NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

Error behaviour of exponential Runge-Kutta methods

by

Anne Kværnø

PREPRINT
NUMERICS NO. 6/2005

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/numerics/2005/N6-2005.ps
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

ERROR BEHAVIOUR OF EXPONENTIAL RUNGE-KUTTA METHODS

ANNE KVÆRNØ

Abstract. The error behaviour of exponential Runge-Kutta methods is studied by use of a modification of the B-series theory. In the modified series the stiffness is isolated from the elementary differentials and included into the coefficients of the series. This makes it possible to study the stiffness dependence of each term separately, and thereby gain better insight into how methods might behave when applied to stiff equations.

1. Introduction

The aim of this paper is to get a better understanding of the error behaviour of exponential Runge-Kutta (RK) methods. Such methods are derived to solve semilinear problems of the form

$$y' = Ly + f(t,y), \qquad y(t_0) = y_0 \tag{1}$$

where L is a matrix and $f(t,y)$ is some nonlinear term. We consider s-stage exponential RK-methods defined by

$$\begin{aligned} Y_i &= \mathrm{e}^{c_i hL} y_0 + h \sum_{j=1}^{s} a_{ij}(hL) f(t_0 + c_j h, Y_j), \qquad i = 1, 2, \cdots, s, \\ y_1 &= \mathrm{e}^{hL} y_0 + h \sum_{i=1}^{s} b_i(hL) f(t_0 + c_i h, Y_i). \end{aligned} \tag{2}$$

for which the coefficients of the methods are given in a Butcher tableau as

$$\begin{array}{c|ccc} c_1 & a_{11}(z) & \cdots & a_{1s}(z) \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1}(z) & \cdots & a_{ss}(z) \\ \hline & b_1(z) & \cdots & b_s(z) \end{array} \quad \text{or in matrix form as} \quad \begin{array}{c|c} c & A(z) \\ \hline & b^T(z) \end{array}$$

Exponential integrators were originally introduced as a method to solve stiff problems by use of explicit methods, without the familiar stability restrictions, see [4, 12, 14, 15]. The recent wave of publications seems to be more directed towards time integration of spatially discretized partial differential equations. A review of exponential integrators and related methods can be found in [13].

Classical local order theory has been derived by several authors, e.g. [1, 8, 10], based on B-series and rooted trees. Such analysis is crucial for proving consistency and convergence

Date: September 3, 2005.
2000 *Mathematics Subject Classification.* 65L05, 65L06.

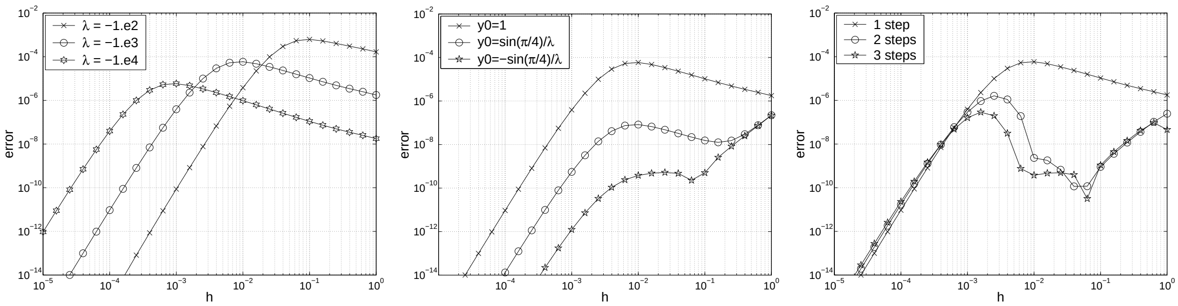


FIGURE 1. Stepsize versus error for the model equation (3).
Left: The effect of stiffness. Middle: The effect of initial values.
Right: How the error changes over the first few steps.

of the methods, but it is not able to describe how the error behaves for stiff problems using reasonable large stepsizes. Stiff order conditions can be found in [8, 9, 10].

In [5] Cox and Matthew describe and develop exponential integrators in the context of a simple model ODE for the evolution of a single Fourier mode, given by

$$y' = \lambda y + f(t, y), \qquad y(t_0) = y_0, \quad \lambda \in \mathbb{C}^-, \qquad |\lambda| >> 1. \tag{3}$$

The extension of the methods to systems of equations is immediate. But if the methods are applied to the simple problem (3), local error behaviour can be observed which can not be completely explained by any of the order theories mentioned above. The following example illustrates this.

Example 1. Consider the equation (3), with $f(t, y) = y + \sin(t)$ and $t_0 = \pi/4$. The error is measured after one step of the Cox and Matthews' 3th order exponential RK method, see Table 1, and the results are given in Figure 1. To the left, the effect of the stiffness parameter λ is demonstrated. In this case, the initial value $y_0 = 1$. The picture in the middle shows the effect of using different initial values. In this case λ is constant, $\lambda = -10^3$. To the right, we can see how the error changes over the first few steps, using $y_0 = 1$ and $\lambda = -10^3$.

The aim of this paper is to derive a theory to explicitly describe these kinds of error behaviour by rooted tree analysis. In the classical B-series theory, see [2, 6], the exact and the numerical solutions are written as series of the form

$$\sum_{\tau \in T} \alpha(\tau) h^{|\tau|} F(\tau)(t_0, y_0).$$

When applied to (3) elementary differential $F(\tau)(y_0)$ depends on powers of λ, making the truncated series unsuiteable for describing the error behaviour when $|\lambda|$ is large and h moderate.

For the model problem (3), we show that this problem can be overcomed by using a modification of the B-series, described roughly by

$$\sum_{\tau \in T} \varphi(\tau)(z) h^{|\tau|} F(\tau)(t_0, \cdot\,; y_0), \qquad z = \lambda h,$$

in which $F(\tau)(t_0, \cdot\,; y_0)$ is independent of λ, but the coefficients $\alpha(\tau)$ are replaced by functions of z, that is $\varphi(\tau)(z)$. The difference of these functions for the exact and the numerical solution then gives a quite precise description of the error for different values of z.

$$
\begin{array}{c|ccc}
0 & & & \\
\frac{1}{2} & \frac{1}{2}\phi_1(\frac{z}{2}) & & \\
1 & -\phi_1(z) & 2\phi_1(z) & \\
\hline
 & \phi_1(z) - 3\phi_2(z) + 4\phi_3(z) & 4\phi_2(z) - 8\phi_3(z) & -\phi_2(z) + 4\phi_3(z)
\end{array}
$$

CM3, s=3

$$
\begin{array}{c|cc}
\frac{1}{3} & \frac{1}{2}\phi_1(\frac{z}{3}) - \frac{1}{6}\phi_2(\frac{z}{3}) & \frac{1}{6}\phi_1(\frac{z}{3}) - \frac{1}{6}\phi_2(\frac{z}{3}) \\
1 & \frac{3}{2}\phi_1(z) - \frac{3}{2}\phi_2(z) & -\frac{1}{2}\phi_1(z) + \frac{3}{2}\phi_2(z) \\
\hline
 & \frac{3}{2}\phi_1(z) - \frac{3}{2}\phi_2(z) & -\frac{1}{2}\phi_1(z) + \frac{3}{2}\phi_2(z)
\end{array}
$$

Radau IIA, s=2

$$
\begin{array}{c|ccc}
0 & & & \\
\frac{1}{3} & \frac{1}{3}\phi_1(\frac{z}{3}) & & \\
\frac{2}{3} & 0 & \frac{2}{3}\phi_1(\frac{2z}{3}) & \\
\hline
 & \frac{8\mathrm{e}^z - 9\mathrm{e}^{\frac{2}{3}z} + 1}{8z} & 0 & \frac{9(\mathrm{e}^{\frac{2z}{3}} - 1)}{8z}
\end{array}
$$

CMO3, s=3

TABLE 1. Examples of exponential RK methods.

