Tendler-like Formulas for Stiff ODEs

Elmar Klausmeier

Contributing authors: Elmar.Klausmeier@gmail.com;

Abstract

This paper proves a convergence result for a general class of methods for the solution of ordinary differential equations (initial value problems). The proof uses standard results from the theory of matrix polynomials. We present new cyclic linear multistep formulas of orders 3 to 9 for stiff equations, which, order by order, outperform the cyclic composite multistep methods of Tendler with respect to the Widlund-wedge angle and Widlund-distance. The Tendler formulas had already outperformed the BDF order by order. We present numerical accuracy comparisons on Dahlquist's and Runge's test equations with the BDF, the original Tendler, the new methods, and the Tischer methods.

Keywords: cyclic linear multistep methods, matrix polynomials, stiff differential equations, convergence analysis, Widlund-wedge

Mathematics Subject Classification: 65L04 65L06

CR: G.1.7

1. Introduction

We consider the numerical solution of the ordinary differential equation, the initial value problem

$$\dot{y}(t) = f(t, y(t)), \quad y(t_0) = y_0 \in \mathbb{C}^n, \quad t \in [t_0, t_{\mathrm{end}}] \subset \mathbb{R}.$$

Assume that $f : [t_0, t_{\mathrm{end}}] \times \mathbb{C}^n \to \mathbb{C}^n$ is Lipschitz-continuous in y for all t, and continuous in t. We will use cyclic linear multistep methods for this, i.e., we will employ a fixed set of linear multistep methods and cycle through them.

Cyclic linear multistep methods are attractive for a number of reasons:

1. They are not subject to Dahlquist's first and second barriers.
2. They share the same low overhead per step as linear multistep methods.

3. They offer more parameters to attain desirable stability or accuracy properties than linear multistep methods.

Below is an example of an order-four cyclic composite method, which we call eTendler4 (enhanced Tendler4). It cycles through three different linear multistep methods for some fixed step size $h \in \mathbb{R}$, $m = 1, 2, \ldots$:

$$\begin{aligned}
3y_{3m-3} - 16y_{3m-2} + 36y_{3m-1} - 48y_{3m} &+ 25y_{3m+1} \\
&= 12h\dot{y}_{3m+1}, \\
16y_{3m-2} - 90y_{3m-1} + 234y_{3m} &- 214y_{3m+1} + 54y_{3m+2} \\
&= -84h\dot{y}_{3m+1} + 36h\dot{y}_{3m+2}, \\
15y_{3m-1} - 94y_{3m} &+ 162y_{3m+1} - 114y_{3m+2} + 31y_{3m+3} \\
&= 48h\dot{y}_{3m+1} - 60h\dot{y}_{3m+2} + 24h\dot{y}_{3m+3}.
\end{aligned}$$

The first stage of eTendler4 is the standard BDF4 (the classic backward differentiation formula of order 4). Each stage is a 4-step linear multistep formula.

In addition to the initial value y_0, the above difference method needs three starting values y_1, y_2, and y_3 and will then produce discrete values y_k for $k = 4, 5, \ldots$, as approximations of order $\mathcal{O}(h^4)$ to the desired function y, $\dot{y}_k = f(t_k, y_k)$, $t_k = t_0 + kh$. These starting values can be generated using lower-order methods with fewer required starting values and a smaller step size $|h|$.

In a later chapter it will be shown that the above method converges for sufficiently small $|h|$ for a wide range of functions f.

The difference equation above can be expressed in matrix-vector form:

$$A_2 Y_m + A_1 Y_{m-1} + A_0 Y_{m-2} = h\left(B_2 \dot{Y}_m + B_1 \dot{Y}_{m-1} + B_0 \dot{Y}_{m-2}\right),$$

with

$$A_0 = \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_1 = \begin{pmatrix} -16 & 36 & -48 \\ 16 & -90 & 234 \\ 0 & 15 & -94 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 25 & 0 & 0 \\ -214 & 54 & 0 \\ 162 & -114 & 31 \end{pmatrix},$$

and

$$B_0 = \mathbf{0}, \quad B_1 = \mathbf{0}, \quad B_2 = \begin{pmatrix} 12 & 0 & 0 \\ -84 & 36 & 0 \\ 48 & -60 & 24 \end{pmatrix},$$

and

$$Y_{m-2} = \begin{pmatrix} y_{3m-5} \\ y_{3m-4} \\ y_{3m-3} \end{pmatrix}, \quad Y_{m-1} = \begin{pmatrix} y_{3m-2} \\ y_{3m-1} \\ y_{3m} \end{pmatrix}, \quad Y_m = \begin{pmatrix} y_{3m+1} \\ y_{3m+2} \\ y_{3m+3} \end{pmatrix}, \quad \dot{Y}_m = \begin{pmatrix} f(t_{3m+1}, y_{3m+1}) \\ f(t_{3m+2}, y_{3m+2}) \\ f(t_{3m+3}, y_{3m+3}) \end{pmatrix}.$$

Stability of this cyclic method is now governed by the two matrix polynomials

$$\rho(\mu) = A_2\mu^2 + A_1\mu + A_0, \quad \sigma(\mu) = B_2\mu^2 + B_1\mu + B_0.$$

All stages in eTendler4 are implicit and must therefore be solved iteratively (fixed-point or Newton).

The stages of a composite method might be explicit or implicit. They might have different orders, and might require a different number of starting values.

2. Consistency and stability

1. Consider the multistep method

$$L(y, t_0, h) = \sum_{i=0}^{k} \left(\alpha_i y(t_0 + ih) - h\beta_i \dot y(t_0 + ih)\right) \in \mathbb{C}^n.$$

and

$$\rho(\mu) := \sum_{i=0}^{k} \alpha_i \mu^i \in \mathbb{C}, \qquad \sigma(\mu) := \sum_{i=0}^{k} \beta_i \mu^i \in \mathbb{C}.$$

The consistency conditions are:

$$C_{p,k} \cdot \begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \left(A \,|\, B\right) \begin{pmatrix}\alpha\\ \beta\end{pmatrix} = 0.$$

The consistency matrix $C_{p,k}$ for a linear k-step method of consistency order p is given by the following equation

$$\left(\begin{array}{cccccc|cccccc} 1 & 1 & 1 & 1 & \dots & 1 & 0 & 0 & 0 & 0 & \dots & 0\\ 0 & 1 & 2 & 3 & \dots & k & -1 & -1 & -1 & -1 & \dots & -1\\ 0 & 1^2 & 2^2 & 3^2 & \dots & k^2 & 0 & -2\cdot 1 & -2\cdot 2 & -2\cdot 3 & \dots & -2\cdot k\\ 0 & 1^3 & 2^3 & 3^3 & \dots & k^3 & 0 & -3\cdot 1^2 & -3\cdot 2^2 & -3\cdot 3^2 & \dots & -3\cdot k^2\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 1^p & 2^p & 3^p & \dots & k^p & 0 & -p1^{p-1} & -p2^{p-1} & -p3^{p-1} & \dots & -pk^{p-1} \end{array}\right) \cdot \begin{pmatrix}\alpha_0\\ \alpha_1\\ \vdots\\ \alpha_{k-1}\\ \alpha_k\\ \beta_0\\ \beta_1\\ \vdots\\ \beta_{k-1}\\ \beta_k\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\\ \vdots\\ 0\end{pmatrix}.$$

2. Theorem: The following eight propositions are equivalent. Therefore any of these statements can be used as a definition for consistency of order p.

(1) $C_{p,k}\binom{\alpha}{\beta} = \mathbf{0}$, so $(\alpha, \beta)^\top \in \ker C_{p,k}$.

(2) $\sum_{i=0}^k \alpha_i i^q = q \sum_{i=0}^k \beta_i i^{q-1}$, for $q = 0, 1, \dots, p$.

(3) $\rho(e^h) - h\sigma(e^h) = \mathcal{O}(h^{p+1})$, for $h \to 0$.

(4) $\zeta = 1$ is a zero of multiplicity at least p of the function

$$h \mapsto \frac{\rho(\zeta)}{\ln \zeta} - \sigma(\zeta),$$

so $\rho(\zeta)/\ln\zeta = \sigma(\zeta) + \mathcal{O}\big((\zeta-1)^p\big)$, for $\zeta \to 1$.

(5) $L(f,t,h) = \mathcal{O}(h^{p+1}), \quad \forall f \in C^{p+2}(G,\mathbb{C})$.

(6) The monomials up to the degree p lie in the kernel of the $h = 1$ cut of L; therefore $L(t^i, t_0, 1) = 0$, for $i = 0, 1, \ldots, p$.

(7) $L(f, t_0, h) = \mathcal{O}(h^{p+1})$, for the special function $t \mapsto f(t) = e^t$.

(8) $L(y, t_0, h) = c_{p+1} h^{p+1} y^{(p+1)}(t_0) + \mathcal{O}(h^{p+2})$ with

$$c_{p+1} = \frac{1}{(p+1)!} \sum_{i=0}^{k} \big(\alpha_i i^{p+1} - (p+1)\beta_i i^p\big).$$

The factor c_{p+1} is called the *unscaled error constant.*

Proof Mostly trivial reformulations. See Hairer/Wanner/Norsett (2008), Thm. III.2.4. □

3. Our composite method is

$$\sum_{i=0}^{\kappa} A_i Y_{n+i} = h\,\varphi(Y_{n+\kappa}, \ldots, Y_n) = h \cdot \left(\sum_{i=0}^{\kappa} B_i \dot{Y}_{n+i} \right), \quad n = 0, 1, 2, \ldots.$$

$\varphi(\cdot)$ also depends on h, t_{n+i}, and f, but for brevity we omit that. This general method is called *general linear method* in [5] and Hairer/Wanner/Norsett (2008). The characteristic polynomial is

$$\det Q(\mu, H) := \det \left\{ \sum_{i=0}^{\kappa} (A_i - HB_i)\mu^i \right\}.$$

The stability region of a method is the area

$$\{H \in \mathbb{C} : \det Q(\mu, H) = 0 \wedge |\mu| < 1\}.$$

This set is not necessarily connected, i.e., it might contain holes. See, for example, the two block-implicit methods of order 8 and 10 from [4].

