A[d^Z])$$

Since $\psi_h^A[d^Z] \le D \sup_{0\le j\le m_h^*} |d_j^Z|$,

we have:

2.10.2. Corollary. If A satisfies the root condition, and if the A-method has order of consistency q then it converges (for $h \to 0$). Its order of convergence is equal to the order

---

[2] Due to Spijker [17], a method is stable w.r.t. a functional ψ_h if (in our notation) for any $z_h, w_h, \delta_h \in \mathbb{G}_h$ that satisfy (2.4) and (2.7): $\exists c>0, h_1>0: (\forall\, h\in(0,h_1])(||w-z||_h \le c\psi_h[\delta])$.

of $\psi_h^A[d^Z]$ $(h \to 0)$ which is at least equal to q.

Furthermore, for $||A^j|| \le D$, $h_1 DK_2 < 1$, $h_1 \le h_0$ we have

$$\exists c_1, c_2 > 0 : (\forall\, h \in (0, h_1])\ \left(c_1 \psi_h^A[d^Z] \le ||z - z||_h \le c_2\, \psi_h^A[d^Z]\right). \qquad (2.10)$$

It is shown in [2, p.42] that $c_1 \ge \left(1+(b-a)D(K_1+K_2)\right)^{-1}$ and

$$c_2 < e^{(b-a)u} \cdot (1-hDK_2)^{-1}(1+hDK_1) \text{ with } u=(1-hDK_2)^{-1}D(K_1+K_2). \qquad (2.11)$$

The order of convergence can be (q+1) in certain cases, as will be seen in §4.

2.11. Definition. We call an A-method strongly (resp. weakly) stable if A satisfies the strong (resp. weak) root condition. The term "stable" always stands for "stable w.r.t. ψ_h^A" (with the appropriate matrix A).

2.12. If the A-method is implicit, at each step j, an iteration

$$z_j^{(0)} = Pz_{j-1};\ z_j^{(i)} = Az_{j-1} + h\phi(x^*_{j-1}, z_{j-1}, z_j^{(i-1)}; h),\ i=1(1)n, \qquad (2.12)$$

is necessary with a suitable predictor matrix P.

(2.12) may be written as relaxation method:

$$z_j^{(0)} = Pz_{j-1} + h\bar{\phi}(x^*_{j-1}, z_{j-1}; h);\ z_j^{(i)} = z_j^{(i-1)} + h\left(\phi_j^{(i-1)} - \phi_j^{(i-2)}\right);\ i=1(1)n, \qquad (2.13)$$

with $h\phi_j^{(-1)} := Rz_{j-1} + h\,\bar{\phi}(x^*_{j-1}, z_{j-1}; h)$ and $z_j = z_j^{(n)}$, $R := P - A$.

This relates the concept of A-methods to Stetter's "generalized Nordsieck methods" [18, pg.357 ff]. For adequately chosen P, (2.13) has programming benefits; however, there is no advantage in analyzing the method in this form.

3. Examples

The aim of this paragraph is to give a rough idea of the flexibility of the concept of A-methods and the variety of methods that fall into this class.

3.1. Linear multistep methods

With

$$\zeta(h) = \begin{pmatrix} \eta_{k-1} \\ \eta_{k-2} \\ \vdots \\ \eta_1 \\ \eta_0 \end{pmatrix} \quad ; \; z_j = \begin{pmatrix} y_{j+k-1} \\ y_{j+k-2} \\ \vdots \\ y_{j+1} \\ y_j \end{pmatrix} \tag{3.1}$$

$$A = \begin{pmatrix} -\alpha_{k-1} & -\alpha_{k-2} & \cdots & -\alpha_1 & -\alpha_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \; ; \; \phi_j = \begin{pmatrix} \beta_k f_{j+k-1} + \ldots + \beta_0 f_{j-1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

(2.4) represents the (one-stage) linear k-step method

$$y_{j+k-1} + \sum_{\ell=0}^{k-1} \alpha_\ell y_{j+\ell-1} = h \sum_{\ell=0}^{k} \beta_\ell f_{j+\ell-1} \; ; \; j = 1,2,\ldots$$

with $y_i = \eta_i(h)$, $i = 0,1,\ldots,(k-1)$.

The associated discretization vector is $Z_j = (y_{j+k-1}, y_{j+k-2}, \ldots, y_j)^T$.

The root condition reduces to Dahlquist's well-known root condition for linear k-step methods.

We also might have chosen the discretization vector (2.3.a).

Then

$$\zeta(h) = \begin{pmatrix} \eta_{k-1} \\ \eta_{k-2} \\ \vdots \\ \eta_0 \\ hf(x_{k-1}, \eta_{k-1}) \end{pmatrix} \qquad z_j = \begin{pmatrix} y_{j+k-1} \\ y_{j+k-2} \\ \vdots \\ y_j \\ hf_{j+k-1} \end{pmatrix}$$

(3.2)

$$A = \begin{pmatrix} -\alpha_{k-1} & -\alpha_{k-2} & \cdots & -\alpha_1 & -\alpha_0 & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}; \quad \phi_j = \begin{pmatrix} \sum_{\ell=0}^{k} \beta_\ell f_{j+\ell-1} \\ 0 \\ 0 \\ \vdots \\ 0 \\ hf_{j+k-1} \end{pmatrix}$$

3.2. Runge-Kutta methods

The R.K. method

$$y_0=\eta_0;\ y_j=y_{j-1}+\frac{h}{6}\{k_1(x_{j-1},y_{j-1})+4k_2(x_{j-1},y_{j-1})+k_3(x_{j-1},y_{j-1})\}$$

with $k_1=f(x,y)$; $k_2=f(x+\frac{h}{2}, y+\frac{hk_1}{2})$; $k_3=f(x+h\ ,\ y-hk_1+2hk_2)$

is a 1-stage 1-step A-method with A=1 if the discretization vector $Z_j=(Y_j)$ is used.

However, with the discretization vector

$$Z_j=(Y_{j-1/2},\ Y_j,\ Y_j)^T \qquad \text{(see (2.3.e)}$$

it becomes the 3-stage, 1-step A-method:

$$\begin{pmatrix} y_{j-1/2} \\ \bar{y}_j \\ y_j \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{j-3/2} \\ \bar{y}_{j-1} \\ y_{j-1} \end{pmatrix} + h \begin{pmatrix} \frac{1}{2} f_{j-1} \\ -f_{j-1} + 2hf_{j-1/2} \\ \frac{1}{6} f_{j-1} + \frac{4}{6} f_{j-1/2} + \frac{1}{6} \bar{f}_j \end{pmatrix}$$

$$j = 1,2,\ldots$$

$y_0=\eta_0$; $y_{-3/2}$ and $\bar{y}_{-1}$: arbitrary.

Note that z_j need not approximate Z_j with the same order in all components (Cooper's [6] considerations on order vectors become relevant here).

3.3. Linear cyclic methods

Donelson and Hansen's [7] linear 3-cyclic 3-step method

$$\begin{aligned} 33y_{3j} + 24y_{3j-1} - 57y_{3j-2} &= h(10f_{3j} + 57f_{3j-1} + 24f_{3j-2} - f_{3j-3}) \\ 125y_{3j+1} - 144y_{3j} - 117y_{3j-1} + 136y_{3j-2} &= 3h(14f_{3j+1} + 39f_{3j} - 48f_{3j-1} - 15f_{3j-2}) \\ 58y_{3j+2} + 531y_{3j+1} - 306y_{3j} - 283y_{3j-1} &= 3h(3f_{3j+2} + 102f_{3j+1} + 177f_{3j} + 28f_{3j-1}) \end{aligned} \qquad (3.3)$$

$$y_0 = \eta_0;\ y_1 = \eta_1;\ y_2 = \eta_2 \qquad j=1(1)m_h^*$$

may be written as $Lz_j = Uz_{j-1} + h\,\phi_1(x_j^*, z_{j-1}, z_j; h), z_0=\zeta(h)$

with

$$\zeta = \begin{pmatrix} \eta_0 \\ \eta_1 \\ \eta_2 \end{pmatrix};\ z_j = \begin{pmatrix} y_{3j} \\ y_{3j+1} \\ y_{3j+2} \end{pmatrix};\ L = \begin{pmatrix} 33 & 0 & 0 \\ -144 & 125 & 0 \\ -306 & 531 & 58 \end{pmatrix};\ U = \begin{pmatrix} 0 & 57 & -24 \\ 0 & -136 & 117 \\ 0 & 0 & 283 \end{pmatrix}$$

and thus reduces to (2.4) with $A=L^{-1}U$.