As an introduction, the idea is demonstrated on a linear problem in section 3. The main results, that is the series of the exact and numerical solutions of (3) ,are derived in section 4. In section 5, we study the series for the smooth solution, as well as situations when the initial value is on or off the smooth solution manifold. Section 6 presents numerical experiments to illuminate the relevance of the theoretical results. Some concluding remarks are given in section 7.

2. Examples of exponential Runge-Kutta methods.

Before studying the error behaviour, we would like to present some exponential RK-methods which will be used throughout the paper to illustrate the theoretical results. The methods are all of classical order 3, they have been chosen because of their diverse error behaviour when applied to the model problem. The methods are

CM3: The third order explicit method developed by Cox and Matthew [5].

Radau IIA: An implicit third order exponential integrator based on the Radau IIA quadrature, constructed by Hochbruck and Ostermann [9].

CMO3: A third order explicit method constructed by Celledoni, Marthinsen and Owren [3]. Contrary to the methods above, CMO3 was derived not primarely for solving (1), but as a Lie group method for solving ordinary differential equations on manifolds.

The tableaux of the methods are given in Table 1. The coefficients are expressed in terms of the function $\phi_q(z)$, defined by

$$
\phi_q(z) = \frac{1}{(q-1)!}\frac{1}{h^q}\mathrm{e}^{\lambda h}\int_0^h \mathrm{e}^{-\lambda\sigma}\sigma^{q-1}d\sigma = \frac{1}{z^q}\left(\mathrm{e}^z - \sum_{j=0}^{q-1}\frac{z^j}{j!}\right), \qquad q = 1, 2, \cdots, \tag{4}
$$

using $z = \lambda h$.

q	$\phi_q(z)$	CM3 $\psi_q(z)$	Radau IIA $\psi_q(z)$	CMO3 $\psi_q(z)$
1	$\frac{e^z-1}{z}$	$\frac{e^z-1}{z}$	$\frac{e^z-1}{z}$	$\frac{e^z-1}{z}$
2	$\frac{e^z-1-z}{z^2}$	$\frac{e^z-1-z}{z^2}$	$\frac{e^z-1-z}{z^2}$	$\frac{3e^{\frac{2}{3}z}-3}{4z}$
3	$\frac{e^z-1-z-\frac{z^2}{2}}{z^3}$	$\frac{e^z-1-z-\frac{z^2}{2}}{z^3}$	$\frac{(4-z)e^z-4-3z}{6z^2}$	$\frac{e^{\frac{2}{3}z}-1}{4z}$
4	$\frac{e^z-1-z-\frac{z^2}{2}-\frac{z^3}{6}}{z^4}$	$\frac{(6-z)e^z-6-5z-2z^2}{12z^3}$	$\frac{(13-4z)e^z-13-9z}{54z^2}$	$\frac{e^{\frac{2}{3}z}-1}{18z}$
5	$\frac{e^z-1-z-\frac{z^2}{2}-\frac{z^3}{6}-\frac{z^4}{24}}{z^5}$	$\frac{(14-3z)e^z-14-11z-4z^2}{96z^3}$	$\frac{(40-13z)e^z-40-27z}{648z^2}$	$\frac{e^{\frac{2}{3}z}-1}{108z}$

TABLE 2. Weight functions for the linear problem

3. The linear case

As an introduction to the idea of this paper, consider the the linear problem

$$y' = \lambda y + f(t), \qquad y(t_0) = y_0, \qquad \lambda \in \mathbb{C}^-. \tag{5}$$

The exact solution is given by

$$y(t_0+h) = e^{\lambda h} y_0 + e^{\lambda h}\int_0^h e^{-\lambda\sigma} f(t_0+\sigma)d\sigma.$$

By using the series expansion of $f(t_0+\sigma)$ and integrating each term separately we get

$$y(t_0+h) = e^z y_0 + \sum_{q=1}^{\infty} \phi_q(z) h^q f^{(q-1)}(t_0), \qquad z = \lambda h. \tag{6}$$

where $\phi_q(z)$ is given by (4).

A similar series can be derived for the numerical solution,

$$y_1 = e^z y_0 + h\sum_{i=1}^{s} b_i(z) f(t_0+c_i h) = e^z y_0 + \sum_{q=1}^{\infty}\psi_q(z) h^q f^{(q-1)}(t_0) \tag{7}$$

where

$$\psi_q(z) = \frac{1}{(q-1)!}\sum_{i=1}^{s} b_i(z) c_i^{q-1}.$$

Table 2 lists the functions ϕ_q as well as ψ_q for the methods given in Table 1. The local truncation error is given by

$$y(t_0+h) - y_1 = \sum_{q=1}^{\infty} \mathcal{E}_q(z) h^q f^{(q-1)}(t_0)$$

where the *error functions* $\mathcal{E}_q$ are given by

$$\mathcal{E}_q(z) = \phi_q(z) - \psi_q(z).$$

Obviously, the error is of order $\varrho+1$ independent of the stiffness parameter λ if

$$\mathcal{E}_q(z) = 0, \qquad q = 1, 2, \cdots \varrho,$$

and for the three methods under consideration

$$\varrho^{\text{CM3}} = 3, \qquad \varrho^{\text{Radau IIA}} = 2 \qquad \text{and} \qquad \varrho^{\text{CMO3}} = 1.$$

But useful information can be achieved by detailed examination of the error functions in the extreme cases, like the nonstiff, the strongly damped and the highly oscillatory case:

The nonstiff case. This situation is characterised by $|z|$ small, such that the error functions can be studied in terms of their series expansions. Since $z = \lambda h$, the nonstiff order of the local error is $p+1$ if

$$\mathcal{E}_q(z) = \mathcal{O}(z^{p-q+1}), \qquad q = \varrho+1, \cdots, p.$$

We then get the following expansion for $\mathcal{E}_q(z)$:

q	CM3	Radau IIA	CMO3
2	0	0	$\frac{1}{216}z^2 + \mathcal{O}(z^3)$
3	0	$\frac{1}{72}z + \mathcal{O}(z^2)$	$-\frac{1}{72}z + \mathcal{O}(z^2)$
4	$\frac{1}{720}z + \mathcal{O}(z^2)$	$-\frac{1}{216} + \mathcal{O}(z)$	$\frac{1}{216} + \mathcal{O}(z)$
5	$-\frac{1}{2880} + \mathcal{O}(z)$	$-\frac{1}{405} + \mathcal{O}(z)$	$\frac{7}{3240} + \mathcal{O}(z)$

which gives the following expressions for the errors:

$$y(x_0+h) - y_1 = \begin{cases} \big(\frac{\lambda}{720} f'''(t_0) - \frac{1}{2880} f^{(4)}(t_0)\big)\, h^5 + \mathcal{O}(h^6) & \text{for CM3} \\ \big(\frac{\lambda}{72} f''(t_0) - \frac{1}{216} f'''(t_0)\big)\, h^4 + \mathcal{O}(h^5) & \text{for RADAU IIA} \\ \big(\frac{\lambda^2}{216} f'(t_0) - \frac{\lambda}{72} f''(t_0) + \frac{1}{216} f'''(t_0)\big)\, h^4 + \mathcal{O}(h^5) & \text{for CMO3} \end{cases}.$$

Thus for $|\lambda|$ small, the methods Radau IIA and CMO3 will have approximately the same error behaviour. However, the λ^2 - term in the error of CMO3 will dominate for larger values of $|\lambda|$. For the linear problem, CM3 is of nonstiff order 5.