The *stability mountain* is the set

$$\{(H, |\mu|) \in \mathbb{C} \times \mathbb{R}^{\geq} : \det Q(\mu, H) = 0\}.$$

This mountain is interesting for observing how fast $|\mu|$ decays, i.e., the gradient.

4. Definition. Slightly deviating and expanding from [22]. For real values $\alpha \ge 0$, $\delta \ge 0$, $r \ge 0$, let

$$\mathcal{A}[\alpha] = \{z \in \mathbb{C}^- : |\arg(-z)| \le \alpha \wedge z \ne 0\}, \quad \mathcal{S}[\delta] = \{z \in \mathbb{C} : \operatorname{Re} z \le -\delta \le 0\}.$$

1. A method is D-stable if the Jordan triple for the matrix polynomial $Q(\mu, 0)$ has all its eigenvalues in the closed unit disc with all eigenvalues of magnitude 1 lying in 1×1 blocks, see [5], Thm 142C.
2. The eigenvalues not equal to 1 for $Q(\mu, 0)$ with the greatest magnitude are called the parasitic roots.
3. A method is $A[\alpha]$-stable if $\mathcal{A}[\alpha]$ is a subset of the stability region.
4. The largest α for which a method is $A[\alpha]$-stable is called the Widlund-wedge angle α.
5. A method is $S[\delta]$-stable if $\mathcal{S}[\delta]$ is a subset of the stability region.
6. The smallest δ for which a method is $S[\delta]$-stable is called the Widlund-distance δ.
7. If $\alpha = 90°$ or $\delta = 0$ then the method is called A-stable.
8. An $A[\alpha]$-stable method with the additional property that for $\operatorname{Re} H \to -\infty$ all $|\mu| < r$ is called $A_\infty^r[\alpha]$-stable. The smallest r is of interest.
9. An $S[\delta]$-stable method with the additional property that for $\operatorname{Re} H \to -\infty$ all $|\mu| < r$ is called $S_\infty^r[\delta]$-stable.
10. For $r < 1$ we abbreviate with $A_\infty[\alpha]$- and $S_\infty[\delta]$-stable. For $r = 0$ the limit $r \to 0$ is meant. An $A_\infty^0[90°]$-stable method is called L-stable, see [11], definition 3.7.

5. Examples. The above stability characteristics manifest in all shapes and forms.

1. The explicit Euler method, $y_{n+1} = y_n + hf(t_n, y_n)$, is D-stable but not A-stable, nor $A[\alpha]$-stable.
2. The implicit Euler method (=BDF1), $y_{n+1} = y_n + hf(t_{n+1}, y_{n+1})$, and the BDF2 are D-stable and A-stable.
3. The BDFi are D-stable for $i = 1, \ldots, 6$, and are unstable for all $i \ge 7$, see [10], III.3, Thm 3.4.
4. The BDFi ($i = 1, \ldots, 6$) are $A_\infty^0[\alpha]$-stable with α being 90°, 90°, 86.03°, 73.35°, 51.84°, 17.84°, respectively, see [1].
5. All Tischer cyclic composite formulas, see [23] and [24], are $A_\infty^0[\alpha]$-stable and $S_\infty^0[\delta]$-stable and have equal to zero parasitic roots. These properties are highly desirable for the integration of stiff systems.
6. The 5-step cyclic linear multistep method with 5 stages of order seven from [12] is $A[44.8°]$- and $S[6]$-stable. The α given in [12] is slightly smaller.
7. The linear multistep formula of order nine from [9] used in DSTIFF is $S_\infty^{0.989}[2.086]$-stable.
8. The DSTIFF formula of order 10 from [9] is $A_\infty^{0.878}[63.74°]$-stable.
9. The new formulas eTendler8 and eTendler9 are $S_\infty^0[\delta]$-stable but not $A[\alpha]$-stable.

3. Convergence result

The proof of the convergence result is conducted in the setting of matrix polynomials, see [7] or [8].

For the differential equation we can confine ourselves to the scalar case $n = 1$, i.e., $\mathbb{C}$ instead of $\mathbb{C}^n$. Otherwise we would have to add $A_i \otimes I_{n\times n}$, and $B_i \otimes I_{n\times n}$, etc. everywhere.

6. Notation. Let (P_1, C_1, R_1) be the first companion triple for the matrix polynomial

$$\rho(\mu) := I\mu^\ell + A_{\ell-1}\mu^{\ell-1} + \cdots + A_1\mu + A_0 \in \mathbb{C}^k,$$

of degree $\ell \geq 1$.

Let $|t_{\text{end}} - t_0| \neq 0$. The step size h is such that $|t_{\text{end}} - t_0|/|h|$ is an integer. N is defined by $N|h| = |t_{\text{end}} - t_0|$.

Let

$$P_1 := \begin{pmatrix} I & 0 & \dots & 0 \end{pmatrix} \in \mathbb{C}^{k\times k\ell}, \qquad R_1 := \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I \end{pmatrix} \in \mathbb{C}^{k\ell\times k}.$$

Let $I = I_{k\times k}$. The first companion matrix is

$$C_1 := \begin{pmatrix} 0 & I & 0 \dots & 0 \\ 0 & 0 & I \dots & 0 \\ \vdots & \vdots & \vdots \ \ddots & \vdots \\ & & \dots & I \\ -A_0 & -A_1 & \dots & -A_{\ell-1} \end{pmatrix} \in \mathbb{C}^{k\ell\times k\ell}.$$

Further

$$\overline{\mathbf{P}}_1 := \operatorname*{diag}_{\nu=0}^{N} P_1 = \begin{pmatrix} I\ 0 \dots 0 & & & \\ & I\ 0 \dots 0 & & \\ & & \ddots & \\ & & & I\ 0 \dots 0 \end{pmatrix} \in \mathbb{C}^{(N+1)k\times(N+1)k\ell},$$

and

$$\overline{\mathbf{R}}_1 := \operatorname{diag}\left(I_{k\ell\times k\ell},\ \operatorname*{diag}_{\nu=1}^{N} R_1\right) = \begin{pmatrix} I_{k\ell\times k\ell} & & \\ & 0 & \\ & \vdots & \\ & 0 & \\ & I & \\ & & \ddots & \\ & & & 0 \\ & & & \vdots \\ & & & 0 \\ & & & I \end{pmatrix} \in \mathbb{C}^{(N+1)k\ell\times(N+\ell)k}.$$

Let

$$\mathbf{R} := \operatorname*{col}_{\nu=0}^{N+\ell-1} r_\nu = \begin{pmatrix} r_0 \\ \vdots \\ r_{N+\ell-1} \end{pmatrix} \in \mathbb{C}^{(N+\ell)k}.$$

7. Definition. Let T be an arbitrary square matrix of size $k \times k$. The *bidiagonal operator* $[T]$ for the matrix T of size $(N+1)k \times (N+1)k$, is defined as follows:

$$[T] := \begin{pmatrix} I & & & \\ -T & I & & \\ & \ddots & \ddots & \\ & & -T & I \end{pmatrix}, \qquad [T]^{-1} = \begin{pmatrix} I & & & \\ T & I & & \\ \vdots & \vdots & \ddots & \\ T^N & T^{N-1} & \dots & I \end{pmatrix}.$$

See [18].

For the product we have: $[C_1]^{-1}\overline{\mathbf{R}}_1 \in \mathbb{C}^{(N+1)k\ell\times(N+\ell)k}$.

Let (X, T, Y) be an arbitrary standard triple. Due to biorthogonality we have

$$\left(\operatorname*{col}_{i=0}^{\ell-1} XT^i\right)^{-1} = \left(\operatorname*{row}_{i=0}^{\ell-1} T^iY\right) B,$$

with the block-Hankel-matrix B

$$B = \begin{pmatrix} A_1 & \dots & A_\ell \\ \vdots & \iddots & \\ A_\ell & & \end{pmatrix}, \qquad A_\ell = I.$$

See [8], Prop. 2.1.

Further

$$\overline{\mathbf{X}} := \operatorname*{diag}_{\nu=0}^{N} X = \begin{pmatrix} X & & & \\ & X & & \\ & & \ddots & \\ & & & X \end{pmatrix}$$

and

$$\overline{\mathbf{Y}} := \operatorname{diag}\left[\left(\operatorname*{col}_{i=0}^{\ell-1} XT^i\right)^{-1}, \operatorname*{diag}_{\nu=1}^{N} Y\right] = \begin{pmatrix} \left(\operatorname{col}_{i=0}^{\ell-1} XT^i\right)^{-1} & & & \\ & Y & & \\ & & \ddots & \\ & & & Y \end{pmatrix}.$$

The special handling of the block matrix for $\overline{\mathbf{R}}_1$ and $\overline{\mathbf{Y}}$ in the first "diagonal element" has its root in the solution representation of the difference equation for matrix polynomials of the form

$$x_n = XJ^n \left(\operatorname*{col}_{i=0}^{\ell-1} XJ^i\right) \begin{pmatrix} y_0 \\ \vdots \\ y_{\ell-1} \end{pmatrix} + X \sum_{\nu=0}^{n-1} J^{n-1-\nu} Y y_{\nu+\ell}.$$

For the case $\ell = 1$ we have $\rho(\mu) = I\mu - A$ and the two matrices P_1 and R_1 reduce to the identity matrices of size $n \times n$. The biorthogonality relation reduces to $X = Y^{-1}$ or $X^{-1} = Y$.