The associated discretization vector is (2.3.d) with $r=k=3$.

A has the eigenvalues $\mu_1=1$, $\mu_2=\frac{361}{7975}$, $\mu_3=0$; hence it is stable (although its three stages do not satisfy Dahlquist's root condition).

3.4. Combinations of linear multi-step with R.K. methods

Cyclic application of a R.K. step of the form

$$y_j = y_{j-1} + h\psi_{j-1}$$

followed by two steps with the formula of Milne-Simpson yields the A-method

$$\begin{pmatrix} y_{3j-2} \\ y_{3j-1} \\ y_{3j} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{3j-5} \\ y_{3j-4} \\ y_{3j-3} \end{pmatrix} + \frac{h}{3} \begin{pmatrix} 3\psi_{3j-3} \\ f_{3j-1} + 4f_{3j-2} + f_{3j-3} \\ 3\psi_{3j-3} + f_{3j} + 4f_{3j-1} + f_{3j-2} \end{pmatrix} \quad (3.4)$$

$$j=1,2,\ldots$$

$y_0 = \eta_0$; y_{-1} and y_{-2}: arbitrary

This is a 3-stage 1-step A-method whose A satisfies the strong root condition.

Obviously, any cyclic combination of n A_k-methods $(A_k \in R(s,s), k=1,2,\ldots,n)$ yields an A-method with $A = A_k A_{k-1} \cdots A_1$.

3.5. Methods with correction terms

Consider the method

$$y_{j+2}=\frac{1}{44}(141y_{j+1}-120\bar{y}_j+23\bar{\bar{y}}_{j-1})+\frac{h}{22}(18f_{j+2}-33f_{j+1})$$

$$\bar{y}_{j+1}=\frac{1}{44}(113y_{j+1}-84\bar{y}_j+15\bar{\bar{y}}_{j-1})+\frac{h}{22}(6f_{j+2}-33f_{j+1}) \qquad (3.5)$$

$$\bar{\bar{y}}_j=\frac{1}{44}(69y_{j+1}-40\bar{y}_j+15\bar{\bar{y}}_{j-1})+\frac{h}{22}(6f_{j+2}-33f_{j+1})$$

$$j=1,2,\ldots$$

$$\bar{\bar{y}}_0=\eta_0;\ y_1=\bar{y}_1=\eta_1;\ y_2=\eta_2$$

where earlier approximations y_{j+1} and y_j are "corrected" before y_{j+3} is calculated (the calculation of $\bar{\bar{y}}_j$ and $\bar{y}_{j+1}$ needs no additional function evaluation). Such methods were first considered by Gear [8].

This method may be written in A-method form as

$$\begin{pmatrix} y_{j+2} \\ \bar{y}_{j+1} \\ \bar{\bar{y}}_j \\ hf_{j+2} \end{pmatrix} = \frac{1}{44}\begin{pmatrix} 141 & -120 & 23 & -66 \\ 113 & -84 & 15 & -66 \\ 69 & -40 & 15 & -66 \\ 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} y_{j+1} \\ \bar{y}_j \\ \bar{\bar{y}}_{j-1} \\ hf_{j+1} \end{pmatrix}+\frac{h}{11}f_{j+2}\cdot\begin{pmatrix} 9 \\ 3 \\ 3 \\ 11 \end{pmatrix}$$

$$j=1(1)m_h^*.$$

It is a 3-stage 3-step method; the associate discretization vector is $z_j=\left(Y_{j+2},\ Y_{j+1},\ Y_j,\ hf(x_{j+2},Y_{j+2})\right)^T$.

The method is strongly stable since A has the eigenvalues $\mu_1=1$, $\mu_2=\mu_3=\mu_4=0$.

3.6. Discontinuous methods

In all previous examples, ϕ was the restriction to $I_h^* \times \mathbb{R}^S \times \mathbb{R}^S \times (0,h_o]$ of a continuous function in $[a,b] \times \mathbb{R}^S \times \mathbb{R}^S \times (0,h_o]$. Such methods will be called continuous; they are discontinuous otherwise.

As a simple example, consider the (primitive cyclic) method:

$$Y_0=\eta_0;\ Y_j=Y_{j-1}+h\phi_j;\ \phi_j=\begin{cases} f(x_{j-1},Y_{j-1}) & \text{for odd } j \\ f(x_j,Y_j) & \text{for ever } j \end{cases} \qquad (3.6.a)$$

$$j=1(1)m_h.$$

This is a 1-stage 1-step method with A=1; the associate discretization vector is $Z_j=(Y_j)$.

With the discretization vector $Z_j=(Y_{2j-1},Y_{2j})^T$ we obtain the 2-stage 1-step method:

$$\begin{pmatrix} Y_{2j-1} \\ Y_{2j} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Y_{2j-3} \\ Y_{2j-2} \end{pmatrix} + h \begin{pmatrix} f_{2j-1} \\ f_{2j} \end{pmatrix} \quad ;\ Y_0=\eta_0;\ Y_{-1}\text{: arbitrary} \qquad (3.6.b)$$
$$j=1,2,\ldots,m_h^*.$$

(3.6.a) is a discontinuous and (3.6.b) a continuous representation of the same method. Another example for discontinuous methods are the Butcher methods of sec. 4.6.

4. The Order Condition

In this paragraph, we inquire under what conditions the order of convergence of an A-method is higher than its order of consistency, generalizing earlier results [1].

4.1. Lema. Let $A \in \mathbb{R}(s,s)$ satisfy the root condition with $|\mu_k| = 1$ for $k=1(1)s_1 \le s$ and let be $t_h \in \mathfrak{C}_h$ uniformly bounded for $h \in H := (0,h_o]$.

Then $e_h(x_j^*) := \sum_{\ell=0}^{j} A^{j-\ell}\, t_h(x_\ell^*)$, $j=o(1)m_h^*$ (4.1)

are uniformly bounded in H if and only if

$$\exists c>0: \left(\forall\, h\in H\right)\left(\left|\sum_{\ell=1}^{j} \mu_k^{j-\ell}\, p_k^T t_h(x_\ell^*)\right| \le c,\ j=1(1)m_h^*,\ k=1(1)s_1\right) \qquad (4.2)$$

where p_k are normed, linear independent, left eigenvectors to the eigenvalues μ_k, $k=1(1)s_1$.

Obs.: (4.1) is equivalent to $e_0=t_h(x_0^*)$, $e_j-Ae_{j-1}=t_h(x_j^*)$, $j=1(1)m_h^*$. (4.3)
The sum in (4.2) starts with $\ell=1$, in order to make it independent of the starting values.

Proof. Let be $\{u_i,\ i=1(1)s\}$ the basis of the normed principal vectors of A.

Then, $t_h(x_\ell^*) = \sum_{i=1}^{s} a_i(x_\ell^*)u_i$

$$|e_h(x_j^*)| \le \sum_{i=1}^{s} \left|\sum_{\ell=0}^{j} a_i(x_\ell^*)A^{j-\ell}u_i\right| = \sum_{i=1}^{s} S_i(h)$$

(a) For $k=1(1)s_1$: $A^{j-\ell}u_k=\mu_k^{j-\ell}u_k$ and $a_k(x_\ell^*)=(p_k^Tu_k)^{-1}(p_k^Tt_h(x_\ell^*))$.

Hence, $S_k(h) = \left|\sum_{\ell=0}^{j} \mu_k^{j-\ell}(p_k^Tu_k)^{-1}\left(p_k^Tt_h(x_\ell^*)\right)u_k\right|$

$S_k(h)$ are bounded for $h\in H$ if and only if (4.2) holds.

(b) For $i=(s_1+1)(1)s$: $S_i(h) = \left|\sum_{\ell=0}^{j} a_i(x_\ell^*)A^{j-\ell}u_i\right| \le b_i\|T\|_\infty \left|\sum_{\ell=0}^{j} N^\ell v_i\right|$ (4.4)

where $N=T^{-1}AT$ is the Jordan normal form of A, $v_i=T^{-1}u_i$ the Jordan basis and $b_i=\sup_{h\in H}\sup_{x\in I_h^*}|a_i(x)|$. The sum in (4.4) is bounded for $j\to\infty$ since $|\mu_i| < 1$.