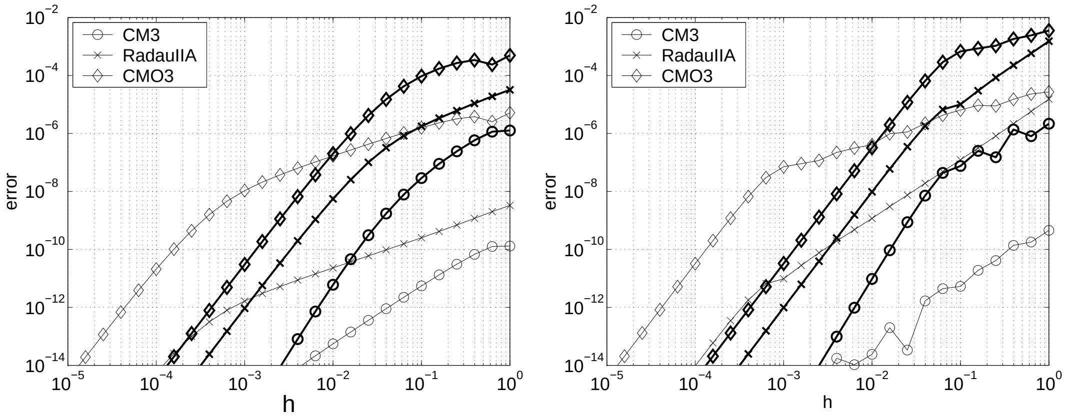
Rapid decay. In this case we assume $\text{Re}(z) << 0$, such that all transients represented by exponential functions are completely damped. The error functions can be written as

q	CM3	Radau IIA	CMO3
2	0	0	$-\frac{1}{4z} + \mathcal{O}\big(\frac{1}{z^2}\big)$
3	0	$-\frac{1}{3z^2} + \mathcal{O}\big(\frac{1}{z^3}\big)$	$-\frac{1}{4z} + \mathcal{O}\big(\frac{1}{z^2}\big)$
4	$-\frac{1}{12z^2} + \mathcal{O}\big(\frac{1}{z^3}\big)$	$-\frac{7}{27z^2} + \mathcal{O}\big(\frac{1}{z^3}\big)$	$-\frac{1}{9z} + \mathcal{O}\big(\frac{1}{z^2}\big)$
5	$-\frac{5}{96z^2} + \mathcal{O}\big(\frac{1}{z^3}\big)$	$-\frac{17}{162z^2} + \mathcal{O}\big(\frac{1}{z^3}\big)$	$-\frac{7}{216z} + \mathcal{O}\big(\frac{1}{z^2}\big)$

and the local truncation error is

$$y(t_0+h) - y_1 = \begin{cases} -\frac{1}{12\lambda^2} h^2 f'''(t_0) + \mathcal{O}\big(\frac{h}{\lambda^3} + \frac{h^3}{\lambda^2}\big) & \text{for CM3} \\ -\frac{1}{3\lambda^2} h f''(t_0) + \mathcal{O}\big(\frac{1}{\lambda^3} + \frac{h^2}{\lambda^2}\big) & \text{for Radau IIA} \\ -\frac{1}{\lambda} h f'(t_0) + \mathcal{O}\big(\frac{1}{\lambda^2} + \frac{h^2}{\lambda}\big) & \text{for CMO3} \end{cases}.$$

Only CM3 has a stiff local order of 2. Of the remaining methods, Radau IIA has an advantage because of the $1/\lambda^2$ dependency.



$\lambda = -10^{-2}$ (**——**), $\lambda = -10^{-4}$ (——). $\lambda = 10^{-2}\, i$ (**——**), $\lambda = 10^{-4}\, i$ (——).

FIGURE 2. Local truncation error.

Rapid oscillations. At last we assume $|z|$ large and λ imaginary. The exponentials will represent rapid oscillations in the error functions, which are dominated by the terms:

q	CM3	Radau IIA	CMO3
2	0	0	$-\frac{3e^{\frac{2}{3}z}+1}{4z} + \mathcal{O}\big(\frac{1}{z^2}\big)$
3	0	$\frac{e^z}{6z} + \mathcal{O}\big(\frac{1}{z^2}\big)$	$-\frac{e^{\frac{2}{3}z}+1}{4z} + \mathcal{O}\big(\frac{1}{z^2}\big)$
4	$\frac{e^z-1}{12z^2} + \mathcal{O}\big(\frac{1}{z^3}\big)$	$\frac{2e^z}{27z} + \mathcal{O}\big(\frac{1}{z^2}\big)$	$-\frac{e^{\frac{2}{3}z}+2}{18z} + \mathcal{O}\big(\frac{1}{z^2}\big)$
5	$\frac{3e^z-5}{96z^2} + \mathcal{O}\big(\frac{1}{z^3}\big)$	$\frac{13e^z}{648} + \mathcal{O}\big(\frac{1}{z^2}\big)$	$-\frac{2e^{\frac{2}{3}z}+7}{216z} + \mathcal{O}\big(\frac{1}{z^2}\big)$

and the absolute value of the local truncation error is

$$|y(t_0+h) - y_1| = \begin{cases} \frac{1}{12|\lambda|^2} M h^2 f'''(t_0) + \mathcal{O}\big(\frac{h^3}{|\lambda|^2} + \frac{h}{|\lambda|^3}\big), & \text{with } M \in [0,2] \quad \text{for CM3} \\ \frac{1}{6|\lambda|} h^2 f''(t_0) + \mathcal{O}\big(\frac{h^3}{|\lambda|} + \frac{h}{|\lambda|^2}\big) & \quad\quad\quad\quad\quad\quad\quad \text{for Radau IIA} \\ \frac{1}{2|\lambda|} M h f'(t_0) + \mathcal{O}\big(\frac{h^2}{|\lambda|} + \frac{1}{|\lambda|^2}\big), & \text{with } M \in [1,2] \quad \text{for CMO3.} \end{cases}$$

Radau IIA method is the only of the three methods that will demonstrate a significant different error behaviour in the decaying and the oscillating case.

The theoretical results are confirmed by the following experiment:

Example 2. Consider the example given by Cox and Matthew [5]:

$$y' = \lambda y + \sin(t), \qquad y(0) = y_0$$

with exact solution

$$y(t) = y_0 e^{\lambda t} + \frac{e^{\lambda t} - \lambda \sin t - \cos t}{1 + \lambda^2}.$$

The equation is solved by each of the three exponential Runge-Kutta methods, starting from $t_0 = \pi/4$, and the error is measured after one step. Two real and two imaginary values of λ have been used to represent the two stiff cases. The results are presented in Figure 2.

Even for this simple example, we observe that the error behaviour depends on both λ and h. We also observe that the theory gives a quite precise description of the observed results.

4. The semilinear case

We will now consider the semilinear scalar equation

$$y' = \lambda y + f(t,y), \quad y(t_0) = y_0, \qquad \lambda \in \mathbb{C}^-. \tag{8}$$

Assuming f sufficiently smooth, (8) possesses a smooth solution which can be written as an asymptotic series in powers of $1/\lambda$:

$$y_s(t) = y_{s,0}(t) + \frac{1}{\lambda} y_{s,1}(t) + \frac{1}{\lambda^2} y_{s,2}(t) + \cdots . \tag{9a}$$

By inserting the series of $y_s(t)$ into (8) and comparing equal powers of $1/\lambda$, we obtain the following expressions for the first three terms

$$y_{s,0}(t) = 0, \quad y_{s,1}(t) = -f, \quad y_{s,2}(t) = -f_t + f_y f \quad \cdots , \tag{9b}$$

where the function f and its differentials are all evaluated in $(t, y_{s,0}(t)) = (t, 0)$. The solution $y_s(t)$ is frequently referred to as the "smooth solution manifold" of (8). If $\mathrm{Re}(\lambda) < 0$ solutions are attracted to the smooth manifold, if λ is purely imaginary, solutions oscillates around it. An initial value is called consistent if $y_0 = y_s(t_0)$.