8. Theorem: (Discrete Lemma of Gronwall) Let $0 \le \eta_0 \le \eta_1 \le \ldots \le \eta_m$ be $(m+1)$ positive numbers. Furthermore $\delta \ge 0$, $h_j \ge 0$ and $x_{j+1} = x_j + h_j$. Assume

$$\varepsilon_0 \le \eta_0 \qquad \text{and} \qquad \varepsilon_{j+1} \le \eta_j + \delta \sum_{\nu=0}^{j} h_\nu \varepsilon_\nu, \qquad j = 0, \ldots, m-1.$$

Then

$$\varepsilon_j \le \eta_j \, e^{\delta\cdot(x_j - x_0)}, \qquad j = 0, \ldots, m.$$

Proof See [26]. The case $\delta = 0$ is simple, due to $e^0 = 1$. Hence, let $\delta > 0$. Starting the induction with $j = 0$ is obvious, again due to $e^0 = 1$. We perform induction from j to $j+1$, assuming $\delta > 0$. We have

$$\begin{aligned}
\varepsilon_{j+1} &\le \eta_{j+1} + \delta \sum_{\nu=0}^{j} h_\nu \, \varepsilon_\nu \\
&\le \eta_{j+1} + \delta \sum_{\nu=0}^{j} h_\nu \, \eta_\nu \, e^{\delta\cdot(x_\nu - x_0)} \\
&\le \eta_{j+1} \cdot \left(1 + \delta \sum_{\nu=0}^{j} h_\nu \, e^{\delta\cdot(x_\nu - x_0)} \right) \\
&\le \eta_{j+1} \, e^{\delta\cdot(x_{j+1} - x_0)}.
\end{aligned}$$

This is so because for the sum in parentheses we can estimate (the sum of a strictly monotonically increasing function)

$$\sum_{\nu=0}^{j} h_\nu \, e^{\delta\cdot(x_\nu - x_0)} \le \int_{x_0}^{x_{j+1}} e^{\delta\cdot(t-x_0)} \, dt = \frac{1}{\delta} \left(e^{\delta(x_{j+1}-x_0)} - 1 \right).$$

□

9. Theorem: (Properties of col, row, diag, $[\cdot]$) We have

1. $\operatorname{col} A_\nu B_\nu = \operatorname{diag} A_\nu \ \operatorname{col} B_\nu$.
2. $\operatorname{col} A_\nu B = (\operatorname{col} A_\nu)\, B$; right distributivity of col-operator.
3. $\operatorname{row} A_\nu B_\nu = \operatorname{row} A_\nu \ \operatorname{diag} B_\nu$.
4. $\operatorname{row} A B_\nu = A \ \operatorname{row} B_\nu$; left distributivity of row-operator.
5. $\operatorname{diag} A_\nu B_\nu = \operatorname{diag} A_\nu \ \operatorname{diag} B_\nu$.
6. $\left[S^{-1} T S\right] = \operatorname{diag} S^{-1} \ [T] \ \operatorname{diag} S$.
7. $\left[S^{-1} T S\right]^{-1} = \operatorname{diag} S^{-1} \ [T]^{-1} \operatorname{diag} S$.

Proof Trivial computations. □

10. Theorem: (Solution of difference equation) The general solution of the difference equation

$$x_{n+\ell} + A_{\ell-1} x_{n+\ell-1} + \cdots + A_0 x_n = y_n, \qquad n = 0, 1, \ldots, N$$

is

$$x_n = P_1 C_1^n z_0 + P_1 \sum_{\nu=0}^{n-1} C_1^{n-1-\nu} R_1 y_\nu.$$

Proof See [8], Thm. 1.6. □

11. Theorem: (Representation theorem) Prerequisites: $\hat{u}_n$ and u_n are the solutions of the two difference equations

$$\left.\begin{aligned}
\hat{u}_{n+\ell} + A_{\ell-1} \hat{u}_{n+\ell-1} + \cdots + A_0 \hat{u}_n &= h\, \varphi(\hat{u}_{n+\ell}, \ldots, \hat{u}_n) + r_{n+\ell} \\
u_{n+\ell} + A_{\ell-1} u_{n+\ell-1} + \cdots + A_0 u_n &= h\, \varphi(u_{n+\ell}, \ldots, u_n)
\end{aligned}\right\} \qquad n = 0, 1, \ldots, N.$$

The “disturbances” $r_{n+\ell}$ correspond to $\hat{u}_{n+\ell}$. We will use

$$\left.\begin{aligned} \delta_{n+\ell} &:= \hat{u}_{n+\ell} - u_{n+\ell}, \\ \hat{\delta}_{n+\ell} &:= \varphi(\hat{u}_{n+\ell},\ldots,\hat{u}_n) - \varphi(u_{n+\ell},\ldots,u_n) \end{aligned}\right\} \qquad n = 0,\ldots,N.$$

The difference equation for $\hat{u}_n$ has the starting values $\hat{u}_i := u_i + r_i$, for $i = 0,\ldots,\ell-1$. Let $\delta_i := r_i$, for $i = 0,\ldots,\ell-1$, and $r_\nu := \delta_\nu := \hat{\delta}_\nu := 0$, for $\nu > N$.

Proposition:

$$\begin{aligned} \delta_n &= P_1 C_1^n \begin{pmatrix} \delta_0 \\ \vdots \\ \delta_{\ell-1} \end{pmatrix} + P_1 \sum_{\nu=0}^{n-\ell} C_1^{n-1-\nu} R_1 \left(r_{\nu+\ell} + h\hat{\delta}_{\nu+\ell} \right) \\ &= P_1 C_1^n \begin{pmatrix} \delta_0 \\ \vdots \\ \delta_{\ell-1} \end{pmatrix} + P_1 \sum_{\nu=0}^{n-\ell} C_1^{n-1-\nu} R_1 r_{\nu+\ell} + h\, P_1 \sum_{\nu=0}^{n-\ell} C_1^{n-1-\nu} R_1 \hat{\delta}_{\nu+\ell}. \end{aligned}$$

Proof Follows immediately from the previous theorem. ☐

12. Theorem: Prerequisite: Let $C_1^i := \mathbf{0} \in \mathbb{C}^{k\ell\times k\ell}$, when $i < 0$.

Proposition: The reduced stability functional is norm-equivalent to the original stability functional, i.e.,

$$\left|[C_1]^{-1}\overline{\mathbf{R}}_1\mathbf{R}\right| \sim \left|\overline{\mathbf{P}}_1[C_1]^{-1}\overline{\mathbf{R}}_1\mathbf{R}\right|.$$

Proof In two parts. We estimate each against the other.

(1) The estimation $\left|\overline{\mathbf{P}}_1[C_1]^{-1}\overline{\mathbf{R}}_1\mathbf{R}\right| \le \left|\overline{\mathbf{P}}_1\right| \left|[C_1]^{-1}\overline{\mathbf{R}}_1\mathbf{R}\right|$ is obvious. The row-norm of $\overline{\mathbf{P}}_1$ is independent of N.

(2) We use

$$\left|C_1^n z_0\right| \le \left|C_1^{\ell-1}\right| \, \left|C_1^{n-\ell+1} z_0\right| = \left|C_1^{\ell-1}\right| \max_{i=0}^{\ell-1} \left|P_1 C_1^{n+i-\ell+1} z_0\right|,$$

due to

$$\left|C_1^n z_0\right| = \max_{i=0}^{\ell-1} \left|P_1 C_1^{n+i} z_0\right|.$$

We can ”extract” $C_1^{\ell-1}$ because the sup-norm for $\left|\overline{\mathbf{P}}_1[C_1]^{-1}\overline{\mathbf{R}}_1\mathbf{R}\right|$ still goes over all rows. Finally

$$\left|C_1^{n-1-\nu} R_1 r_{\nu+\ell}\right| \le \left|C_1^{\ell-1}\right| \, \left|C_1^{n-\ell-\nu} R_1 r_{\nu+\ell}\right| = \left|C_1^{\ell-1}\right| \max_{i=0}^{\ell-1} \left|P_1 C_1^{n-\ell-\nu+i} R_1 r_{\nu+\ell}\right|,$$

due to $r_\nu := 0$, for $\nu > N$. ☐

13. Theorem: (Estimation theorem) Prerequisite: Let $\varphi(\cdot)$ be Lipschitz-continuous in each component with Lipschitz constant K_i. The values $\delta_{\nu+\ell}$ and $\hat{\delta}_{\nu+\ell}$ are as above.