4.2. Definition. An A-method with l.d.e.s

$$d_j = h^q t_h(x_j^*), \quad j=1(1)m_h^*$$

is said to satisfy the order condition iff (4.2) holds.

We are now able to specify necessary and sufficient conditions such that a method with order of consistency q converges with order (q+1).

4.3. Theorem. Let be z_h obtained from an A-method with A satisfying the root condition and with l.d.e.s

$$d_0^z = h^{q+1} t_h(x_0^*) \quad \text{(starting values of order (q+1))},$$
$$d_j^z = h^q t_h(x_j^*), \quad j=1(1)m_h, \tag{4.5}$$

with $t_h \in \mathbb{C}_h$ uniformly bounded for $h \in (0,h_o]$.

z_h then converges (for $h \to 0$) with order (q+1) to Z_h if and only if the order condition (4.2) is satisfied.

Proof. From (2.10) we obtain for sufficiently small $h > 0$:

$$c_1 \psi_h^A[d] \le \|z-z\|_h \le c_2 \psi_h^A[d]$$

with $$\psi_h^A[d] = \sup_{0\le j\le m_h^*} \left| A^j d_o^z + h \sum_{\ell=1}^{j} A^{j-\ell} d_\ell^z \right| = h^{q+1} \sup \left| e_h(x_j^*) \right|.$$

Hence, $\|z-z\|_h = 0(h^{q+1})$ $(h \to 0)$ due to lema 4.1.

For many applications the order condition (4.2) is unhandy and unnecessarily general. It simplifies by imposing additional restrictions on t_h or A.

Let be $$t_h(x_j^*) = g_h(x_j^*)t + hs_h(x_j^*), \quad j=1(1)m_h^*, \tag{4.6}$$

with constant $t \in \mathbb{R}^s$, $g_h: I_h^* \to \mathbb{R}$, $s_h \in \mathbb{C}_h$ uniformly bounded for $h \in (0,h_o]$.

4.3.1. Corollary. Under assumption (4.6), let A satisfy the strong root condition. The order condition (4.2) then reduces to

either $\exists\ c>0: \left(\forall\ h\in(0,h_o]\right)\left(\left|\sum_{\ell=1}^{j} g_h(x_\ell^*)\right| < c,\ j=1(1)m_h^*\right)$ (4.7)

and/or $p^T t=0$ for all $p^T=p^T A,\ |p|=1$. (4.8)

Note that (4.7) can only be satisfied by discontinuous methods (sec. 3.6).

(4.8) is equivalent to $\operatorname{rank}(I-A,t) = s-n$, (4.8.a)

n: number of linear independent left eigenvectors to $\mu=1$.

Proof. Since $\mu_k=1$, $k=1(1)s_1$, (4.2) yields

$$A_{kj} := \left|\sum_{\ell=1}^{j} \mu_k^{j-\ell} p_k^T t_h(x_\ell^*)\right| = \left|\sum_{\ell=1}^{j} p_k^T t_h(x_\ell^*)\right| = \left|(p_k^T t)\left(\sum_{\ell=1}^{j} g_h(x_\ell^*)\right) + h\sum_{\ell=1}^{j} p_k^T s_h(x_\ell^*)\right|.$$

The second sum is bounded due to the factor h.

Corollary 4.3.1. permits an elegant theoretical approach to maximum order linear cyclic methods [1], but it is also decisive for classical algorithms as will be seen in §7. An equivalent condition to (4.8) is given in Skeel [16].

4.3.2. Corollary. Under assumption (4.6), let g_h be the restriction to I_h^* of a function $g:[a,b] \to \mathbb{R}$, $g\not\equiv 0$, independent of h and with bounded derivative (frequently, g is the (q+1) - th derivative of Y). The order condition (4.2) then reduces to

$$p^T t = 0 \quad \text{for all } p^T = p^T A,\ |p| = 1. \tag{4.8}$$

Note that the eigenvalues $\mu_i \neq 1$ do not appear in condition (4.8).

Proof. For $k=1(1)s_1 \leq s$:

$$A_{kj} := \left| \sum_{\ell=1}^{j} \mu_k^{j-\ell} p_k^T t_h(x_\ell^*) \right| = \left| (p_k^T t) \sum_{\ell=1}^{j} \mu_k^{j-\ell} g(x_\ell^*) + \sum_{\ell=1}^{j} h\mu_k^{j-\ell} p_k^T s_h(x_\ell^*) \right|$$

$$= \left| (p_k^T t)\ S_1 + S_2 \right|$$

$|S_2|$ is bounded due to the factor h.

(a) For $\mu_k=1$, S_1 is unbounded for $j \to \infty$; hence, in this case, $A_{kj} \leq$ const. if and only if (4.8) is satisfied.

(b) For $\mu_k \neq 1$, $|\mu_k| = 1$, as g has a bounded derivative:

$$S_1 = \sum_{\ell=1}^{j} \mu_k^{j-\ell} g(x_\ell^*) = \sum_{\ell=1}^{j} \mu_k^{j-\ell} g(x_{\ell-1}^*) + S^*;$$

$$S^* := hr \sum_{\ell=1}^{j} \mu_k^{j-\ell} g'(\xi_\ell) \quad ; \text{ r as in (2.2.b)}$$

$$= \mu_k^{-1} \sum_{\ell=0}^{j-1} \mu_k^{j-\ell} g(x_\ell^*) + S^* = \mu_k^{-1} \left(S_1 + \mu_k^j g(x_0^*) - g(x_j^*) + \mu_k S^* \right)$$

$$|S_1| = |\mu_k^{-1} - 1|^{-1} \left| \mu_k^j g(x_0^*) - g(x_j^*) + \mu_k S^* \right|; \quad S^* \text{: bounded.}$$

Hence, S_1 is bounded in case (b), which finishes the proof.

Corollary 4.3.1. excludes weakly stable and corollary 4.3.2 discontinuous methods.

Methods that satisfy the order condition are distinguished by the fact that their l.d.e.s accumulate in a particularly favourable way. The following theorem presents a complementary aspect by showing that any method with order of consistency q that satisfies the order condition may be interpreted as a method that has order of consistency (q+1) w.r.t. a "neighbouring" discretization W_h.

4.4. Theorem. Let a stable A-method have l.d.e.s

$$d_0^Z = h^{q+1} t_h(x_0^*);\ d_j^Z = h^q t_h(x_j^*),\ j=1(1)m_h^*$$

with $t_h \in \mathbb{C}_h$ bounded for $h\in(0,h_o]$.

If and only if the order condition is satisfied, a discretization W_h exists

$$W_j = Z_j - e_j h^{q+1},\ e_h \in \mathbb{C}_h, \text{uniformly bounded for } h\in(0,h_o], \quad (4.9)$$

$$e_0 = t_h(x_0^*);\ e_j - Ae_{j-1} = t_h(x_j^*),\ j=1(1)m_h^*, \quad (4.10)$$

with respect to which the method has order of consistency (q+1).

Proof. By definition and from lema 4.1:

$$Z_0=\zeta(h)+h^{q+1}t_0;\ Z_j=AZ_{j-1}+h\phi(x_{j-1}^*,Z_{j-1},Z_j;h)+h^{q+1}t_j,\ j=1(1)m_h^*,$$

$$(Z_j-h^{q+1}e_j)=A(Z_{j-1}-h^{q+1}e_{j-1})+h\phi(x_{j-1}^*,Z_{j-1}-h^{q+1}e_{j-1},Z_j-h^{q+1}e_j;h)+hd_j^W$$

$$d_j^W:=\phi(x_{j-1}^*,Z_{j-1},Z_j;h)-\phi(x_{j-1}^*,Z_{j-1}-h^{q+1}e_{j-1},Z_j-h^{q+1}e_j;h)$$

$$= O(h^{q+1}) \quad \text{due to (2.5)}.$$

Note that the transformation (4.9) need not be performed in order to have order of convergence (q+1) nor need e_h be known! Theorem 4.4 offers, however, the possibility to obtain a discretization with respect to which the method does not satisfy the order condition; this will be of interest in a later context (sec. 6.3.).

We might consider W_h the "natural" discretization for the method under consideration. However, W_h is, in general, not a suitable discretization for practical use.