As in the linear case, we would like to express the exact and the numerical solution as power series of h, in which the coefficients of the series are functions of z. However, from the Runge-Kutta theory, we know that each term corresponding to a certain power of h might split into several terms, so that the exact and the numerical solution are better represented by B-series, [2, 7]. These considerations motivate the following preliminary definition:

Definition 1. A *GB-series* is a formal series in h of the form

$$GB(\varphi, Q_0; z, h) = \sum_{\tau \in T} \varphi(\tau)(z)\, h^{|\tau|} F(\tau)(Q_0), \tag{10}$$

where

- T is an index set.
- $\varphi(\tau)(z)$ is a function of z (elementary weight function).
- $|\tau|$ is a non-negative integer (order).
- $F(\tau)(Q_0)$ is a differential operator on f (elementary differential), evaluated at some point Q_0.

Our aim is to construct GB-series of the exact solution $y(t_0 + h)$ of (8) as well as the numerical solution y_1 given by (2) using $L = \lambda$, that is

$$y(t_0 + h) = GB(\varphi_e, Q_0; h\lambda, h, y_0) = \sum_{\tau \in T} \varphi_e(\tau)(h\lambda)\, h^{|\tau|} F(\tau)(Q_0), \tag{11}$$

$$y_1 = GB(\psi, Q_0; h\lambda, h, y_0) = \sum_{\tau \in T} \psi(\tau)(h\lambda)\, h^{|\tau|} F(\tau)(Q_0). \tag{12}$$

such that the elementary differentials $F(\tau)(Q_0)$ of f are independent on the stiffness parameter λ.

Obviously, the exact and numerical solution of the linear equation, given by (6) and (7), can both be considered as GB-series. In this case T is the set of all nonnegative integers and $Q_0 = t_0$. In the semilinear case, it will be more convenient to let T be a certain set of multicoloured rooted trees. As in (9) the elementary differentials will usually be evaluated at $(t_0, y_{s,0}(t_0)) = (t_0, 0)$. We will then use the notation $Q_0 = (t_0, 0; y_0)$ where y_0 is the initial value of the problem.

The following lemma, which can be established with a minimum knowledge of the GB-series, is a prerequisite for the construction of the series (11) and (12).

Lemma 1. *Assume that $\tilde{y}(h)$ can be expressed as a GB-series around $Q_0 = (t_0, 0; y_0)$, that is $\tilde{y}(h) = GB(\varphi, Q_0; z, h)$. Then the function $f(t + \gamma h, \tilde{y}(h))$ can be written as a formal series*

$$f(t_0 + \gamma h, \tilde{y}(h)) = \sum_{w \in W} \alpha_f(w) \varphi_f(w)(z) h^{|w|_f} F_f(w)(Q_0) \tag{13}$$

where the index set W and the terms α, ψ_f, $|\cdot|_f$ and F_f are given as follows::

a) W is the collection of multisets of indices so that
- $\{\emptyset\} \in W$
- $w = \{\bullet^k, \tau_1, \cdots, \tau_m\} \in W$ *if* $\tau_1, \cdots, \tau_m \in T$.

Here $\bullet$ is one particular index, but $\bullet \notin T$.

In the following, $w = \{\bullet^k, \tau_1, \cdots, \tau_m\} \in W$ and among $\tau_1, \cdots, \tau_m$ there are q distinct elements, each of multiplicity r_j, $j = 1, \cdots, q$. Then

b) $\alpha_f(\{\emptyset\}) = 1, \qquad \alpha_f(w) = \dfrac{1}{k!} \displaystyle\prod_{j=1}^{q} \dfrac{1}{r_j!}.$

c) $\varphi_f(\{\emptyset\}) = 1, \qquad \varphi_f(w) = \gamma^k \displaystyle\prod_{i=1}^{m} \varphi(\tau_i).$

d) $|\{\emptyset\}|_f = 0, \qquad |w|_f = k + \displaystyle\sum_{i=1}^{m} |\tau_i|.$

e) $F_f(\{\emptyset\}) = f(t_0, 0), \qquad F_f(w)(Q_0) = \dfrac{\partial^{k+m} f}{\partial t^k \partial y^m}(t_0, 0)\Big(F(\tau_1)(Q_0), \cdots, F(\tau_m)(Q_0)\Big).$

Proof. Let z be fixed, and do a multivariable Taylor expansion of f around $(t_0, 0)$. Then

$$\begin{aligned} f(t_0 + \gamma h, \tilde{y}(h)) &= \sum_{p=0}^{\infty} \sum_{k+m=p} \frac{1}{k!m!} f_{ktmy}(t_0, 0)(\gamma h)^k (\tilde{y}(h))^m \\ &= \sum_{p=0}^{\infty} \sum_{k+m=p} \frac{1}{k!m!} f_{ktmy}(t_0, 0)(\gamma h)^k \Big(\sum_{\tau \in T} \varphi(\tau)(z) h^{|\tau|} F(\tau)(Q_0)\Big)^m \\ &= \sum_{p=0}^{\infty} \sum_{k+m=p} \frac{1}{k!m!} f_{ktmy}(t_0, 0)(\gamma h)^k \sum_{\tau_1, \cdots, \tau_m \in T} \prod_{i=1}^{m} \varphi(\tau_i)(z) h^{|\tau_i|} F(\tau_i)(Q_0) \end{aligned}$$

The last sum is here taken over all possible ordered sets of $\tau_1, \cdots, \tau_m$. If this set consists of q distinct elements, each of multiplicity r_j, then the same term appear exactly $m!/(r_1! \cdots r_q!)$ times. Using the multilinearity of the differentials f_{ktmy} we get

$$\begin{aligned} &f(t + \gamma h, \tilde{y}(h)) = f(t_0, 0) \\ &\quad + \sum_{\{\bullet^k, \tau_1, \cdots, \tau_m\} \in W} \frac{1}{k! r_1! \cdots r_q!} \gamma^k \Big(\prod_{i=1}^{m} \varphi(\tau_i)(z)\Big) h^{k + \sum_{i=1}^{m} |\tau_i|} f_{ktmy}(t_0, 0)(F(\tau_1)(Q_0) \cdots F(\tau_m)(Q_0)) \end{aligned}$$

where $\bullet^k$ represents the k times derivatives of f with respect to t. Comparing this with (13) proves the lemma. $\square$

Remark 1. In the transient case it might be advantageous to do the expansions around (t_0, y_0) rather than around $(t_0, 0)$. This can be done by writing the GB-series as

$$\tilde{y}(h) = y_0 + \varphi(\star)(z)y_0 + \sum_{\tau \in T \setminus \{\star\}} \varphi(\tau)(z)h^{|\tau|}F(\tau)(Q_0), \qquad Q_0 = (t_0, y_0; y_0),$$

and $\star$ is the index representing y_0. In this case, all the elementary differentials is computed in (t_0, y_0), but the lemma is not altered in any other ways.

The next step is to derive the GB-series of the exact solution, (11). In the following, we will use $Q_0 = (t_0, 0; y_0)$. The series will be derived from the the variation-of-constants formula, which gives the exact solution of (8) as

$$y(t_0 + h) = \mathrm{e}^{\lambda h} y_0 + \mathrm{e}^{\lambda h} \int_0^h \mathrm{e}^{-\lambda \varsigma} f(t_0 + \varsigma, y(t_0 + \varsigma)) d\varsigma. \tag{14}$$

Immediately, we can conclude that one term in the series is $\mathrm{e}^z y_0$, or more formally

$$\star \in T, \quad \varphi_e(\star)(z) = \mathrm{e}^z, \quad |\star| = 0 \quad \text{and} \quad F(\star)(Q_0) = y_0. \tag{15}$$

We also have that $f(t_0 + \tau, y(t_0 + \tau)) = f(t_0, 0) + \cdots$, thus a second term is given by

$$\mathrm{e}^z \int_0^h \mathrm{e}^{-\lambda \varsigma} f(t_0, 0) d\varsigma = \frac{\mathrm{e}^z - 1}{z} h f(t_0, 0)$$

so that

$$\circ \in T, \quad \varphi_e(\circ)(z) = \frac{\mathrm{e}^z - 1}{z}, \quad |\circ| = 1 \quad \text{and} \quad F(\circ)(Q_0) = f(t_0, 0). \tag{16}$$

The complete GB-series of the exact solution is given by the following theorem:

Theorem 1. *The exact solution $y(t_0 + h)$ of (8) can be written as a GB-series of the form (11) around $Q_0 = (t_0, 0; y_0)$, with*

a) $\star \in T$, $\circ \in T$ and $[\bullet^k, \tau_1, \cdots, \tau_m] \in T$ if $\tau_1, \cdots, \tau_m \in T$.