Proposition:

$$\begin{aligned}\sum_{\nu=0}^{n-\ell}|\hat\delta_{\nu+\ell}| &\le K_\ell|\delta_n| + \left(\sum_{i=0}^{\ell}K_i\right)\left(\sum_{\nu=0}^{n-1}|\delta_\nu|\right)\\ &\le \left(\sum_{i=0}^{\ell}K_i\right)\left(\sum_{\nu=0}^{n}|\delta_\nu|\right)\\ &\le (\ell+1)\cdot\left(\max_{i=0}^{\ell}K_i\right)\sum_{\nu=0}^{n}|\delta_\nu|.\end{aligned}$$

Proof For $\nu = 0,\ldots,n-1$ we have

$$\begin{aligned}|\hat\delta_{\nu+\ell}| &= |\varphi(\hat u_{\nu+\ell},\ldots,\hat u_\nu) - \varphi(u_{\nu+\ell},\ldots,u_\nu)|\\ &\le K_0|\delta_\nu| + K_1|\delta_{\nu+1}| + \cdots + K_\ell|\delta_{\nu+\ell}|.\end{aligned}$$

For ease of notation $\delta_\nu \leftarrow |\delta_\nu|$ and $\hat\delta_\nu \leftarrow |\hat\delta_\nu|$. Hence,

$$\begin{aligned}\sum_{\nu=0}^{n-\ell}\hat\delta_\nu &\le (K_0\delta_0+\cdots+K_\ell\delta_\ell) + (K_0\delta_1+\cdots+K_\ell\delta_{\ell+1}) + \cdots + (K_0\delta_{n-\ell+1}+\cdots+K_\ell\delta_n)\\ &= K_0\,(\delta_0+\cdots+\delta_{n-\ell+1})\\ &\quad + K_1\,(\delta_1+\cdots+\delta_{n-\ell+2})\\ &\quad\quad + \quad \cdots\\ &\quad\quad\quad + K_\ell\,(\delta_\ell+\cdots+\delta_n)\,.\end{aligned}$$

Summation and estimate shows the first claim. The second estimate follows from the first. □

14. Main theorem: Prerequisites: The function φ is Lipschitz-continuous in each component with Lipschitz constants K_i, i.e.,

$$|\varphi(u_\ell,\ldots,\hat u_i,\ldots,u_0) - \varphi(u_\ell,\ldots,u_i,\ldots,u_0)| \le K_i\cdot|\hat u_i - u_i|\,, \quad \text{for} \quad i=0,\ldots,\ell.$$

The powers of the matrix C_1 are bounded by D, $|C_1^\nu| \le D$, $\forall\nu\in\mathbb{N}$. Let ξ and $\hat\xi$ be

$$\xi := |P_1|\,D\,|R_1|\,K_\ell, \qquad \hat\xi := |P_1|\,D\,|R_1|\left(\sum_{i=0}^{\ell}K_i\right).$$

The value $\hat\xi$ is a multivariate function: $\hat\xi = \hat\xi(P_1, D, R_1, K_0,\ldots,K_\ell)$.

Assume

$$|h| < \begin{cases}1/\xi, & \text{if } \xi > 0;\\ \infty, & \text{if } \xi = 0.\end{cases}$$

Proposition: (1) The two difference equations

$$\begin{aligned}\hat u_{n+\ell} + A_{\ell-1}\hat u_{n+\ell-1} + \cdots + A_0\hat u_n &= h\,\varphi(\hat u_{n+\ell},\ldots,\hat u_n) + r_{n+\ell}\\ u_{n+\ell} + A_{\ell-1}u_{n+\ell-1} + \cdots + A_0u_n &= h\,\varphi(u_{n+\ell},\ldots,u_n)\end{aligned}$$

have a unique solution $u_{n+\ell}$ for each n , and $\hat{u}_{n+\ell}$.

(2) For the maximal norm deviation $|\hat{u}_n - u_n|$ we have the two-sided estimation with respect to the error terms r_n,

$$c_1 \left|\overline{\mathbf{P}}_1 [C_1]^{-1} \overline{\mathbf{R}}_1 \mathbf{R}\right| \le \left|\hat{U} - U\right| \le c_2 \left|\overline{\mathbf{P}}_1 [C_1]^{-1} \overline{\mathbf{R}}_1 \mathbf{R}\right| \le c_3 N \left|\mathbf{R}\right|.$$

We use $U = (u_1, \ldots, u_N)$, and $\hat{U} = (\hat{u}_1, \ldots, \hat{u}_N)$.

(3) The positive constants c_i, for $i = 1, 2, 3$, are given by

$$c_1 = \frac{1}{1 + \hat{\xi}|t_{\mathrm{end}} - t_0|}, \qquad c_2 = \frac{1}{1 - |h|\xi} \exp \frac{\hat{\xi}\,|t_{\mathrm{end}} - t_0|}{1 - |h|\xi}, \qquad c_3 = c_2 \left|P_1\right| D \left|R_1\right|.$$

(4) The estimate by (2) is independent from the choice of standard triple, i.e.,

$$\overline{\mathbf{X}}_1 [T_1]^{-1} \overline{\mathbf{Y}}_1 \mathbf{R} = \overline{\mathbf{X}}_2 [T_2]^{-1} \overline{\mathbf{Y}}_2 \mathbf{R},$$

for two arbitrary standard triples (X_1, T_1, Y_1) and (X_2, T_2, Y_2) for the matrix polynomial ρ.

(5) The reduced functional $\left|[C_1]^{-1} \overline{\mathbf{R}}_1 \mathbf{R}\right|$ is also a stability functional and equivalent to the unreduced functional, independent of N, i.e.,

$$\left|\overline{\mathbf{P}}_1 [C_1]^{-1} \overline{\mathbf{R}}_1 \mathbf{R}\right| \sim \left|[C_1]^{-1} \overline{\mathbf{R}}_1 \mathbf{R}\right|.$$

(6) Reduced stability functionals are each equivalent when changing standard triples. They are not necessarily equal. We have

$$\left|[T_1]^{-1} \overline{\mathbf{Y}}_1 \mathbf{R}\right| \sim \left|[T_2]^{-1} \overline{\mathbf{Y}}_2 \mathbf{R}\right|.$$

Proof For abbreviation, we use

$$\delta_{n+\ell} := \hat{u}_{n+\ell} - u_{n+\ell}, \qquad \hat{\delta}_{n+\ell} := \varphi(\hat{u}_{n+\ell}, \ldots, \hat{u}_n) - \varphi(u_{n+\ell}, \ldots, u_n).$$

For (1): For each n both difference equations can be written as

$$\hat{u}_{n+\ell} = \hat{F}(\hat{u}_{n+\ell}) := h\varphi(\hat{u}_{n+\ell}, \ldots\ ) + \hat{\psi}, \quad \text{resp.} \quad u_{n+\ell} = F(u_{n+\ell}) := h\varphi(u_{n+\ell}, \ldots\ ) + \psi,$$

These are contractive if $|h| K_\ell < 1$. Therefore, we have uniqueness due to $|h|\xi < 1$.

For (2): a) According to the representation theorem

$$\begin{aligned}
\left| P_1 C_1^n \begin{pmatrix} r_0 \\ \vdots \\ r_{\ell-1} \end{pmatrix} + P_1 \sum_{\nu=0}^{n-1} C_1^{n-1-\nu} R_1 r_{\nu+\ell} \right| &\le |\delta_n| + |h|\,|P_1|\,D\,|R_1| \sum_{\nu=0}^{n-\ell} \left|\hat{\delta}_{\nu+\ell}\right| \\
&\le |\delta_n| + |h|\,|P_1|\,D\,|R_1| \left(\sum_{i=0}^{\ell} K_i\right) \sum_{\nu=0}^{n-1} |\delta_\nu| \\
&\le |\delta_n| + |t_{\text{end}} - t_0|\,|P_1|\,D\,|R_1| \left(\sum_{i=0}^{\ell} K_i\right) \sup_{\nu=0}^{n-1} |\delta_\nu| \\
&\le \sup_{\nu=0}^{n} |\delta_\nu| + |t_{\text{end}} - t_0|\,|P_1|\,D\,|R_1| \left(\sum_{i=0}^{\ell} K_i\right) \sup_{\nu=0}^{n} |\delta_\nu| \\
&= \Big(1 + \hat{\xi}|t_{\text{end}} - t_0|\Big) \sup_{\nu=0}^{n} |\delta_\nu|.
\end{aligned}$$

Here we used

$$\sum_{\nu=0}^{n-\ell} \left|\hat{\delta}_{\nu+\ell}\right| \le K_\ell |\delta_n| + \left(\sum_{i=0}^{\ell-1} K_i\right) \left(\sum_{\nu=0}^{n-1} |\delta_\nu|\right)$$

of the above estimation theorem. We employed $N|h| = |t_{\text{end}} - t_0|$ and finally $\sum_{\nu=0}^{n} |\delta_\nu| \le N \sup_{\nu=0}^{n-1} |\delta_\nu|$. Multiplying by

$$\frac{1}{1 + \hat{\xi}\,|t_{\text{end}} - t_0|}$$

then this gives the first inequality from (2), and also gives c_1.

b) Again using the representation theorem and using norms

$$\begin{aligned}
|\delta_n| \le |h|\;|P_1|\,D\,|R_1| \sum_{\nu=0}^{n-\ell} \left|\hat{\delta}_{\nu+\ell}\right| + \left| P_1 C_1^n \begin{pmatrix} r_0 \\ \vdots \\ r_{\ell-1} \end{pmatrix} + P_1 \sum_{\nu=0}^{n-1} C_1^{n-1-\nu} R_1 r_{\nu+\ell} \right| \\
\le |h|\; \underbrace{|P_1|\,D\,|R_1| \left(\sum_{i=0}^{\ell} K_i\right)}_{=\hat{\xi}} \sum_{\nu=0}^{n-1} |\delta_\nu| + |h|\; \underbrace{|P_1|\,D\,|R_1|\,K_\ell}_{=\xi} |\delta_n| + \left|\overline{\mathbf{P}}_1 [C_1]^{-1} \overline{\mathbf{R}}_1 \mathbf{R}\right|,
\end{aligned}$$

We make use of the estimation theorem once more

$$(1 - |h|\xi)\,|\delta_n| \le |h|\hat{\xi} \sum_{\nu=0}^{n-1} |\delta_\nu| + \left|\overline{\mathbf{P}}_1 [C_1]^{-1} \overline{\mathbf{R}}_1 \mathbf{R}\right|.$$

As $|h| < 1/\xi$, we have $1 - |h|\xi > 0$. Using the discrete lemma of Gronwall and using

$$\varepsilon_{j+1} \leftarrow |\delta_n|, \qquad \eta_j \leftarrow \frac{\left|\overline{\mathbf{P}}_1 [C_1]^{-1} \overline{\mathbf{R}}_1 \mathbf{R}\right|}{1 - |h|\xi}, \qquad \delta \leftarrow \frac{\hat{\xi}}{1 - |h|\xi}, \qquad h_\nu \leftarrow |h|,$$

we get the estimate

$$|\delta_n| \le \frac{\left|\overline{\mathbf{P}}_1 [C_1]^{-1} \overline{\mathbf{R}}_1 \mathbf{R}\right|}{1 - |h|\xi} \exp \frac{\hat{\xi}\,|t_{\text{end}} - t_0|}{1 - |h|\xi}.$$

This shows the constant c_2. The constant c_3 results from a typical estimation.