In an effort to reduce the k-step Adams-Moulton-Nordsieck method with order of consistency (k+1), Stetter [18, p.352] calculates the e_j such that (4.10) holds. In that special case this is possible because the order condition is satisfied; in general, the e_j are unbounded.

In a fairly simple way, the order condition solves the problem of finding the order of convergence of all (stable) A-methods that satisfy (4.5). This assumption may be relaxed if order p convergence is not needed at all gridpoints $x_j \in I_h$ but only at a few selected ones; we consider a special class of such methods next.

4.5. Theorem. Let z_j, $j=1(1)(m_h^*-1)$, be obtained from a stable A-method and the last approximation z_m, $m=m_h^*$, from a B-method.

(a) Let the l.d.e.s be

$$d_0^z = h^q t_h(x_o^*) \tag{4.11}$$

$$d_j^z = h^q\left(t_h^1(x_j^*) + t_h^2(x_j^*)\right) \quad , \; j=1(1)(m-1) \tag{4.12}$$

$$d_m^z = -h^{q-1}B\tilde{e}_{m-1}$$

with t_h^1, $t_h^2 \in \mathbb{C}_h$ uniformly bounded for $h \in (0,h_o]$

and $\tilde{e}_0=t_h(x_0^*)$, $\tilde{e}_j=A\tilde{e}_{j-1}+ht_h^1(x_j^*)$, $j=1(1)(m-1)$. (4.13)

(b) Let be

$$\phi(x_{j-1}^*,z_{j-1},z_j;h)=\phi(x_{j-1}^*,z_{j-1}-\tilde{e}_{j-1}h^q,z_j-\tilde{e}_jh^q;h)-h^qt_h^2(x_j^*)+0(h^{q+1}), \tag{4.14}$$

$$j=1(1)(m-1).$$

Then $(z_j-z_j) = \tilde{e}_jh^q$

and $|z_m-z_m| \le (1-hK_2)^{-1}K_1|\tilde{e}_{m-1}|h^{q+1}$ for $hK_2 < 1$.

Note that $\tilde{e}_jh^q$, $\tilde{e}_j=t_h(x_0^*)+h\sum_{\ell=1}^{j}A^{j-\ell}t_h^1(x_\ell^*)$, is the exact global error of z_j, $j=0(1)(m-1)$.

Proof. By definition of the l.d.e.s for $j=1(1)(m-1)$:

$$Z_j = AZ_{j-1} + h\phi(x^*_{j-1}, Z_{j-1}, Z_j; h) + h^{q+1}t^1_h(x^*_j) + h^{q+1}t^2_h(x^*_j).$$

With (4.13) and (4.14):

$$(Z_j - \tilde{e}_j h^q) = A(Z_j - \tilde{e}_{j-1}h^q) + h\phi(x^*_{j-1}, Z_{j-1} - \tilde{e}_{j-1}h^q, Z_j - \tilde{e}_j h^q; h)$$

Hence,
$$z_j = Z_j - \tilde{e}_j h^q . \tag{4.15}$$

This proves the first part of the theorem.

By definition of d^Z_m:

$$Z_m = BZ_{m-1} + h\tilde{\phi}(x_{j-1}, Z_{m-1}, Z_m; h) - h^q B\tilde{e}_{m-1}$$

$$z_m = Bz_{m-1} + h\tilde{\phi}(x_{j-1}, z_{m-1}, z_m; h).$$

Subtraction yields with (4.15) and (2.5):

$$|Z_m - z_m| \le h^{q+1} K_1 |\tilde{e}_{m-1}| + K_2 h |Z_m - z_m|$$

which completes the proof.

Theorem 4.5 generalizes and gives a new approach to methods that were originally presented by Butcher [5] and then analyzed by Stetter, who already observed that their basic idea extends to certain predictor-corrector procedures [18, p.299].

4.6. Example. Consider a procedure consisting of three explicit R.K. methods:

- a first step (or initial value) with l.d.e:

$$d_o^Z = h^q Q(x_o) + 0(h^{q+1})$$

- (m-1) steps with the 2nd method and l.d.e.s

$$d_j^Z = h^q \left(\frac{d}{dx} Q(x) - f_Y(x, Y(x)) \right)_j + 0(h^{q+1}), \quad j=1(1)(m-1)$$

- a last step with the 3rd method and l.d.e.

$$d_m^Z = -h^{q-1} Q(x_m)$$

where $Q \in C^1[a,b]$.

Then, $\tilde{e}_j = Q(x_{j+1}) + 0(h)$, $j=1(1)(m-1)$, $B=1$ and (4.14) is satisfied, thus, theorem 4.5 applies.

The above procedure characterizes Butcher's "effective order" methods (see Stetter [18, p.160]). This name refers to their order of convergence being different from their order of consistency. However, the same applies to methods that satisfy the order condition. Therefore, it seems reasonable to substitute this nomenclature and to call them Butcher methods.

5. Applications of the Order Condition

5.1. The classical Adams-Moulton-Nordsieck methods are perhaps the most important procedures that satisfy the order condition.

As an example, we consider Nordsieck's form of the 3-step Adams-Moulton method given by

$$\begin{pmatrix} y_j \\ hy'_j \\ \frac{1}{2}h^2y''_j \\ \frac{1}{6}h^3y'''_j \end{pmatrix} = \begin{pmatrix} 1 & 5/8 & 2/8 & -1/8 \\ 0 & 0 & 0 & 0 \\ 0 & -3/4 & -1/2 & 3/4 \\ 0 & -1/6 & -1/3 & 1/2 \end{pmatrix} \begin{pmatrix} y_{j-1} \\ hy'_{j-1} \\ \frac{1}{2}h^2y''_{j-1} \\ \frac{1}{6}h^3y'''_{j-1} \end{pmatrix} + hy'_j \begin{pmatrix} 3/8 \\ 1 \\ 3/4 \\ 1/6 \end{pmatrix} \qquad (5.1)$$

$$j=3(1)m_h$$

$y_j^{(i)}$ denote approximations of $Y_j^{(i)}$; $y'_j=f(x_j,y_j)$.

For sufficiently differentiable Y the l.d.e.s are

$$d_j=h^3Y^{(4)}(x_j)t+0(h^4);\ t=\frac{1}{144}(-3;0;18;20)^T,\ j=3(1)m_h.$$

Thus the method has order of consistency q=3. It is strongly stable as A has eigenvalues $\mu_1=1$, $\mu_2=\mu_3=\mu_4=0$; furthermore, $p^T=p^TA$ for $p^T=(1;1/2;1/6;0)$.

Hence, (4.8) is satisfied and - with starting values of order 4 - the method converges with order (q+1) = 4 (theorem 4.3 and corollary 4.3.1).

Obs.: The above notation differs from Gear's [9]; we shall come back to this point in section 7.5.

5.2. The linear cyclic method (3.3) is strongly stable; for sufficiently differentiable Y its l.d.e.s have order q=5 and the form

$$d_j = h^5 Y^{(6)}(x_j^*)t + 0(h^6) \quad , \quad j=1(1)m_h^*$$

with $t = L^{-1}t^*$, $t^* = c(33;261;-225)^T$, $c\neq 0$.

With $A = L^{-1}U$ we obtain

$$\operatorname{rank}(I-A,t)=\operatorname{rank}(L-U,t^*)=\operatorname{rank}\begin{pmatrix} 33 & -57 & 24 & 33c \\ -144 & 261 & -117 & 261c \\ -306 & 531 & -225 & -225c \end{pmatrix}$$

$$=\operatorname{rank}\begin{pmatrix} -57 & 24 & 33 \\ 261 & -117 & 261 \\ 531 & -225 & -225 \end{pmatrix} = 2 .$$

Thus, (4.8.a) is satisfied and, due to theorem 4.3, the method converges with order q+1=6, if the starting values have order 6.

A detailed treatment of linear cyclic methods is given in Albrecht [2].