In the following, $\tau = \{\bullet^k, \tau_1, \cdots, \tau_m\} \in T$ and among $\tau_1, \cdots, \tau_m$ there are q distinct elements, each of multiplicity r_j, $j = 1, \cdots, q$. Then

b) $F(\star)(Q_0) = y_0$, $F(\circ)(Q_0) = f(t_0, 0)$ and

$$F(\tau)(Q_0) = f_{ktmy}(t_0, 0)\big(F(\tau_1)(Q_0) \cdots F(\tau_m)(Q_0)\big),$$

c) $|\star| = 0$, $|\circ| = 1$ and $|\tau| = k + \sum_{i=1}^{m} |\tau_i| + 1$,

d) $\varphi_e(\star)(z) = e^z$, $\varphi_e(\circ)(z) = \dfrac{e^z - 1}{z}$. Otherwise, $\varphi_e(\tau)(z)$ is given by

$$\varphi_e(\tau)(z) = \frac{1}{z^{|\tau|}} e^z \int_0^z \sigma^{|\tau| - 1} e^{-\sigma} \Gamma(\tau)(\sigma) d\sigma,$$

where

$$\Gamma(\tau)(z) = \alpha(\tau) \prod_{i=1}^{m} \varphi_e(\tau_i)(z), \qquad \alpha(\tau) = \frac{1}{k!} \prod_{j=1}^{q} \frac{1}{r_j!}.$$

Proof. The theorem will be proved by induction. Assume that it is possible to write the exact solution as some GB-series (11). Then the first two terms are given by (15) and (16). For the remaining terms, replace $y(t_0+\varsigma)$ in the integral of (14) by its GB-series and use Lemma 1. Thus (14) can be written as

(17)
$$\sum_{\tau\in T}\varphi_e(\tau)(z)h^{|\tau|}F(\tau)(Q_0) = \mathrm{e}^z y_0 + \mathrm{e}^z\int_0^h \mathrm{e}^{-\lambda\varsigma}\sum_{w\in W}\alpha_f(w)\varphi_f(w)(\lambda\varsigma)\tau^{|\tau|_f}F_f(w)(Q_0)d\varsigma$$
$$= \mathrm{e}^z y_0 + \sum_{w\in W}\left(\frac{\mathrm{e}^z}{z^{|w|_f+1}}\int_0^z \sigma^{|w|_f}\mathrm{e}^{-\sigma}\alpha_f(w)\varphi_f(w)(\sigma)d\sigma\right)h^{|w|_f+1}F_f(w)(Q_0).$$

Each term on the left hand side have a corresponding term on the right. Assume that $\tau_1,\tau_2,\cdots,\tau_m\in T$ and their corresponding functions F, $|\cdot|$ and φ_e are given by the theorem. Then $w=\{\bullet^k,\tau_1,\cdots,\tau_m\}\in W$, and there must be a corresponding term on the left hand side represented by an index $\tau\in T$, which conveniently will be denoted by $\tau=[\bullet^k,\tau_1,\cdots,\tau_m]$. Comparing the terms on each side, we get

$$|\tau| = |w|_f+1,\quad F(\tau)(Q_0)=F_f(w)(Q_0)$$
$$\text{and}\quad \varphi_e(\tau) = \frac{\mathrm{e}^z}{z^{|w|_f+1}}\int_0^z \sigma^{|w|_f}\mathrm{e}^{-\sigma}\alpha_f(w)\varphi_f(w)(\sigma)d\sigma.$$

Inserting the expressions for $|\cdot|_f$, F_f, α_f and φ_f from Lemma 1 completes the proof. □

Remark 2. If the elementary differentials are evaluated in $(t_0,y_0;y_0)$ rather than in $(t_0,0;y_0)$, then the GB-series for the exact solution can be written as

$$y(t_0+h) = y_0 + (\mathrm{e}^z-1)y_0 + \sum_{\tau\in T\setminus\{\star\}}\varphi_e(\tau)(z)h^{|\tau|}F(\tau)(Q_0)$$

Thus $\varphi_e(\star)(z) = \mathrm{e}^z-1$, consequently one of the initial values for the recursion is changed, but the reccurence formula itself is unaltered.

Using standard notation, an index $\tau=[\bullet^k,\tau_1,\cdots,\tau_m]\in T$ is associated with a rooted tree obtained by connecting k $\bullet$'s and the roots of $\tau_1,\cdots,\tau_m$ by $k+m$ branches to a new white vertex $\circ$ which becomes the root of τ. In Table 3 the trees with less than four vertices are listed, together with their corresponding terms.

The last step is to prove that the numerical solution can be written as GB-series similar to those of the exact solution. The formula (2) applied to (3) can be written more compactly as

$$Y = \mathrm{e}^{cz}y_0 + hA(z)f(t_0+ch,Y),$$
$$y_1 = \mathrm{e}^z y_0 + hb^T(z)f(t_0+ch,Y), \tag{18}$$

by using $Y=[Y_1,\cdots,Y_s]^T$ for the vector of stage values, $\mathrm{e}^{cz}=[\mathrm{e}^{c_1z},\cdots,\mathrm{e}^{c_sz}]^T$ and $f(t_0+ch,Y)=[f(t_0+c_1h,Y_1),\cdots,f(t_0+c_sh,Y_s)]^T$. Further $\mathbb{1}_s=[1,1,\cdots,1]^T\in\mathbb{R}^s$. The multiplication of a product of vectors by a matrix is defined by: If A is an $s\times s$ matrix and d_i, $i=1,\cdots,m$ are vectors of dimension s, then

$$A\prod_{i=1}^m d_i = A\left(\prod_{i=1}^m d_i\right).$$

That is, form the element-by-element product of the vectors before multiplying by A. We now establish the following result:

Theorem 2. *The stage vector Y as well as the numerical solution y_1 given by* (18) *can be expressed as GB-series similar to those of the exact solution of* (8) *, that is*

$$Y = GB(\Psi, Q_0; z, h), \qquad y_1 = GB(\psi, Q_0; z, h)$$

with Q_0, T, F and $|\cdot|$ given by Theorem 1, and

$$\Psi(\star)(z) = e^{cz}, \quad \Psi(\circ)(z) = A(z)\mathbb{1}_s \quad \textit{and} \quad \Psi(\tau)(z) = \alpha(\tau)\ A(z)\Big(c^k \prod_{i=1}^{m} \Psi(\tau_i)\Big),$$

$$\psi(\star) = e^{z}, \quad \psi(\circ) = b^T(z)\mathbb{1}_s \quad \textit{and} \quad \psi(\tau) = \alpha(\tau)\ b^T(z)\Big(c^k \prod_{i=1}^{m} \Psi(\tau_i)\Big),$$

where $\tau = [\bullet^k, \tau_1, \cdots, \tau_m] \in T$ with q distinct subtrees τ_j, each of multiplicity r_j, and $\alpha(\tau) = \frac{1}{k!}\prod_{j=1}^{q}\frac{1}{r_j!}$.