For (4): The standard triple (X_1, T_1, Y_1) is similar to the standard triple (X_2, T_2, Y_2) iff

$$X_2 = X_1 S, \qquad T_2 = S^{-1} T_1 S, \qquad Y_2 = S^{-1} Y_1,$$

or

$$X_1 = X_2 S^{-1}, \qquad T_1 = S T_2 S^{-1}, \qquad Y_1 = S Y_2.$$

Now we have

$$\begin{aligned}
\overline{\mathbf{X}}_1 [T_1]^{-1} \overline{\mathbf{Y}}_1 \mathbf{R} &= \left(\operatorname*{diag}_{\nu=0}^{N} X_1 \right) [T_1]^{-1} \operatorname{diag} \left[\left(\operatorname*{row}_{i=0}^{\ell-1} T_1^i Y \right) B,\ \operatorname*{diag}_{\nu=1}^{N} Y_1 \right] \mathbf{R} \\
&= \left(\operatorname*{diag}_{\nu=0}^{N} (X_2 S^{-1}) \right) [S T_2 S^{-1}]^{-1} \operatorname{diag} \left\{ \operatorname*{row}_{i=0}^{\ell-1} \left[\left(S T_2 S^{-1} \right)^i S Y_2 \right] B,\ \operatorname*{diag}_{\nu=1}^{N} (S Y_2) \right\} \mathbf{R} \\
&= \left(\operatorname*{diag}_{\nu=0}^{N} X_2 \right) [T_2]^{-1} \operatorname{diag} \left[\left(\operatorname*{row}_{i=0}^{\ell-1} T_2^i Y \right) B,\ \operatorname*{diag}_{\nu=1}^{N} Y_2 \right] \mathbf{R} \\
&= \overline{\mathbf{X}}_2 [T_2]^{-1} \overline{\mathbf{Y}}_2 \mathbf{R}.
\end{aligned}$$

For (5): This was already proved.

For (6): As in (4)

$$\begin{aligned}
[T_1]^{-1} \overline{\mathbf{Y}}_1 \mathbf{R} &= [S T_2 S^{-1}]^{-1} \operatorname{diag} \left\{ \operatorname*{row}_{i=0}^{\ell-1} \left[\left(S T_2 S^{-1} \right)^i S Y_2 \right] B,\ \operatorname*{diag}_{\nu=1}^{N} S Y_2 \right\} \mathbf{R} \\
&= \left(\operatorname*{diag}_{\nu=0}^{N} S \right) [T_2]^{-1} \operatorname{diag} \left[\operatorname*{row}_{i=0}^{\ell-1} \left(T_2^i Y_2 \right) B,\ \operatorname*{diag}_{\nu=1}^{N} Y_2 \right] \overline{\mathbf{Y}}_2 \mathbf{R} \\
&= \left(\operatorname*{diag}_{\nu=0}^{N} S \right) [T_2]^{-1} \overline{\mathbf{Y}}_2 \mathbf{R}.
\end{aligned}$$

Multiplying from the left with $\operatorname{diag}_{\nu=0}^{N} S^{-1}$ then this gives

$$[T_2]^{-1} \overline{\mathbf{Y}}_2 \mathbf{R} = \left(\operatorname*{diag}_{\nu=0}^{N} S^{-1} \right) [T_1]^{-1} \overline{\mathbf{Y}}_1 \mathbf{R}.$$

Therefore, both stability functionals are equivalent. □

Similar results can be found in [2] and [3], who also analyzes the left eigenvectors, which under some circumstances allow a higher convergence rate.

With the main theorem we now have the promised convergence result for a consistent method, i.e., we have the classical result:

- Consistency + D-Stability ⇒ Convergence.

4. New cyclic linear multistep formulas

We now stick to cyclic linear multistep methods of the form

$$\sum_{j=-k+1}^{\ell} [\alpha_{ij} y_{m\ell+j} - h \beta_{ij} \dot{y}_{m\ell+j}] = 0, \quad i = 1, \dots, \ell.$$

Building on the work of [20] we combine the base formulas given in his chapter 4.2 and use the same restrictions R1-R4. Linear combinations of elements in the kernel of $C_{p,k}$ are also in the kernel.

15. Restrictions. Now for $i = 1, \dots, \ell$:

- R1. $\alpha_{ij} = \beta_{ij} = 0$, for $i < j$
- R2. $\beta_{ij} = 0$ for $j \leq 0$ and $\beta_{ii} \neq 0$
- R3. $\alpha_{ij} = 0$ for $j < -k + i$
- R4. $p_i \geq k$, where p_i is the consistency order of the i-th stage

By design of R1-R4 all Tendler-like cycles start with the BDF of the same order. Further, if the Tendler-like formula is $A[\alpha]$-stable, it is also automatically $A^0_\infty[\alpha]$-stable.

Restriction R3 has its origin in an implementation detail of STINT, where the predictor for $y_{m\ell+i}$ for each stage i in the cycle is built *only* from the backward differences of $y_{m\ell+i-1}$, see [21]. In principal, however, at each stage $i > 1$, all the previous values are available. In contrast, the Tischer formulas are not subject to the restriction R3.

16. Characteristics. The new cyclic linear multistep formulas are called eTendler3–9 (enhanced Tendler3–9). These kind of formulas are a natural extension of the BDF to cyclic form. From an implementation viewpoint, they are advantageous because each cycle requires no derivatives from the previous cycle. Therefore, no interpolation and storage for $f(t_n, y_n)$-values is required. In contrast, Tischer's formulas need storage and interpolation for both, y_n and $f(t_n, y_n)$.

It is worth noting that although the new formulas all start with BDFi and the BDFi are not D-stable for $i \geq 7$, the new cyclic formulas are all D-stable. In a cyclic method one or all constituent multistep methods might be unstable, nevertheless the cycle itself can be D-stable.

For Tischer's formulas see [23], [24], and [25]. All Tischer formulas have a cycle length of two.

While Tendler apparently searched in an entirely manual way, we searched in an incremental and random machine-assisted way. I.e., we searched in a hypercube of the parameter space by either providing a starting point, or letting those starting points be chosen by a random number generator. Our search criteria were Widlund-wedge α, Widlund-distance δ, and parasitic root modulus.

The formulas from Tendler from 1973 were our baseline. Clearly, we wanted to improve them, or at least find possible limits. So in table 1 we summarize their characteristics:

- p is the convergence order
- ℓ is the cycle length of the Tendler formulas
- abs(root) is the magnitude of the parasitic root
- α is the Widlund-wedge angle
- δ is the Widlund-distance

The new cyclic linear multistep formulas have better stability characteristics order by order than the original Tendler formulas. They in turn have better stability characteristics than the BDF.

17. Higher order. While a formula not being $A[\alpha]$-stable is of limited value for a pure stiff ODE solver, it is nevertheless of interest for a type-insensitive code, see [17] and [14], switching between fixed point and Newton iteration. The type-insensitive

Table 1 Widlund-wedge angle for Tendler and Tischer formulas

p	ℓ	abs(root)	α(Tendler)	δ(Tendler)	α(Tischer)	δ(Tischer)
1	3	0	90°	0	90°	0
2	3	0.333333333	90°	0	90°	0
3	3	0.55371901	89.427°	0.004776	90°	0
4	3	0.35406989	80.882047°	0.244157	90°	0
5	4	0.42931855	77.477315°	1.421472	86.649352°	0.040844
6	4	0.52827598	63.245842°	2.933167	76.311756°	0.280752
7	4	0.66669430	33.531759°	10.179501	57.663061°	0.959187
8	n/a	n/a	n/a	n/a	22.149242°	2.534082

Table 2 Widlund-wedge angle for new formulas

p	ℓ	abs(root)	α(new)	δ(new)	Comment
3	3	0.70756795	89.72423°	0.00164	worse root modulus, better δ
4	3	0.28351644	84.91216°	0.07106	better modulus, better α, better δ
5	**3**	0.48870093	77.81321°	0.42370	shorter cycle length, better α, better δ
6	4	0.29026688	71.63806°	1.03854	better modulus, better α, better δ
7	4	0.57300425	55.13529°	3.87902	better modulus, better α, better δ
8	4	0.61600197	none	15.05503	no other formula with $\alpha > 0.1$ found
9	**5**	0.76270334	none	38.22753	cycle length > 4 seems to be required

code LSODA from [15] has been found useful and competitive in [19]. Higher order methods are required for orbit calculations with higher precision, see [13] chapter 4.1.6.

The stability mountain of the new order 3 method eTendler3 is shown in the first figure.