5.3. Consider the 1-step, r-stage block-implicit method

$$\begin{pmatrix} y_{rj-r+1} \\ y_{rj-r+2} \\ \vdots \\ y_{rj} \end{pmatrix} = \begin{pmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 0 & 1 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{rj-2r+1} \\ y_{rj-2r+2} \\ \vdots \\ y_{rj-r} \end{pmatrix} + h \begin{pmatrix} \beta_0^1 & \dots & \beta_r^1 \\ \beta_0^2 & \dots & \beta_r^2 \\ \dots & \dots & \dots \\ \beta_0^r & \dots & \beta_r^r \end{pmatrix} \begin{pmatrix} f_{rj-r} \\ f_{rj-r+1} \\ \vdots \\ f_{rj} \end{pmatrix} \quad (5.2)$$

$$j=1(1)m_h^*$$

$z_0=(0;\dots;0;\eta_0)^T \in \mathbb{R}^r$.

Let r be even, and let the β_i^j be determined such that the last stage has order of consistency q=(r+2) and all other stages q=(r+1) (the method is then uniquely defined).

Theorem. Under the above assumptions and for sufficiently differentiable Y, method (5.2) converges with order (r+2).

It should be remembered that order of convergence p implies that all components of z_j converge with order p to the resp. components of z_j.

Proof. The method is strongly stable and $p^T = (0;\ldots;0;1)$ is left eigenvector of A to $\mu=1$. The l.d.e.s have order q=(r+1):

$$d_j = h^{r+1} Y^{(r+2)}(x_j^*)t + 0(h^{r+2}),\quad j=1(1)m_h^*$$

$$t = (t_1; \ldots; t_{r-1}; 0)^T,\quad t_i: \text{error const. of i-th stage.}$$

Hence, $p^T t=0$ and the theorem follows from 4.3.

For odd r, the order condition cannot be satisfied; the corresponding methods therefore converge with order (r+1). Shampine and Watts [14] showed that the procedure (5.2) is A-stable for r=2(1)8.

5.4. The (discontinuous) primitive cyclic method (3.6.a) is strongly stable and, for sufficiently differentiable Y, its l.d.e.s have order q=1:

$$d_j = \begin{cases} \frac{1}{2} \\ -\frac{1}{2} \end{cases} Y''(x_j)h + 0(h^2) \qquad \begin{matrix} \text{for odd } j \\ \\ \text{for even } j. \end{matrix}$$

Hence, corollary 4.3.1 applies and condition (4.7) is satisfied with $g_h(x_j)=(-1)^{j+1}Y''(x_j)$. Due to theorem 4.3, the method converges with order (q+1) = 2.

In its continuous representation (3.6.b), the same method has

$$d_j = \tfrac{1}{2}h\, Y''(x_j)\ t + 0(h^2)\ ;\ t = \begin{pmatrix}1\\0\end{pmatrix},\quad j=1(1)m_h^*.$$

$A = \begin{pmatrix}0 & 1\\ 0 & 1\end{pmatrix}$ has the eigenvalues $\mu_1=1$, $\mu_2=0$ and the left eigenvector $p^T = (0;1)$. Hence, $p^T t=0$ which again proves the order of convergence (q+1) = 2.

5.5. Method (3.4) is an example of a procedure that does not satisfy assumption (4.6). Therefore, the order condition applies - if at all - in its general form (4.2). With $\mu_1=1$, $\mu_2=\mu_3=0$, $p_1^T=(0;0;1)$ it reduces to

$$\left(\forall\, h\in(0,h_0]\right)\left(\left|\sum_{\ell=0}^{j} p_1^T t_h(x_\ell^*)\right| = \left|\sum_{\ell=0}^{j} t_h^{(3)}(x_\ell^*)\right| < c\right),\quad j=1(1)m_h^*, \qquad (5.3)$$

where $t_h^{(3)}$ denotes the 3rd component of t_h.

Assume that the R.K. formula in (3.4) is a 5-stage order 4 method designed such that the last formula of (3.4) has order 5 (this cannot be achieved with a 4-stage order 4 R.K. formula!)

The method then has order of consistency q = 4 and $t_h^{(3)}(x_\ell^*)=0(h)$, $\ell=1(1)m_h^*$.

Thus, (5.3) is satisfied for a constant $c>0$, and the method has order of convergence $(q+1)=5$.

This is only a simple example out of a large variety of methods that is generated by composite application of different integration formulae, and that may contain efficient alternatives to actually applied procedures. The aim of this paper is to provide the means for their theoretical analysis.

As step changing automatically implies the transition to a multi-stage methods, even if we continue with the "same" formula (see sec. 8.1), we may use the order condition to combine step changes with an increase of order; yet, this possibility needs further investigation.

6. Error Control

The rest of this paper will be primarily concerned with the practical consequences of the A-method theory, beginning with its implications to error control:

6.1. The usual way to control the global discretization error at x_m^*, $E_m := (Z_m - z_m)$, is to estimate the principal l.d.e. (sec. 2.7) and to monitor stepsize and order according to it. An essential part of all modern codes is touched if we ask for the justification of this procedure.

There is no doubt that the global error E_m is composed of the l.d.e.s d_ℓ^Z, $\ell=0(1)m$; however, they do not accumulate in a purely additive way (safe is special cases such as R.K. methods or Adams methods). Essential for the influence of the d_ℓ^Z on E_m is, therefore, not only their size but how they accumulate. If $|hDK_2|< 1$, this depends upon the A-norm of d_h

$$\psi_h^A[d] = \sup_{0\le j\le m^*} |A^j d_0 + h \sum_{\ell=1}^{j} A^{j-\ell} d_\ell| \qquad (6.1)$$

as is seen from corollary 2.10.2. Rather than monitor the method, in the j-th step, by the quantity $|d_j^Z|$, we should estimate the effect of the j-th step on (6.1).

6.2. This reasoning affects, in particular, our definition of an "error constant" for general A-methods.

In the case of the linear 3-cyclic method (3.3), for example, a stepsize and order control based upon the l.d.e.s of the stages (in analogy to the procedure in linear multistep methods) would fail since its "error constants" $c_6^{(i)}$ (i=1,2,3) do not reflect the composite structure of the method and, worse, the order of the l.d.e.s does not even indicate the correct order of convergence.

This excludes also the suggestive idea of defining an error constant for multistage methods simply by $|t|$, if the method has l.d.e.s of the form

$$d_j = h^q y^{(q+1)}(x_j^*)t + 0(h^{q+1}), \quad j=1(1)m_h^*. \tag{6.2}$$

When analyzing the accumulation of the l.d.e.s, we have to distinguish between methods that satisfy an order condition and those that do not.

6.3. Consider a strongly stable A-method, with l.d.e.s

$$d_j^Z = h^q t_h(x_j^*), \quad j=0(1)m_h^* \tag{6.3}$$

that does not satisfy the order condition.

Let be μ_i, $i=1(1)s$, the eigenvalues of A with $\mu_k=1$ for $k=1(1)s_1$;

p_k, $k=1(1)s_1$, linear independent normed left eigenvectors to $\mu=1$;

u_k, $k=1(1)s_1$, linear independent normed eigenvectors to $\mu=1$;

$\{u_i, i=1(1)s\}$ the basis of the normed principal vectors of A.

Then, $t_h(x_\ell^*) = \sum_{i=1}^{s} a_i(x_\ell^*)u_i$,

$$a_k(x_\ell^*) := (p_k^T u_k)^{-1}\left(p_k^T t_h(x_\ell^*)\right), \quad k=1(1)s_1.$$

Let be $\tilde{e}_j := A^j t_0 + h\sum_{\ell=1}^{j} A^{j-\ell} t_\ell$, $j=0(1)m_h^*$

$$= \sum_{i=1}^{s} a_i(x_0^*)A^j u_i + h\sum_{i=1}^{s} S_i(h); \quad S_i = \sum_{\ell=1}^{j} a_i(x_\ell^*)A^{j-\ell}u_i, \quad i=1(1)s.$$

The sums S_i, $i=(s_1+1)(1)s$, are bounded due to (4.4), and from $|\mu_i| < 1$ follows $|A^j u_i| = o(h)$ $(h\to 0)$, $i=(s_1+1)(1)(s)$.

Hence, $\tilde{e}_j = \sum_{k=1}^{s_1}\left(a_k(x_0^*)u_k + (p_k^T u_k)^{-1}u_k\, h\sum_{\ell=1}^{j}(p_k^T t_\ell)\right) + O(h)$.

As $\psi_h^A[d] = h^q \sup_{0\le j\le m_h^*}|\tilde{e}_j|$, the effect of the j-th step on ψ_h^A is $h\sum_{k=1}^{s_1}(p_k^T u_k)^{-1}(p_k^T t_j)u_k$, in essence.