Proof. Assume that the GB-series of Y and y_1 exist. Insert those into (18) and use Lemma 1. We then get

$$\sum_{\tau\in T} \Psi(\tau)(z)h^{|\tau|}F(\tau)(Q_0) = \mathrm{e}^{cz}y_0 + A(z)\sum_{w\in W} \alpha_f(w)\Psi_f(w)h^{|w|_f+1}F_f(w)(Q_0),$$

$$\sum_{\tau\in T} \psi(\tau)(z)h^{|\tau|}F(\tau)(Q_0) = \mathrm{e}^{z}y_0 + b^T(z)\sum_{w\in W} \alpha_f(w)\Psi_f(w)h^{|w|_f+1}F_f(w)(Q_0).$$

These expressions are similar to (17), which was used to generate the GB-series for the exact solution. They only differ in the weight functions. We conclude that Y and y_1 can be written as GB-series similar to those of the exact solution, but with weight functions given by

$$\Psi(\star)(z) = \mathrm{e}^{cz}, \quad \Psi(\circ)(z) = A(z)\mathbb{1}_s, \quad \Psi(\tau)(z) = A(z)\alpha_f(w)\Psi_f(w)$$
$$\psi(\star)(z) = \mathrm{e}^{z}, \quad \psi(\circ)(z) = b^T(z)\mathbb{1}_s, \quad \psi(\tau)(z) = b^T(z)\alpha_f(w)\Psi_f(w)$$

where α_f and Ψ_f are given by Lemma 1. □

Remark 3. If the elementary differentials are evaluated in (t_0, y_0) the theorem is still valid, with the only difference that $\Psi(\star)(z) = \mathrm{e}^{cz} - \mathbb{1}_s$ and $\psi(\star)(z) = \mathrm{e}^{z} - 1$, see the remark after Theorem 1.

Table 3 lists the weight functions for the trees of less than four vertices. Error functions $\mathcal{E}(\tau)(z) = \varphi_e(\tau)(z) - \psi(\tau)(z)$ for the three methods of Table 1 can be found in Table 6 - 8 in the appendix.

5. Solutions on the smooth manifold

In the introductory example we saw how the error behaviour did not only depend on h and λ, but also on the initial value. The behaviour also changes over the first few steps. In the rapid decay case, this is related to how the numerical solution approaches the exact smooth solution $y_s(t)$, given by (9). In this section, we will first describe the solution $y_s(t)$ in terms of GB-series. Further, we discuss the series for the numerical solution in the case of

τ	$\lvert\tau\rvert$	$F(\tau)$	$\Gamma(\tau)(z)$	$\varphi_e(\tau)(z)$	$\Psi(\tau)(z)$	$\psi(\tau)(z)$
$\star$	0	y_0		e^z	e^{cz}	e^z
$\circ$	1	f	1	$\frac{\mathrm{e}^z-1}{z}$	$A\mathbb{1}_s$	$b^T\mathbb{1}_s$
	1	$f_y y_0$	e^z	e^z	$A\mathrm{e}^{cz}$	$b^T\mathrm{e}^{cz}$
	2	f_t	1	$\frac{\mathrm{e}^z-1-z}{z^2}$	Ac	b^Tc
	2	$f_y f$	$\frac{\mathrm{e}^z-1}{z}$	$\frac{\mathrm{e}^z(-1+z)+1}{z^2}$	$A^2\mathbb{1}_s$	$b^TA\mathbb{1}_s$
	1	$f_{yy}y_0y_0$	$\frac{1}{2}\mathrm{e}^{2z}$	$\frac{\mathrm{e}^{2z}-\mathrm{e}^z}{2z}$	$\frac{1}{2}A\mathrm{e}^{2cz}$	$\frac{1}{2}b^T\mathrm{e}^{2cz}$
	2	$f_{ty}y_0$	e^z	$\frac{\mathrm{e}^z}{2}$	$A\big(c\cdot\mathrm{e}^{cz}\big)$	$b^T\big(c\cdot\mathrm{e}^{cz}\big)$
	2	$f_{yy}fy_0$	$\mathrm{e}^z\frac{\mathrm{e}^z-1}{z}$	$\frac{\mathrm{e}^{2z}-(z+1)\mathrm{e}^z}{z^2}$	$A\big((A\mathbb{1}_s)\cdot\mathrm{e}^{cz}\big)$	$b^T\big((A\mathbb{1}_s)\cdot\mathrm{e}^{cz}\big)$
	2	$f_yf_yy_0$	e^z	$\frac{\mathrm{e}^z}{2}$	$A^2\mathrm{e}^{cz}$	$b^TA\mathrm{e}^{cz}$
	3	f_{tt}	$\frac{1}{2}$	$\frac{\mathrm{e}^z-(1+z+\frac{1}{2}z^2)}{z^3}$	$\frac{1}{2}Ac^2$	$\frac{1}{2}b^Tc^2$
	3	$f_{ty}f$	$\frac{\mathrm{e}^z-1}{z}$	$\frac{\mathrm{e}^z(-1+\frac{1}{2}z^2)+(1+z)}{z^3}$	$A\big(c\cdot(A\mathbb{1}_s)\big)$	$b^T\big(c\cdot(A\mathbb{1}_s)\big)$
	3	$f_{yy}ff$	$\frac{(\mathrm{e}^z-1)^2}{2z^2}$	$\frac{\mathrm{e}^{2z}-2z\mathrm{e}^z-1}{2z^3}$	$\frac{1}{2}A\big((A\mathbb{1}_s)\cdot(A\mathbb{1}_s)\big)$	$\frac{1}{2}b^T\big((A\mathbb{1}_s)\cdot(A\mathbb{1}_s)\big)$
	3	f_yf_t	$\frac{\mathrm{e}^z-1-z}{z^2}$	$\frac{\mathrm{e}^z(z-2)+(2+z)}{z^3}$	A^2c	b^TAc
	3	f_yf_yf	$\frac{\mathrm{e}^z(-1+z)+1}{z^2}$	$\frac{\mathrm{e}^z(1-z+\frac{1}{2}z^2)-1}{z^3}$	$A^3\mathbb{1}_s$	$b^TA^2\mathbb{1}_s$

TABLE 3. Trees with less than 4 vertices and their corresponding terms, using $Q_0 = (t_0, 0; y_0)$

consistent initial value. At the end of the section, we will use these ideas to explain how the error changes over the first few steps.

GB-series of the smooth solution. Consider again the GB-series of the exact solution given by Theorem 1. From the construction of the trees τ and the weight functions $\varphi_e(\tau)(z)$ we observe the following:

- The weight function $\varphi_e(\tau)(z)$ can be written as

$$\varphi_e(\tau)(z) = \varphi_{tr}(\tau)(z) + \varphi_s(\tau)(z)$$

  where $\varphi_{tr}(\tau)(z)$ includes all the terms involving e^z, and the remaining term $\varphi_s(\tau)(z)$ is a rational function in z.
- $\varphi_s(\tau)(z) = 0$ for all $\tau \in T$ if (at least) one of the vertices in τ is a $\star$.

If $\mathrm{Re}(\lambda) << 0$, $\varphi_{tr}(\tau)(z)$ quickly becomes negligible compared to $\varphi_s(\tau)(z)$ as z grows. Thus the exact solution can be split into a transient and a smooth part,

$$y(t_0 + h) = y_{tr}(t_0 + h) + y_s(t_0 + h),$$

where the rapidly damped transient is given by

$$y_{tr}(t_0+h) = \sum_{\tau \in T} \varphi_{tr}(\tau)(z)\, h^{|\tau|} F(\tau)(Q_0),$$

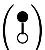
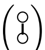
and the smooth solution, to which $y(t_0+h)$ will be attracted, is

$$y_s(t_0+h) = \sum_{\tau \in T_s} \varphi_s(\tau)(z)\, h^{|\tau|} F(\tau)(t_0, 0; \cdot). \tag{19}$$

Here $T_s \subset T$ is the subset of trees in T with no $\star$ vertex. Elementary differentials $F(\tau)(Q_0)$ of trees τ with no $\star$ vertex do not depend on the initial value y_0, this is emphasised by using $Q_0 = (t_0, 0; \cdot)$ for GB-series over T_s.

In the remaining part of this section, we will use the subscript s to denote expressions for which all exponentials e^z has been removed, that is, set to zero.