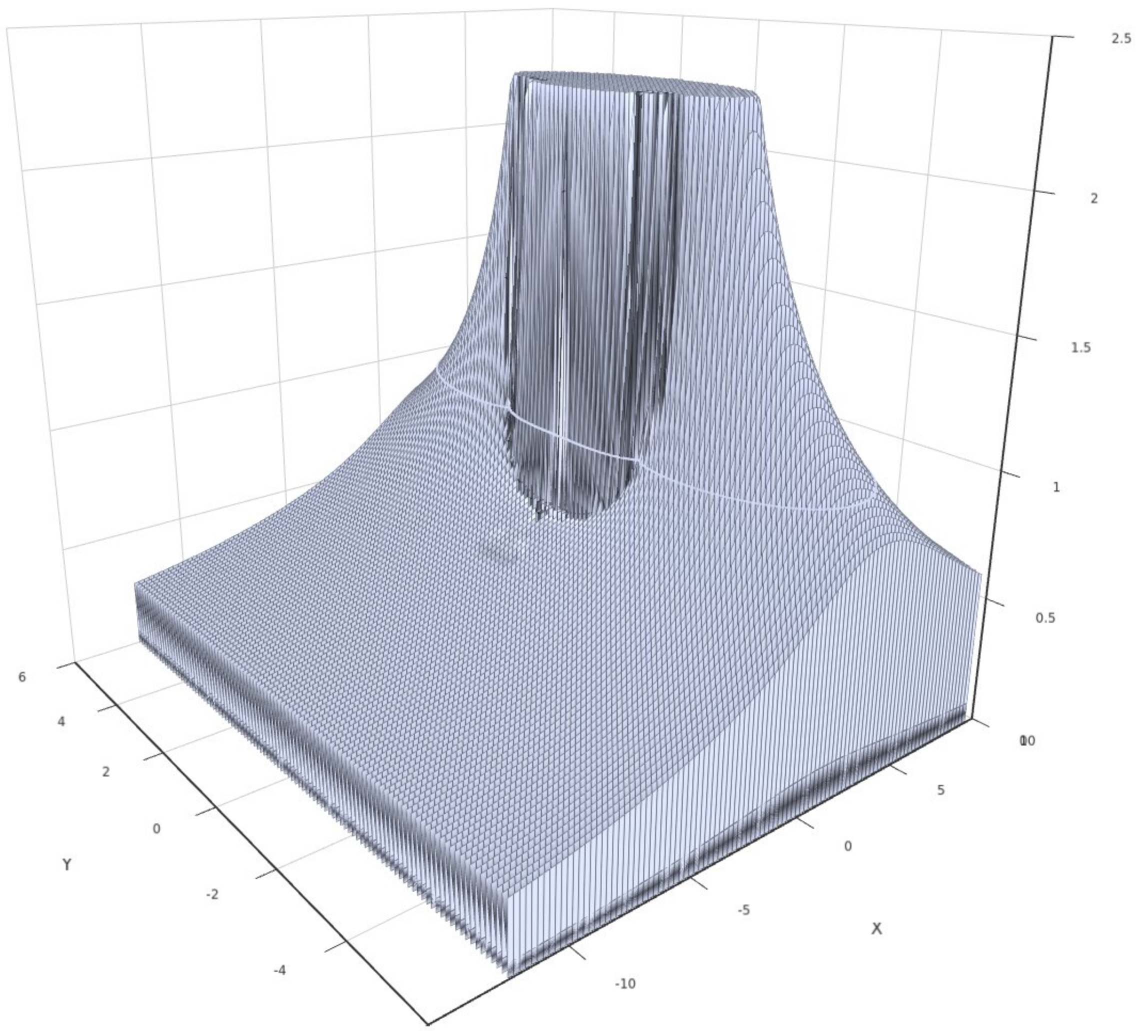

Fig. 1 Stability mountain of eTendler3

Our search results based on restrictions R1-R4 are as follows:

1. We didn't find any formula of order 8 and higher with an α of any significance, even when allowing for a huge number of stages. It is therefore conjectured that there aren't any methods of this kind.
2. We didn't find any formula of order 3 and 4 which actually is A-stable, in contrast to Tischer's results. We conjecture that there isn't any.

18. The formulas. $c_{i,p+1}$ is the unscaled error constant of the i-th stage. $\eta_{i,p+1}$ is the *scaled error constant* for the eTendler formulas. Likewise, $\eta^o_{i,p+1}$ are the scaled error constants for the original Tendler formulas, and $\eta^T_{i,p+1}$ are the scaled error constants for the Tischer formulas.

$$\eta_{i,p+1} = -\frac{1}{(p+1)!}\frac{1}{\alpha_{ii}}c_{i,p+1}$$

The minus-sign is only there as most formulas in this manuscript have negative values.

For linear multistep methods and cyclic methods the α_{ii} cannot be zero. However, for a block-implicit method the α_{ii} can be zero, see the order 2, 4, 6, 8 and 10 methods in [4]. More on the scaled error constant in [23], appendix B, and [3], chapter 6.

The error constants are rounded to five digits after the decimal point. The coefficients of the formulas are exact.

Order 3 and 4.

$p=3$	1	2	3
-2	-2	0	0
-1	9	-153	0
0	-18	750	-23
1	11	-1131	966
2	0	534	-1365
3	0	0	422
-2	0	0	0
-1	0	0	0
0	0	0	0
1	6	-246	-384
2	0	336	-378
3	0	0	264
$\eta_{i,p+1}$	0.13636	0.19569	0.15521
$\eta^o_{i,p+1}$	0.13636	0.13636	0.16667
$\eta^T_{i,p+1}$	1.24411	-0.56732	

$p=4$	1	2	3
-3	3	0	0
-2	-16	16	0
-1	36	-90	15
0	-48	234	-94
1	25	-214	162
2	0	54	-114
3	0	0	31
-3	0	0	0
-2	0	0	0
-1	0	0	0
0	0	0	0
1	12	-84	48
2	0	36	-60
3	0	0	24
$\eta_{i,p+1}$	0.096	0.21111	0.30323
$\eta^o_{i,p+1}$	0.096	0.096	0.10753
$\eta^T_{i,p+1}$	1.25579	-1.30782	

Order 5 and 6. $\eta^o_{4,p+1} = 0.10320$ for $p = 5$ where Tendler's formula needs four stages.

$p=5$	1	2	3
-4	-12	0	0
-3	75	-66	0
-2	-200	425	-93
-1	300	-1200	615
0	-300	2100	-1880
1	137	-1550	2460
2	0	291	-1515
3	0	0	413
-4	0	0	0
-3	0	0	0
-2	0	0	0
-1	0	0	0
0	0	0	0
1	60	-600	540
2	0	180	-540
3	0	0	240
$\eta_{i,p+1}$	0.07299	0.17182	0.16223
$\eta^o_{i,p+1}$	0.07299	0.07299	0.07210
$\eta^T_{i,p+1}$	1.13388	-1.34700	

$p=6$	1	2	3	4
-5	10	0	0	0
-4	-72	38	0	0
-3	225	-276	145	0
-2	-400	875	-1054	41
-1	450	-1600	3350	-289
0	-360	1950	-6200	830
1	147	-1388	7075	-1880
2	0	401	-4970	2935
3	0	0	1654	-1991
4	0	0	0	354
-5	0	0	0	0
-4	0	0	0	0
-3	0	0	0	0
-2	0	0	0	0
-1	0	0	0	0
0	0	0	0	0
1	60	-240	300	300
2	0	180	-600	-240
3	0	0	720	-600
4	0	0	0	180
$\eta_{i,p+1}$	0.05831	0.07838	0.07255	0.10048
$\eta^o_{i,p+1}$	0.05831	0.05900	0.05736	0.10557
$\eta^T_{i,p+1}$	0.94952	-0.92618		

Order 7.

$p=7$	1	2	3	4
-6	-60	0	0	0
-5	490	-280	0	0
-4	-1764	2310	-270	0
-3	3675	-8442	2233	-474
-2	-4900	18025	-8197	3920
-1	4410	-25200	17675	-14413
0	-2940	25830	-25550	31430
1	1089	-14910	23695	-42770
2	0	2667	-12383	36904
3	0	0	2797	-20615
4	0	0	0	6018
-6	0	0	0	0
-5	0	0	0	0
-4	0	0	0	0
-3	0	0	0	0
-2	0	0	0	0
-1	0	0	0	0
0	0	0	0	0
1	420	-4200	2100	-1680
2	0	1260	-2940	3360
3	0	0	1260	-2940
4	0	0	0	2520
$\eta_{i,p+1}$	0.04821	0.08718	0.07955	0.06539
$\eta^o_{i,p+1}$	0.04821	0.04821	0.06281	0.06765
$\eta^T_{i,p+1}$	0.82959	-0.70538		

Order 8.

$p=8$	1	2	3	4
-7	105	0	0	0
-6	-960	10560	0	0
-5	3920	-96740	4350	0
-4	-9408	396116	-40060	11580
-3	14700	-954618	165256	-106094
-2	-15680	1501850	-402822	434406
-1	11760	-1623860	646450	-1046346
0	-6720	1267140	-731500	1640450
1	2283	-701166	591360	-1801730
2	0	200718	-290706	1438794
3	0	0	57672	-782406
4	0	0	0	211346
-7	0	0	0	0
-6	0	0	0	0
-5	0	0	0	0
-4	0	0	0	0
-3	0	0	0	0
-2	0	0	0	0
-1	0	0	0	0
0	0	0	0	0
1	840	-56280	25200	21000
2	0	76440	-64680	2520
3	0	0	24360	-81480
4	0	0	0	81480
$\eta_{i,p+1}$	0.04088	0.04621	0.06424	0.04804
$\eta^T_{i,p+1}$	0.73907	-0.55211		

Order 9.