Thus, the quantity to estimate for error control is

$$\varepsilon_j := \sum_{k=1}^{s_1}(p_k^T u_k)^{-1}(p_k^T t_j). \tag{6.4}$$

In most practical cases we have $s_1=1$.

In the (very common) special case

$$s_1=1;\ d_j^Z = h^q Y^{(q+1)}(x_j^*)t + O(h^{q+1}) \text{ with constant } t\in\mathbb{R}^s \tag{6.5}$$

(6.4) takes the form (omitting the subscripts k)

$$\varepsilon_j = (p^T u)^{-1}(p^T t)\ Y^{(q+1)}(x_j^*) \tag{6.6}$$

6.4. Definition. ϵ_j will be called effective l.d.e.s, and the constant

$$C = (p^T u)^{-1}(p^T t)$$

is the error constant (of an A-method with (6.5)).

In the special case of strongly stable linear multistep methods (3.1) with l.d.e.s (6.2), $t=(c_{q+1};0;\ldots;0)^T \in \mathbb{R}^k$, we have

$$p^T=\left\{1;(1+\alpha_{k-1});\ldots;\left(1+\sum_{i=1}^{k-1}\alpha_i\right)\right\};\ u=(1;1;\ldots;1)^T$$

thus,

$$C = c_{q+1}/(\beta_0+\beta_1+\ldots+\beta_k). \tag{6.7}$$

Hence, definition 6.4 contains the classical error constant for linear multistep methods (see Henrici [11, p.223]) as special case.

Obs.: Extension of the above considerations to weakly stable methods is omitted as it is of little practical interest.

The concept of error constants and, more general, of effective l.d.e.s permits a finer measure of accuracy and is useful for automatic order control in A-methods.

6.5. We consider now the case of a stable A-method with l.d.e.s

$$d_h^Z(x_0^*) = h^{q+1}t_h(x_0^*) \; ; \; d_h^Z = h^q t_h(x_j^*) + h^{q+1}s_h(x_j^*), \; j=1(1)m_h^* \qquad (6.8)$$

that satisfies the order condition (4.2).

Due to theorem 4.4., the method then has order of consistency (q+1) with respect to W_h, $W_j := Z_j - h^{q+1}e_j$. For sufficiently small h it follows from corollary 2.10.2. that

$$c_1\psi_h^A[d^W] \le \sup_{0\le j\le m_h^*} |z_j - z_j - h^{q+1}e_j| \le c_2\psi_h^A[d^W]$$

where $\psi_h^A[d^W]$ depends on the terms $h^{q+1}s_h(x_j^*)$.

Consequently, the global error $E_j := (Z_j - z_j)$ cannot be safely controlled by the principal l.d.e. $h^q t_h$ alone, if the order condition is satisfied; both t_h and s_h must be considered (by monitoring e_j and the principal l.d.e.s with respect to W_h).

As an alternative, the method may be modified such that its order of consistency is maintained but the order condition not satisfied. Of course, this reduces its order of convergence by 1. An example, that concerns one of today's most widely used codes is given in the next paragraph.

Hence, the significance of the concept of order condition lies not only in its capacity to increase the order of a method, but in the possibility of finding out whether conventional error control techniques are permitted. This affects error control in the block-implicit methods (5.2), the linear k-cyclic, k-step methods of order 2k, certain Nordsieck forms, and many other methods.

7. Nordsieck Forms of A-Methods

In this paragraph, linear transformations are performed on A-methods in order to obtain new procedures. This yields, in particular, a very simple approach to the well-known Adams-Moulton-Nordsieck methods and a generalization of its main ideas to A-methods.

7.1. Definition . Two stable methods M_1 and M_2 that can be represented by

$$z_0=\zeta(h) \; ; \; z_j=Az_{j-1}+ h\phi_A(x^*_{j-1},z_{j-1},z_j;h) \qquad j=1(1)m^*_h \tag{7.1}$$

and

$$w_0=\omega(h) \; ; \; w_j=Bw_{j-1}+ h\phi_B(x^*_{j-1},w_{j-1},w_j;h) \tag{7.2}$$

are called equivalent if

$$w_j = Tz_j, \quad j=0(1)m^*_h, \; T\in R(s,s), \text{ regular.}$$

Apparently, if M_1 and M_2 are equivalent,

$$B = TAT^{-1}; \; \phi_B(x^*_{j-1},u,v;h) = T\phi_A(x^*_{j-1},T^{-1}u,T^{-1}v;h).$$

7.2. Let

$$y_{j+k-1} = y_{j+k-2} + h(\beta_k f_{j+k-1} + \ldots + \beta_0 f_{j-1}), \quad j=1(1)m_h^*$$

$$y_i = \eta_i \quad , \; i=0(1)(k-1)$$

represent the k-step Adams-Moulton method of order (k+1).

With the discretization (2.3.b) it takes the form

$$z_0 = \zeta(h); \; z_j = Az_{j-1} + hf_j^* a \quad , \; j=1(1)m_h^* \tag{7.3}$$

$$a \in \mathbb{R}^k, \quad f_j^* := f(x_j^*, y_j^*) = f(x_{j+k-1}, y_{j+k-1})$$

$$z_j = \begin{pmatrix} y_{j+k-1} \\ hf_{j+k-1} \\ hf_{j+k-2} \\ \vdots \\ hf_{j+1} \\ hf_j \end{pmatrix} \; ; \quad A = \begin{pmatrix} 1 & \beta_{k-1} & \beta_{k-2} & \cdots & \beta_1 & \beta_0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \; ; \quad a = \begin{pmatrix} \beta_k \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

Let T be defined by $W_j = TZ_j + 0(h^{k+1})$

where Z_j and W_j are the discretization vectors (2.3.b) and (2.3.c) with r=1. By transformation of (7.3) with

$$z_j = Tw_j \tag{7.4}$$

we obtain the equivalent method (re-labelling w by z):

$$z_j = Bz_{j-1} + hf_j^* b, \quad B = TAT^{-1}, \quad b = Ta. \tag{7.5}$$

This is the k-step Nordsieck form of the Adams-Moulton method (labelled A-M-N method) of order (k+1).

7.3. It must be distinguished between the above classical k-step A-M-N method with order (k+1) , and the k-step A-M-N method of order k that results, if the same transformation (7.4) is applied to the (k-1)-step Adams-Moulton method with discretization (2.3.b) (dimension s=k+1) and order k. Note that this method is a k-step method although its equivalent counterpart is a (k-1)-step procedure (which confirms the previous remark that equivalent methods may have different step-numbers k).

By the following definition, we extend the concept of Nordsieck forms to more general A-methods.

7.4. Definition. All methods with $z_0=\zeta(h)$; $z_j=Az_{j-1}+h\phi(x_j^*,f_{j-1}^*,f_j^*;h)$ with associated discretization vector (2.3.c) are called Nordsieck forms.

Nordsieck forms have the well-known advantage that step changing is particularly simple, consisting of pre-multiplying z_j by the matrix $D(\alpha)=\mathrm{diag}(1,\alpha,\alpha^2,\ldots,\alpha^{s-1})$, where α is the step changing factor.

7.5. The notation (7.5) of the Adams-Moulton-Nordsieck methods differs from Gear's [9]; however, both are equivalent.
To see this, let P be the matrix of an explicit method ("predictor") with order k

$$z_j = Pz_{j-1}, \quad j=1(1)m_h^*, \text{ and } R := P-B.$$

Solving (7.5) by iteration yields

$$z_j^{(0)} = Pz_{j-1};\ z_j^{(i)} = Bz_{j-1} + hf_j^{*(i-1)}b,\ i=1(1)n;\ z_j = z_j^{(n)};$$

$$= (P-R)z_{j-1} + hf_j^{*(i-1)}b \qquad (7.6.a)$$

$$z_j^{(i-1)} = (P-R)z_{j-1} + hf_j^{*(i-2)}b. \qquad (7.6.b)$$

From (7.6.a/b): $z_j^{(i)} = z_j^{(i-1)} + h\left(f_j^{*(i-1)} - f_j^{*(i-2)}\right)b,\ i=1(1)n,$ (7.7)

where $$hf_j^{*(-1)}b := z_j^{(0)} - Pz_{j-1} + Rz_{j-1} = Rz_{j-1}.$$

(7.7) is Gear's representation. Apparently, (7.5) is simpler and, therefore, more suitable for analysis; however, (7.7) has well-known advantages when coded.