For the smooth solution, we might be more interested in the dependency of λ than that of h. It is then convenient to introduce the function $\mathcal{P}_s(\tau)(z) = z^{|\tau|}\varphi_s(\tau)(z)$. From Table 4 we get

$\mathcal{P}_s(\circ) = -1$, $\mathcal{P}_s\Big($  $\Big) = -(1+z)$, $\mathcal{P}_s\Big($  $\Big) = 1$, $\mathcal{P}_s\Big($  $\Big) = -1-z-\frac{1}{2}z^2$, $\mathcal{P}_s\Big($  $\Big) = 1+z, \cdots$,

and the smooth solution can be written as

(20)

$$\begin{aligned} y_s(t_0+h) &= \sum_{\tau \in T_s} \frac{1}{\lambda^{|\tau|}} \mathcal{P}_s(\tau)(z) F(\tau)(t_0,0;\cdot) \\ &= -\frac{1}{\lambda} f - \frac{1}{\lambda^2}\Big((1+z) f_t - f_y f\Big) \\ &\quad - \frac{1}{\lambda^3}\Big((1+z+\frac{1}{2}(z)^2) f_{tt} + (1+z) f_{ty} f - \frac{1}{2} f_{yy} f f + (2+z) f_y f_y f_y\Big) + \mathcal{O}\Big(\frac{1}{\lambda^4}\Big). \end{aligned}$$

To get an expression for the consistent initial value, keep λ fixed and let $h \to 0$:

$$\begin{aligned} y_s(t_0) &= \sum_{\tau \in T_s} \frac{1}{\lambda^{|\tau|}} \mathcal{P}_s(\tau)(0) F(\tau)(t_0,0;\cdot) \\ &= -\frac{1}{\lambda} f - \frac{1}{\lambda^2}(f_t - f_y f) + \frac{1}{\lambda^3}(f_{tt} - f_{ty} f + \frac{1}{2} f_{yy} f f - 2 f_y f_t + f_y f_y f) + \mathcal{O}\Big(\frac{1}{\lambda^4}\Big). \end{aligned} \tag{21}$$

Since $y_s(t_0)$ only depends on t_0 and f, this approach gives a complete characterisation of the smooth solution. (9).

For the numerical solution the situation is more complicated, since the behaviour of the method depends significantly on whether the initial value y_0 is on or off the smooth manifold.

Consistent initial value. Assume that the initial value is consistent, that is $y_0 = y_s(t_0)$. This improves the error behaviour significantly, and not only for h large. To se why, consider the GB-series of the numerical solution given by Theorem 2. For each elementary differential of a $\tau \in T \backslash T_s$, y_0 can be replaced by the series (21). As an example, consider the term corresponding to $\overset{\star}{\circ}$:

$$\begin{aligned}\psi\left(\overset{\star}{\circ}\right)(z)hf_y\,y_0 &= \psi\left(\overset{\star}{\circ}\right)(z)hf_y\left(-\frac{1}{\lambda}f-\frac{1}{\lambda^2}(f_t-f_yf)+\cdots\right)\\ &= \psi\left(\overset{\star}{\circ}\right)(z)\left(-\frac{1}{z}h^2f_yf-\frac{1}{z^2}h^3(f_yf_t-f_yf_yf)+\cdots\right)\\ &= -\frac{1}{z}\psi\left(\overset{\star}{\circ}\right)(z)h^2F\left(\overset{\circ}{\circ}\right)-\frac{1}{z^2}\psi\left(\overset{\star}{\circ}\right)(z)h^3F\left(\overset{\bullet}{\overset{\circ}{\circ}}\right)+\frac{1}{z^2}\psi\left(\overset{\star}{\circ}\right)(z)h^3F\left(\overset{\circ}{\overset{\circ}{\circ}}\right)+\cdots\end{aligned}$$

The term $\overset{\star}{\circ}$ contributes to the terms $\overset{\circ}{\circ}$, $\overset{\bullet}{\overset{\circ}{\circ}}$, $\overset{\circ}{\overset{\circ}{\circ}}$ etc. By replacing each y_0 with the series (21), the numerical solution y_1 can be written as as a GB-series similar to (19), or

$$y_1=\sum_{\tau\in T_s}\psi_c(\tau)(z)h^{|\tau|}F(\tau)(t_0,0;\cdot).$$

The subscript c is used to emphasise that the expression is only valid for a consistent initial value. In this case each $\psi_c(\tau)(z)$ is the sum of $\psi(\tau)(z)$ and terms coming from some $T\backslash T_s$. To find $\psi_c(\tau)(z)$, the following lemma becomes useful:

Lemma 2. *If $y_0=y_s(t_0)$, then for each $\tau^\star\in T\backslash T_s$ the elementary differential can be expanded into the series*

$$F(\tau^\star)(t_0,0;y_s(t_0))=\sum_{u\in U(\tau^\star)}\beta(u;\tau^\star)\frac{1}{\lambda^{\sigma(u;\tau^\star)}}\mathcal{Q}(u;\tau^\star)F(u)(t_0,0;\cdot)$$

The set $U(\tau^\star)\subset T_s$ is defined recursively by

a1) $U(\star)=T_s$.
a2) *If* $\tau^\star=[\bullet^k,\tau_1^\star,\cdots,\tau_l^\star,\tau_{l+1},\cdots,\tau_m]$ *with* $\tau_i^\star\in T\backslash T_s$ *and* $\tau_i\in T_s$, *then* $u=[\bullet^k,u_1,\cdots,u_l,\tau_{l+1},\cdots,\tau_m]\in U(\tau^\star)$ *if* $u_i\in U(\tau_i^\star)$.

The remaining terms are given by

b) $$\sigma(u;\tau^\star)=\begin{cases}|u| & \text{if } u\in U(\star),\\ \sum_{i=1}^{l}\sigma(u_i;\tau_i^\star) & \text{if } \tau^\star \text{ and } u \text{ is given by a2).}\end{cases}$$

c) $$\mathcal{Q}(u;\tau^\star)=\begin{cases}\mathcal{P}_s(u)(0) & \text{if } u\in U(\star),\\ \prod_{i=1}^{l}\mathcal{Q}(u_i;\tau_i^\star) & \text{if } \tau^\star \text{ and } u \text{ is given by a2).}\end{cases}$$

d) $$\beta(u;\tau^\star)=\begin{cases}1 & \text{if } u\in U(\star),\\ \frac{l!}{r_1!\cdots r_q!}\prod_{i=1}^{l}\beta(u_i;\tau_i^\star) & \text{if } \tau^\star \text{ and } u \text{ is given by a2).}\end{cases}$$

In d) we have assumed that among $u_1,\cdots,u_l$ there are q distinct trees, each of multiplicity r_j, $j=1,\cdots,q$.

Proof. Obviously $F(\star)(t_0,0;y_s(t_0))=y_s(t_0)$ and for this tree the result follows from (21). Let $\tau^\star=[\bullet^k,\tau_1^\star,\cdots,\tau_l^\star,\tau_{l+1},\cdots,\tau_m]$, and assume the statement of the theorem to be true for $\tau_i^\star$,

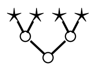
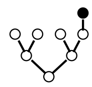
$i = 1, \cdots, l$. From Theorem 1 we get

$$F(\tau^\star)(t_0, 0; y_s(t_0)) = f_{ktmy}(t_0, 0)\Bigg(\prod_{i=1}^{l}\Big(\sum_{u_i \in U(w_i^\star)} \beta(u_i; \tau_i^\star)\frac{1}{\lambda^{\sigma(u_i;\tau_i^\star)}}\mathcal{Q}(u_i; \tau_i^\star)F(u_i)(t_0, 0; \cdot)\Big)$$
$$\cdot \prod_{i=l+1}^{m} F(\tau_i)(t_0, 0; \cdot)\Bigg)$$
$$= \sum_{u \in U(\tau^\star)} \beta(u; \tau^\star)\Bigg(\prod_{i=1}^{m}\frac{1}{\lambda^{\sigma(u_i;\tau_i^\star)}}\mathcal{Q}(u_i; \tau_i^\star)\Bigg)F(u)(t_0, 0; \cdot),$$

using the same argument for $\beta(u; \tau^\star)$ as used in the proof of *b)* of Lemma 1. □

Remark 4. A simple intepretation of the lemma can be given as follows: Assume that $\tau^\star$ has k $\star$-nodes. A tree $u \in U(\tau^\star)$ is found by replacing each of the $\star$-nodes by some $u_i \in T_s$, $i = 1, \cdots, k$. Then $\sigma(u; \tau^\star) = \sum_{i=1}^{k} |u_i|$ and $\mathcal{Q}(u; \tau^\star) = \prod_{i=1}^{k} \mathcal{P}_s(u_i)(0)$. Unfortunately, $\beta(u; \tau^\star)$ still has to be calculated recursively.