$p=9$	1	2	3	4	5
-8	-280	0	0	0	0
-7	2835	-5285	0	0	0
-6	-12960	53730	-13715	0	0
-5	35280	-246960	138885	-24780	0
-4	-63504	677376	-634992	250764	-22331
-3	79380	-1233036	1728720	-1145544	225768
-2	-70560	1569960	-3111108	3115434	-1029642
-1	45360	-1446480	3883740	-5600364	2789808
0	-22680	1028160	-3422160	6991530	-4946214
1	7129	-486351	2295792	-6110664	6531756
2	0	88886	-1194345	3889494	-5933718
3	0	0	329183	-2019384	3364992
4	0	0	0	653514	-1609983
5	0	0	0	0	629564
-8	0	0	0	0	0
-7	0	0	0	0	0
-6	0	0	0	0	0
-5	0	0	0	0	0
-4	0	0	0	0	0
-3	0	0	0	0	0
-2	0	0	0	0	0
-1	0	0	0	0	0
0	0	0	0	0	0
1	2520	-98280	-80640	-40320	-241920
2	0	35280	-63000	-73080	-168840
3	0	0	118440	35280	171360
4	0	0	0	229320	171360
5	0	0	0	0	216720
$\eta_{i,p+1}$	0.03535	0.05198	0.03743	0.03425	0.03217

5. Numerical results

We will now use various formulas and conduct more than 5000 numerical tests. Before implementing new formulas in a variable step size and variable order computer code with convergence tests, nonlinear equation solver, interpolation, dozens of heuristics, type-insensitive switching logic, etc., we want to ascertain that the formulas work as intended. Creating such a computer code is a substantial software development effort.

The first parameterized differential equation tests the resilience against stiffness. The second differential equations tests for accuracy in the presence of large modulus of the higher derivatives.

19. Dahlquist's equation. We tested the BDF, Tendler's formulas, new Tendler-like formulas, and Tischer's formulas on the classical test equation

$$\dot{y}(t) = \lambda\, y(t) \in \mathbb{C}, \quad y(t_0) = e^{\lambda t_0}, \quad t \in [t_0, t_{\mathrm{end}}].$$

We used

$$\lambda = re^{i\varphi}, \quad t_0 = 0, \quad t_{\mathrm{end}} = -40, \quad h < 0.$$

The radius is fixed to $r = 100$, and φ varies from 5° to 90° in steps of 5°. φ directly tests the Widlund-wedge angle of the formula.

The step size h for each formula is chosen as $h = -0.1$, then $h = -0.01$, finally $h = -0.001$.

The above differential equation is now solved with the following formulas:

1. BDF of order 1–6
2. Original Tendler formulas of order 3–7. Note: Tendler's formulas of order 1 and 2 are just the BDF1 and BDF2.
3. New Tendler-like formulas of order 3–9
4. Tischer's formulas of order 2–8 using $s = 0$, see [24] for the meaning of s

This creates 1350 data points. Multiplying by two hardware architectures and two precisions yields 5400 records.

20. Global error. We report the *summed global error* computed as

$$g_{\mathrm{err}} = \sum_{i=0}^{n} |y(t_i) - y_i|, \quad n = \frac{t_{\mathrm{end}} - t_0}{h}.$$

Computations were done in double precision (`double complex` in the C programming language). As the global error $|y(t_i) - y_i|$ per step varies wildly between 10^{-324} and 10^{+304}, we computed

$$\hat{g}_{\mathrm{err}} = \begin{cases} 5, & \text{if } g_{\mathrm{err}} > 30 \text{ or NaN} \\ \log_{10} g_{\mathrm{err}}, & \text{else} \end{cases}$$

Less is better. That is what is shown in the second figure.

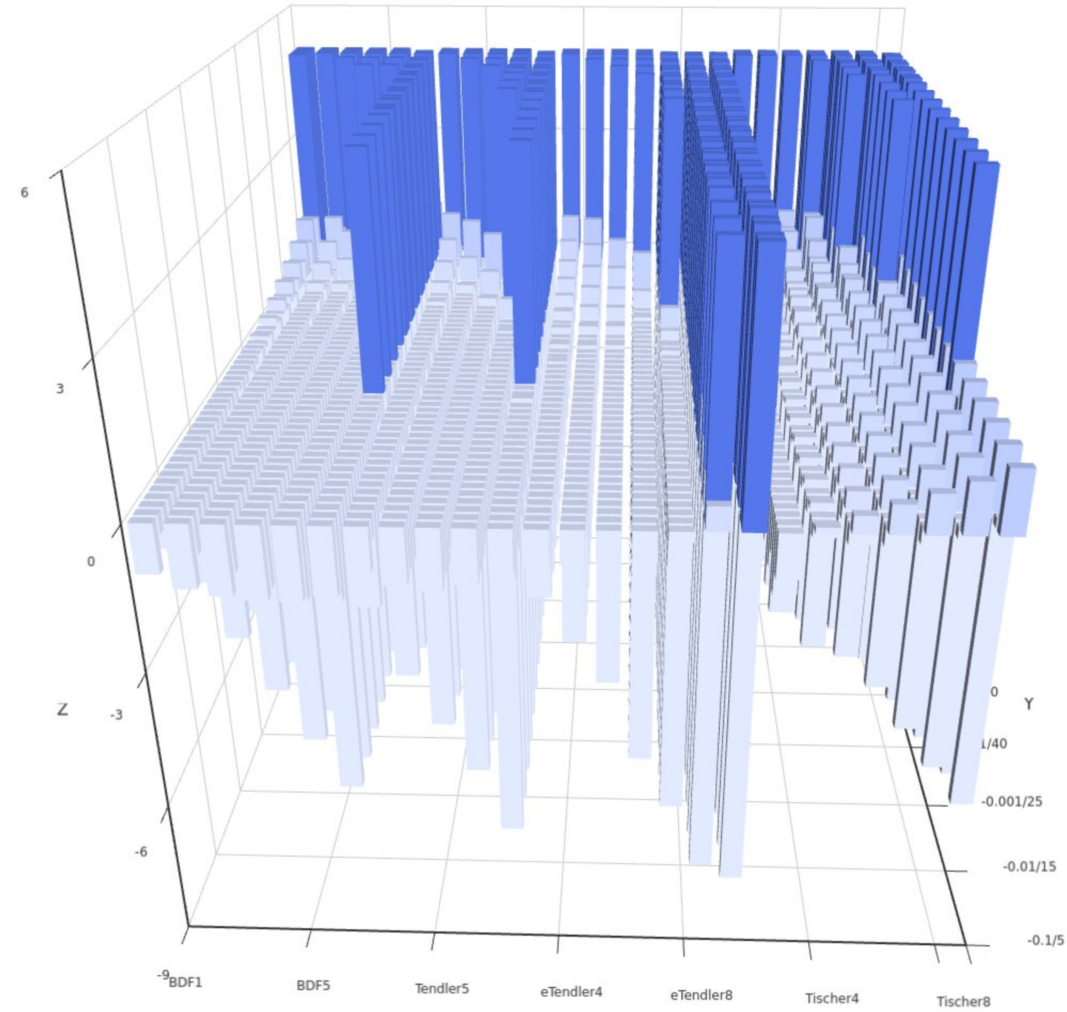

Fig. 2 Accuracy of BDF vs. Tendler vs. New vs. Tischer

One can clearly see that the higher-order methods quickly lose precision when a higher Widlund-wedge angle is required. This is the reason why *none* of the computer codes GEAR, EPISODE, LSODE and CVODE use BDF6. However, in reality, this is mainly a problem for the step size and order control segment to properly switch order.

The second figure also demonstrates that the Tischer formulas produce larger global errors. This is to be attributed to the error constants being an order of magnitude larger than the error constants of the Tendler formulas. That explains why the program ODIOUS, which uses the Tischer formulas, needs a smaller step size, and therefore needs more steps. This is reflected in the results in [25].

Since [6] noted a sensitivity regarding the chosen precision, especially in STINT, which implements Tendler's formulas, we repeated the above tests in single precision (`float complex` in the C language) and on two CPU architectures (AMD Ryzen 5700 and ARM Cortex A77). The qualitative results did not differ in any way. We henceforth conclude that the discontinuity and sensitivity of the step size and order control segment are the reason for the observed behavior in line with the remarks in [6].

21. Runge's equation. The test equation is

$$\dot{y}(t) = \frac{-2t}{(1+t^2)^2} \in \mathbb{R}, \quad y(t_0) = \frac{1}{1+t_0^2}, \quad t \in [t_0, t_{\text{end}}], \quad t_0 = -5, t_{\text{end}} = 5.$$

It has the exact solution $y(t) = 1/(1+t^2)$.

This is a numerical quadrature problem and not really a differential equation. This is Runge's function showcasing Runge's phenomenon.

For the numerical solution we used the step sizes $h = 0.1$, $h = 0.01$, and $h = 0.001$.

Running the same formulas from above creates 75 data points. Multiplying by two precisions gives 150 records.

All formulas produce accurate results and the error is always considerably smaller than one. We therefore show

$$\hat{g}_{\text{err}} = -\log_{10} g_{\text{err}}.$$

Higher is better. The third image shows the results for quadruple precision (`long double`).