It has already been pointed out in sec. 2.12., that any implicit A-method can be written as relaxation method. Hence, the advantages of this form, as explained in Gear [10] for the special case of linear multistep methods, extend to all implicit A-methods.

7.6. Let be

$$w_0=\omega(h);\ w_j=Bw_{j-1}+ h\phi_B(x^*_{j-1},w_{j-1},w_j;h) \tag{7.8}$$

obtained from

$$z_0=\zeta(h);\ z_j=Az_{j-1}+ h\phi_A(x^*_{j-1},z_{j-1},z_j;h) \tag{7.9}$$

by the transformation $w_j = Tz_j$, $j=0(1)m^*_h$.

Assume that $z_h\in \mathbb{G}_h$ converges with order p to the discretization Z_h of Y; we then may interpret w_h as an order p approximation to any other discretization W_h of Y with

$$W_j= TZ_j + c_h(x^*_j)h^r,\ r\geq p,\quad j=0(1)m^*_h \tag{7.10}$$

$$c_h\in \mathbb{G}_h \text{ uniformly bounded in } (0,h_o].$$

With respect to this discretization, method (7.8) may have advantages over (7.9) which motivated the introduction of equivalent methods. They may differ in many aspects such as order of consistency, number of starting values, step changing procedures, and error control techniques.

The case r=p is of special interest; the l.d.e.s of (7.8) with respect to W_h are then

$$d^W_j= Td^Z_j + (c_j-Bc_{j-1})h^{p-1}+ 0(h^p),\ j=1(1)m^*_h. \tag{7.11}$$

If (7.9) has order of consistency p, $d^Z_j=0(h^p)$, then method (7.8) has order of consistency q=(p-1) but, due to construction, it converges with order p to W_h. This proves the following two theorems.

7.7. Theorem. If two equivalent methods have different orders of consistency (but the same order of the starting values), then the lower order method satisfies the order condition.

7.8. Theorem. Let (7.9) have order of consistency p w.r.t. Z_h. If (7.10) holds for r=p, then method (7.8) satisfies the order condition.

From (7.4) follows for the special case of A-M-N methods:

7.9. Corollary. All k-step Adams-Moulton-Nordsieck methods with order (k+1) satisfy the order condition.

7.10. Gear's widely used code DIFSUB[3] is based on Nordsieck forms of Adams-Moulton methods and, in view of the results of sec. 6.5, one may wonder why its error control works.

It has been said in sec. 6.5. that difficulties with conventional error control can be avoided by modifying the method such that the order condition is not satisfied while the order of consistency is maintained. The error control in DIFSUB is correct because the classical k-step A-M-N- procedure of order (k+1) is modified to the k-step, order k A-M-N method of sec. 7.3. As DIFSUB starts with k=1, the lower order of these formulae has no negative consequence.

The following theorem implies that error control of a method may be performed with the effective l.d.e.s of any equivalent method, if both do not satisfy the order condition.

7.11. Theorem. Equivalent methods that do not satisfy the order condition have the same effective l.d.e.s, in particular, - in case (6.5) - the same error constants.

Proof. Let be $B=TAT^{-1}$, $\hat{t}_j=Tt_j$, $\hat{p}_k^T B=\hat{p}_k^T$, $B\hat{u}_k=\hat{u}_k$, $k=1(1)s_1$.

Then $\hat{p}_k^T = p_k^T T^{-1}$, $\hat{u}_k=Tu_k$ and the theorem follows from

$$\hat{\varepsilon}_j := \sum_{k=1}^{s_1} (\hat{p}_k^T \hat{u}_k)^{-1} (\hat{p}_k^T \hat{t}_j) = \varepsilon_j .$$

---

3) Refering to DIFSUB always includes its modifications, e.g. Hindmarsh's code LSODE [12].

8. A Note on Step Changing Effects

This paragraph deals with a generalization of the concept of A-methods which is of interest for the study of step changing effects. A complete theory does not yet exist; the considerations here are fragmentary and should be taken as a first step.

8.1. Our theory generalizes without changes to methods

$$z_0=\zeta(h);\ z_j=Az_{j-1}+h\phi(x^*_{j-1},z_{j-1},z_j;h),\ j=1(1)m^*_h, \tag{8.1}$$

on non-equidistant grids

$$I_h:=\{x_j \mid x_0=a,\ x_j=x_{j-1}+\alpha_j h,\ \alpha_j\geq 0,\ j=1(1)m_h\} \tag{8.2}$$

with $\sum_{j=1}^{m_h}(\alpha_j h) = (b-a)$, $h\in(0,h_o]$; $x^*_j\in I^*_h$ as in (2.2.b).

Then, (8.1) is of discontinuous type, where ϕ depends on the grid factors α_j.

As in §2, the l.d.e.s d_j of (8.1) are defined by

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

where Z_j is the discretization vector (on the grid (8.2)).

Composite application of n A_i-methods, i=1(1)n, on (8.2) generates an A-method either if the methods are used cyclically (then $A=A_n A_{n-1}\ldots A_1$) or if all A_i are equal. In order to specify those compositions that do not result in an A-method, we introduce the term "n-composite" method.

8.2. Definition. The composite application of n A_i-methods is called a n-composite procedure, if it cannot be reduced to an A-method.

8.3. Step changes in an A-method on equidistant grid result in the procedure

$$z_0=\zeta(h);\ z_j=A_j z_{j-1}+ h\phi(x^*_{j-1},z_{j-1},z_j;h),\quad j=1(1)m^*_h, \tag{8.3}$$

on the grid (8.1), where A_j and ϕ usually depend on α_j. In general, (8.3) is n-composite.

As n-composite methods are no A-methods, the stability and convergence considerations presented so far do not strictly apply and must be generalized. From (8.3) we have $z_j=A_jA_{j-1}\cdots A_k z_{k-1}+hG_j$ where G_j depends on A_i,ϕ_i,α_i and h, i=k(1)j. This motivates the following definition which obviously generalizes the root condition.

8.4. Definition. The n-composite method (8.3) is stable, if there is a constant c>0 such that for any $h\in(0,h_o]$ and any sequence $\{\alpha_j\}$ with $\sum_{j=1}^{m^*_h}(\alpha_j h) = (b-a)$

$$D = \sup_{j,k} \|A_jA_{j-1}\cdots A_{k+1}A_k\| \le c\quad ,\ 1\le k\le j\le m^*_h. \tag{8.4}$$

8.5. Usually, $A_j=A$ for n_j steps with fixed stepsize, and $A_j\neq A$ for m_j subsequent steps which characterize the step changing process. The latter matrices will be called step changing matrices.

The current codes possess no particular means to satisfy (8.4) which works if the step increases (α_{j+1}/α_j) are not "too large". (In LSODE, for example, (α_{j+1}/α_j) is limited by 10, except in the starting phase, and in Shampine's code DETEST [15] it is limited by 2). Our aim is to quantify this assertion. This seems quite difficult for all situations that may arise in step changing; it is possible, however, if we restrict to special step changing techniques. Condition (8.4) can be satisfied simply by performing the step change such that the product is equal to some power of a matrix B with bounded $\|B^j\|$; B need not be equal to A. This can be achieved in

two ways:

(a) making the step change matrices equal to A (in this case B=A),

(b) performing, after every step change, p steps with a B-method such that $B^pT = B^p$, (8.5)
where T denotes the product of the step changing matrices.

Stability, as defined in 8.4, thus becomes independent of the grid factors α_j.

In case (a), (8.3) reduces to an A-method on (8.2), and it is seen from corollary 2.10.2. how the α_j must be bounded: As the Lipschitz constant K_2 now depends on α_j, we have as a sufficient condition for convergence:

$$|DK_2(\alpha_j)h| < 1 \quad \text{with } D := \sup_{j \in \mathbb{N}} \|B^j\|. \tag{8.6}$$

It is possible to establish a similar restriction on α_j in case (b). (8.6) excludes <u>stiff</u> problems.

8.6. Example. Consider the i-th step of a linear 3-step method of order 3:

$$z_j := \begin{pmatrix} y_j \\ y_{j-1} \\ y_{j-2} \\ hf_j \end{pmatrix} = \begin{pmatrix} -a_2 & -a_1 & -a_0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_{j-1} \\ y_{j-2} \\ y_{j-3} \\ hf_{j-1} \end{pmatrix} + h \begin{pmatrix} b_3^0 f_j + b_2^0 f_{j-1} + b_1^0 f_{j-2} + b_0^0 f_{j-3} \\ 0 \\ 0 \\ f_j \end{pmatrix} \tag{8.7}$$

A common procedure for a step change $h \leftarrow \alpha_i h$ at x_i would be to calculate new approximations at $(x_i - \alpha_i h)$ and $(x_i - 2\alpha_i h)$ by interpolation, i.e.

$$T(\alpha_i) z_i = \tilde{z}_i = \begin{pmatrix} y_i \\ y_{i-\alpha_i} \\ y_{i-2\alpha_i} \\ hf_i \end{pmatrix} \quad ; \qquad T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ x & x & x & x \\ x & x & x & x \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{8.8}$$

and then to apply (8.7) for $j=(i+1),(i+2),\ldots$ with $z_i \leftarrow \tilde{z}_i$ and $h \leftarrow \alpha_i h$. This would define a n-composite method (with $A_1 = A$, $A_i = AT(\alpha_i)$). Its stability (in the sense of definition 8.4) depends on the size of the α_j and the frequency of step increases. The following two alternatives maintain stability.

(a) The step change is performed with formulae on the non-equidistant grid points $\{x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}\}$ by

$$y_{i+1} = -a_2 y_i - a_1 y_{i-1} - a_0 y_{i-2} + h\left(b_3^1(\alpha_i) f_{i+1} + \ldots + b_0^1(\alpha_i) f_{i-2}\right)$$

$$y_{i+2} = -a_2 y_{i+1} - a_1 y_i - a_0 y_{i-1} + h\left(b_3^2(\alpha_i) f_{i+2} + \ldots + b_0^2(\alpha_i) f_{i-1}\right)$$

where $b_k^r(\alpha_i)$ $(r=1,2;\ k=0,1,2,3)$ are determined such that the formulae have order 3.

This method is discontinuous but not n-composite and

$K_2(\alpha_j) = L \max_{j,r}\left(\left|b_3^{(r)}(\alpha_j)\right|, 1\right)$, r=0,1,2; L:Lipschitz constant of f.

(b) The step change uses $\tilde{z}_i$, but it is performed with two Adams steps with order 3 and stepsize $\alpha_i h$:

$$z_{i+1} = B\tilde{z}_i + \alpha_i h\phi(x^*_i, \tilde{z}_i, z_{i+1}; \alpha_i h) \qquad B := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$z_{i+2} = Bz_{i+1} + \alpha_i h\phi(x^*_{i+1}, z_{i+1}, z_{i+2}; \alpha_i h)$$

Since $B^2T = B^2$ for any T of the form (8.8), we have

$D = \sup_{j \in \mathbb{N}} \left(||B^j|| , ||BT|| \right)$ and thus stability.

If (8.7) is an Adams method itself, then stability is achieved simply by keeping the stepsize fixed for 2 steps[4] after a step increase. This appears to be the principal advantage of Adams methods in comparison to other schemes.

8.7. In DIFSUB, for example, after any step increase, at least (k+1) steps with the current order k are performed without allowing another increment or a change of order. The above analysis reveals that this measure stabilizes the code in its Adams-mode, but not in its BDF-mode, since the BDF-matrices do not have the property (8.5). A stabilizing effect could be achieved in the BDF-mode, if Adams forms were used in the k steps following a step increase, but this would destroy the code's advantages for stiff problems. On the other hand, step changing also effects stiff stability. This might not be noticed as a defect of the method but, instead, be interpreted as an increase of stiffness in the problem, since the effects due to step changes would be undistinguishable from the effects of stiffness.

---

4) Note that this requirement is as restrictive as, for a 3-stage cyclic method, the rule to finish a cycle before the stepsize is increase, which shows that linear cyclic methods have no disadvantages with this respect over one-stage methods.

Conclusions

An unlimited number of new methods can be generated by proper combination of classical procedures and, considering the large variety of possibilities, it seems very likely that some of them will prove superior to the methods that are used today. One reason why this vast area of alternatives yet has been very little explored was probably the absence of adequate theoretical tools for their analysis and comparison - apart from the very satisfactory performance of the existing codes. But, how develop competitive alternatives without proper means to compare methods, to perform error control, or to execute step changes efficiently?

This paper is an attempt to remove these difficulties by amplifying the theoretical basis of a large class of O.D.E. methods that includes almost all relevant procedures (with the exception of multi-derivative methods to which the theory probably may be generalized).

The theory presented does not add new aspects or results to Runge-Kutta methods, due to their special structure, but the A-method notation is useful (see e.g. [6]). It is the natural representation for composite methods of most miscellaneous type, and thus may become the standard notation for O.D.E. methods.

It has been observed in sec. 2.2. that the discretizations Z_j define an interpolation polynomial; hence, A-methods may be interpreted as procedures that generate sequences of polynomial approximations of the solution Y. In this context, Stetter's ideas in [19] become relevant for the A-method theory, and one wonders whether the A-method approach could produce a contribution to this line of research.

Major parts of our analysis rely on the assumption $hK_2D<1$ which is not satisfied for stiff problems. This restriction can be overcome

- however not in an easy manner - if we include the Jacobian of f in a generalization of theorem 2.10.

It is not yet well-seen, how the ideas in §8 can be effectively used in practice, and when stabilizing efforts - beyond those that are already applied in today's codes - are necessary.

Acknowledgement

The author is grateful to A.B.Pereira for helping to understand certain details of DIFSUB, and to R.D.Grigorieff for valuable comments that resulted in various improvements of this paper.

References

[1] Albrecht, P., Explicit,optimal stability functionals and their application to cyclic discretization methods, Computing 19(1978), pp. 233-249.

[2] Albrecht, P., Die numerische Behandlung gewöhnlicher Differentialgleichungen, Akademie-Verlag, Berlin, 1979; Hanser-Verlag München, 1979.

[3] Albrecht, P., Métodos de discretização, IMPA, Rio de Janeiro, 1973.

[4] Butcher, J.C., On the convergence of numerical solutions to ordinary differential equations, Math.Comp. 20 (1966), pp. 1-10.

[5] Butcher, J.C., The effective order of Runge-Kutta methods, Lecture Notes in Math. 109, pp. 133-139, Springer-Verlag, 1969.

[6] Cooper, G.J., Error estimates for general linear methods for ordinary differential equations, SIAM J. Num. Anal. 18 (1981), pp. 65-82.

[7] Donelson, J; Hansen, E., Cyclic composite multistep predictor-corrector methods, SIAM J. Num. Anal.8 (1971), 137-157.

[8] Gear, C.W.;The numerical integration of ordinary differential equations, Math.Comp. 21 (1967), pp. 146-156.

[9] Gear, C.W.; Numerical initial value problems in ordinary differential equations, Prentice-Hall, 1971.

[10] Gear, C.W., Simultaneous numerical solution of differential-algebraic equations, IEEE Trans. Circuit Theory, Vol. CT-18, No.1 (1971), pp. 89-95.

[11] Henrici, P., Discrete variable methods in ordinary differential equations, J. Wiley, N.Y., 1962.

[12] Hindmarsh, A.C., User's manual for LSODE, Nat. Energy Software Center, No. 81-70, 1980.

[13] Pierce, S., Analysis and optimization of cyclic methods for application in orbit computations, Report NASA CR-132826, 1974.

[14] Shampine, L.F.; Watts, H.A., A-stable block implicit one-step methods, BIT 12 (1972), pp. 252-266.

[15] Shampine, L.F.; Gordon, M.K., Computer solution of ordinary differential equations, W.H. Freeman, S.Francisco, 1975.

[16] Skeel, R., Analysis of fixed-stepsize methods, SIAM J. Num. Anal. 13 (1976), pp. 271-300.

[17] Spijker, M.N., On the possibility of two-sided error bounds in the numerical solution of initial value problems, Num. Math. 26 (1976), pp. 271-300.

[18] Stetter, H.J., Analysis of discretization methods for ordinary differential equations, Springer-Verlag, Berlin, 1973.

[19] Stetter, H.J., Interpolation and error estimation in Adams PC-codes, SIAM J. Num. Anal. 16 (1979),pp.311-322.