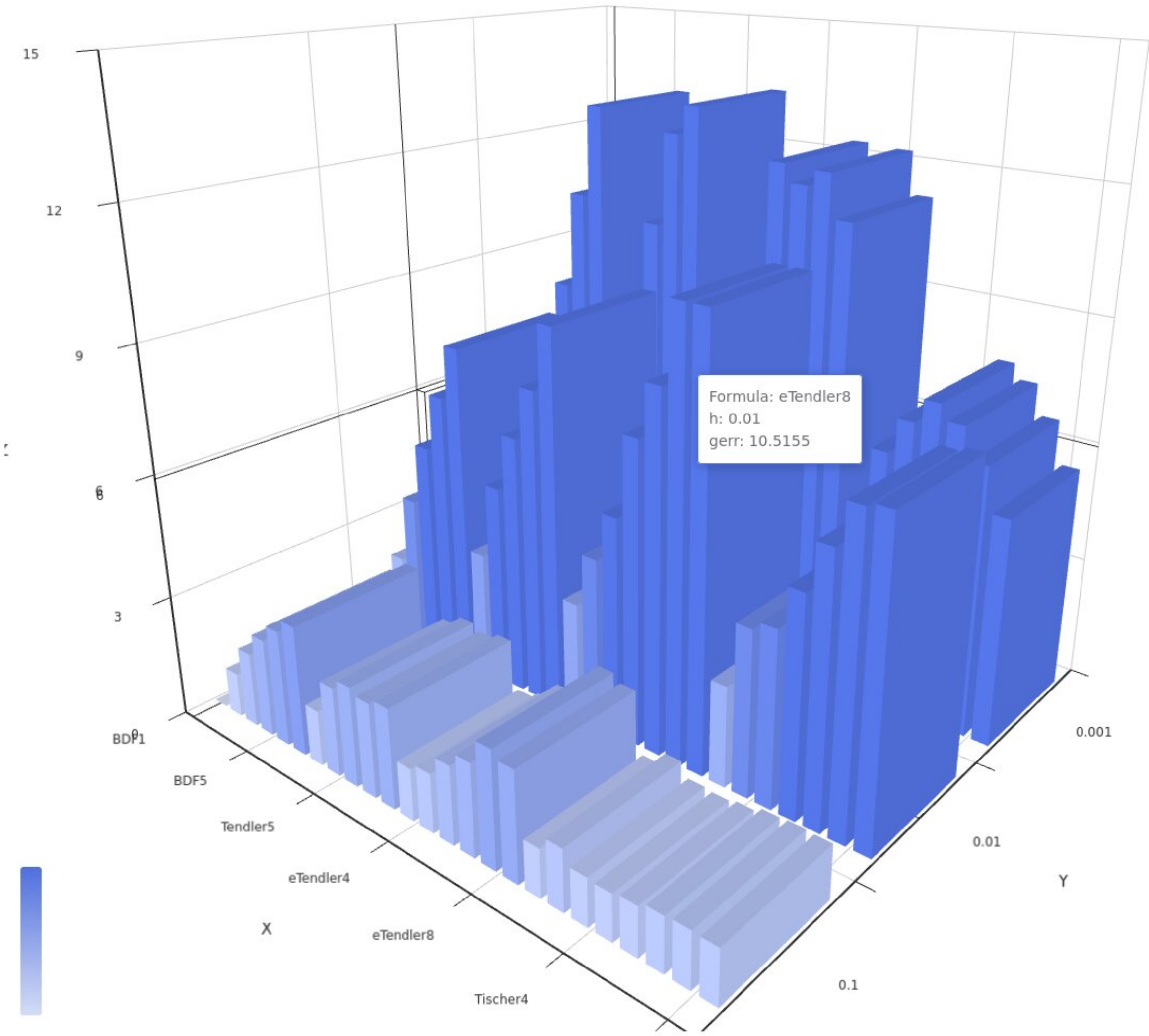

Fig. 3 Runge: Accuracy of BDF vs. Tendler vs. New vs. Tischer

From an accuracy point of view the formulas eTendler6-9 are outperformed by BDF6 and Tendler7 for the step size $h = 0.001$. For $h = 0.01$ eTendler8 shows the best accuracy.

The Tischer formulas again fall short of expectations and produce larger errors than all the other methods. This is to be explained by their larger error constants multiplied by the large magnitude of the higher derivatives of the Runge function. This is in line with the results reported in [25] where the ODIOUS program on average needs 50%-100% more steps and function evaluations than LSODE (based on BDF) excluding problem B5.

6. Summary and conclusions

We have given a convergence proof for general linear methods in the setting of matrix polynomials. Matrix polynomials are a versatile medium to analyze cyclic linear multistep methods.

We have extended the work of Tendler and found new cyclic linear multistep formulas with enhanced Widlund-wedge angle and Widlund-distance. It is conjectured that by lifting restriction R3 we might find even more enhanced formulas.

By comparing the error constants we explain why the Tischer formulas produce larger global errors.

Conflict of interest: There is no conflict of interest.

References

[1] Akrivis, G., Katsoprinakis, E.: Maximum angles of $A(\vartheta)$-stability of backward difference formulae. BIT Numer. Math. **60**, 93–99 (2020), https://doi.org/10.1007/s10543-019-00768-1

[2] Albrecht, P.: Explicit, optimal stability functionals and their application to cyclic discretization methods. Computing **19**, 233–249 (1978), https://doi.org/10.1007/BF02252202

[3] Albrecht, P.: Numerical treatment of O.D.Es.: The theory of A-methods. Numer. Math. **47**, 59–87 (1985), https://doi.org/10.1007/BF01389876

[4] Bickart, T.A., Picel, Z.: High order stiffly stable composite multistep methods for numerical integration of stiff differential equations. BIT Numer. Math. **13**, 272–286 (1973), https://doi.org/10.1007/BF01951938

[5] Butcher, J.C.: Numerical Methods for Ordinary Differential Equations. John Wiley & Sons, Chichester (2016), https://doi.org/10.1002/9781119121534

[6] Gaffney, P.: A Performance Evaluation of Some FORTRAN Subroutines for the Solution of Stiff Oscillatory Ordinary Differential Equations. ACM Trans. Math. Softw. **10**(1), 58–72 (1984), https://doi.org/10.1145/356068.356073

[7] Gohberg, I., Lancaster, P., Rodman, L.: Spectral Analysis of Matrix Polynomials — I. Canonical Forms and Divisors. Linear Algebra Its Appl. **20**, 1–44 (1978), https://doi.org/10.1016/0024-3795

[8] Gohberg, I., Lancaster, P., Rodman, L.: Matrix Polynomials. Society for Industrial and Applied Mathematics (2009), https://doi.org/10.1137/1.9780898719024

[9] Gupta, G.: Description and Evaluation of a Stiff ODE Code DSTIFF. SIAM J. Sci. Comput **6**(4), 939–950 (1985), https://doi.org/10.1137/0906063

[10] Hairer, E., Wanner, G., Nørsett, S.: Solving Ordinary Differential Equations I – Nonstiff Problems. Springer Berlin, Heidelberg (2008), https://doi.org/10.1007/978-3-540-78862-1

[11] Hairer, E., Wanner, G.: Solving Ordinary Differential Equations II – Stiff and Differential-Algebraic Problems. Springer Berlin, Heidelberg (2010), https://doi.org/10.1007/978-3-642-05221-7

[12] Mihelcic, M., Wingerath, K.: $A(\alpha)$-Stable Cyclic Composite Multistep Methods of Orders 6 and 7 for Numerical Integration of Stiff Ordinary Differential Equations. ZAMM **61**, 261–264 (1981), https://doi.org/10.1002/zamm.19810610609

[13] Montenbruck, O., Gill, E.: Satellite Orbits. Springer Berlin, Heidelberg (2000), https://doi.org/10.1007/978-3-642-58351-3

[14] Nørsett, S., Thomsen, P.G.: Switching between modified Newton and fix-point iteration for implicit ode-solvers. BIT Numer. Math. **26**, 339–348 (1986), https://doi.org/10.1007/BF01933714

[15] Petzold, L.: Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations. SIAM J. Sci. Comput. **4**(1), 136–148 (1983), https://doi.org/10.1137/0904010

[16] Shampine, L.F.: Type-sensitive ODE codes based on implicit $A(\alpha)$-stable formulas. Math. Comput. **39**(159), 109–123 (1982), https://doi.org/10.2307/2007622

[17] Skeel, R.: Analysis of Fixed-Stepsize Methods. SIAM J. Numer. Anal. **13**(5), 664–685 (1976), https://doi.org/10.1137/0713055

[18] Städter, P., Schälte, Y., Schmiester, L., Hasenauer, J., Stapor, P.L.: Benchmarking of numerical integration methods for ODE models of biological systems. Scientific Reports 11, 2696 (2021), https://doi.org/10.1038/s41598-021-82196-2

[19] Tendler, J.M.: A Stiffly Stable Integration Process Using Cyclic Composite Methods. Ph.D. thesis, Syracuse University, Syracuse, New York (1973)

[20] Tendler, J.M., Bickart, T.A., Picel, Z.: STINT: STiff ordinary differential equations INTegrator. Program and User Manual. Electrical and Computer Engineering Department, Syracuse University, Syracuse, New York 13210, USA, and Department of Electrical Engineering, Delft University of Technology, Delft, The Netherlands, Technical Report TR-76-12 (1976)

[21] Tendler, J.M., Bickart, T.A., Picel, Z.: A Stiffly Stable Integration Process Using Cyclic Composite Methods. ACM Trans. Math. Softw. **4**(4), 339–368 (1978),

https://doi.org/10.1145/356502.356495

[22] Tischer, P.: The Cyclic Use of Linear Multistep Formulas for the Solution of Stiff Differential Equations. Ph.D. thesis, Department of Computer Science, Monash University, Clayton, Victoria, Australia (1983)

[23] Tischer, P., Sacks-Davis, R.: A New Class of Cyclic Multistep Formulae for Stiff Systems. SIAM J. Sci. Comput. **4**(4), 733–747 (1983), https://doi.org/10.1137/0904051

[24] Tischer, P., Gupta, G.: An evaluation of some new cyclic linear multistep formulas for stiff ODEs. ACM Trans. Math. Softw. **11**(3), 263–270 (1985), https://doi.org/10.1145/214408.214417

[25] Werner, H., Arndt, H.: Gewöhnliche Differentialgleichungen. Springer Berlin, Heidelberg (1986), https://doi.org/10.1007/978-3-642-70338-6