Example 3. Let

$\tau^\star =$  and $u =$ .

Since $\mathcal{P}_s(\circ) = -1$ and $\mathcal{P}_s$ () $= -1$ we have

$$\sigma(u; \tau^\star) = 5, \qquad \mathcal{Q}(u; \tau^\star) = 1, \qquad \beta(u; \tau^\star) = 4$$

We are now ready to present the main result of this section:

Theorem 3. *Assume that the initial value y_0 is consistent, thus satisfying* (21)*. For a $\tau \in T_s$, let*

$$S(\tau) = \{\tau^\star \in T \backslash T_s,\ \textit{such that}\ \tau \in U(\tau^\star)\}$$

where $U(\tau^\star)$ is given by Lemma 2. Then the exact and numerical solution can be written as

$$y(t_0 + h) = \sum_{\tau \in T_s} \varphi_s(\tau)(z)h^{|\tau|}F(\tau)(Q_0)$$
$$y_1 = \sum_{\tau \in T_s} \psi_c(\tau)(z)h^{|\tau|}F(\tau)(Q_0)$$

with

$$\varphi_s(\tau)(z) = \varphi_e(\tau)(z)|_{e^z=0}$$

and

$$\psi_c(\tau)(z) = \psi(\tau)(z) + \sum_{\tau^\star \in S(\tau)} \frac{\psi(\tau^\star)(z)}{z^{|\tau|-|\tau^\star|}}\beta(\tau; \tau^\star)\mathcal{Q}(\tau; \tau^\star).$$

where $\beta(\tau; \tau^\star)$ and $\mathcal{Q}(\tau; \tau^\star)$ are given by Lemma 2.

Proof. The results for the exact solution is given by (9). For the results of the numerical solution, use (12) and Lemma 2 so that the numerical solution can be written as

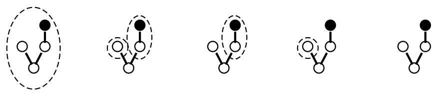
$$\begin{aligned} y_1 &= \sum_{\tau\in T_s} \psi(\tau)(z)h^{|\tau|}F(\tau)(t_0,0;\cdot) + \sum_{\tau^\star\in T\setminus T_s} \psi(\tau^\star)(z)h^{|\tau^\star|}F(\tau^\star)(t_0,0;y_s(t_0)) \\ &= \sum_{\tau\in T_s} \psi(\tau)(z)h^{|\tau|}F(\tau)(t_0,0;\cdot) + \sum_{\tau^\star\in T\setminus T_s} \psi(\tau^\star)(z)h^{|\tau^\star|} \sum_{\tau\in U(\tau^\star)} \frac{\beta(\tau;\tau^\star)}{\lambda^{\sigma(\tau;\tau^\star)}}\mathcal{Q}(\tau;\tau^\star)F(\tau)(t_0,0;\cdot) \\ &= \sum_{\tau\in T_s}\left(\psi(\tau)(z)h^{|\tau|} + \sum_{\tau^\star\in S(\tau)} \frac{\psi(\tau^\star)(z)}{(h\lambda)^{\sigma(\tau;\tau^\star)}}h^{|\tau^\star|+\sigma(\tau;\tau^\star)}\beta(\tau;\tau^\star)\mathcal{Q}(\tau;\tau^\star)\right)F(\tau)(t_0,0;\cdot). \end{aligned}$$

Using the fact that $|\tau^\star| + \sigma(\tau;\tau^\star) = |\tau|$ and $z = \lambda h$ we conclude the proof. □

The set $S(\tau)$ can be found by the following rule:

Partition rule: Given a $\tau \in T_s$, and let w be a subtree of τ, in the sense that w is a connected graph including the root of τ. Let $\tau\setminus w$ be the "forest" collecting the trees left over when w has been removed from τ. Make sure that w is chosen such that $\tau\setminus w \subset T_s$. For each w there is an associated $\tau^\star$, that is the tree consisting of w and all the removed trees replaced by a $\star$. $S(\tau)$ is then the set of all such possible $\tau^\star$'s.

Example 4. The partition of the 4th order tree $\tau = [\circ,[\bullet]]$ is given by



thus

$$\psi_c(\tau)(z) = \frac{1}{z^4}\psi(\star)(z)\mathcal{P}_s\left([\circ,[\bullet]]\right)(0) + \frac{1}{z^3}\psi\left([\star,\star]\right)(z)\mathcal{P}_s(\circ)(0)\mathcal{P}_s\left([\bullet]\right)(0)$$
$$+\frac{1}{z^2}\psi\left([\circ,\star]\right)(z)\mathcal{P}_s\left([\bullet]\right)(0) + \frac{2}{z}\psi\left([\star,[\bullet]]\right)(z)\mathcal{P}_s(\circ)(0) + \psi\left([\circ,[\bullet]]\right)(z).$$

The expressions for $\varphi_s(\tau)$ and $\psi_c(\tau)$ are given in Table 4. For the methods of section 2 the lowest order error terms $\mathcal{E}_c(\tau)(z) = \varphi_s(\tau)(z) - \psi_c(\tau)(z)$ becomes

$$\begin{aligned} &\mathcal{E}_c\left([[\bullet]]\right)(z) = \tfrac{4-5z+z^2}{z^5}e^{2z} - \tfrac{8-4z}{z^5}e^{\frac{3z}{2}} + \tfrac{2+z^2}{z^4}e^{z} + \tfrac{8+4z}{z^5}e^{\frac{z}{2}} - \tfrac{4+5z+z^2}{z^5} && \text{for CM3} \\ (22)\quad &\mathcal{E}_c\left([\bullet,\bullet]\right)(z) = \tfrac{z^2-4z+6}{6z^3}e^{z} - \tfrac{z+3}{3z^3} && \text{for Radau IIA} \\ &\mathcal{E}_c\left([\bullet]\right)(z) = \tfrac{1}{z^2}e^{z} - \tfrac{3}{4z}e^{\frac{2z}{3}} - \tfrac{4+z}{4z^2} && \text{for CMO3} \end{aligned}$$

which is a significant improvement for all three methods. Notice that Theorem 3 is valid also for small values of $|z|$.

Nonconsistent initial value. We now assume y_0 to be different from, but close to $y_s(t_0)$, say $y_s(t_0) - y_0 = \mathcal{O}(h)$. As for the exact solution, if $\mathrm{Re}(\lambda) << 0$ the numerical solution y_1 can be split into a transient and a smooth part, of which the smooth part can be written as

$$y_{s,1} = \sum_{\tau\in T} \psi_s(\tau)(z)h^{|\tau|}F(\tau)(t_0,0;y_0) \qquad (23)$$

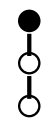
v	$\varphi_s(\tau)(z)$	$\psi_c(\tau)(z)$
	$-\frac{1}{z}$	$-\frac{1}{z}e^z + b^T\mathbb{1}_s$
	$-\frac{1+z}{z^2}$	$-\frac{1}{z^2}e^z + b^Tc$
	$\frac{1}{z^2}$	$\frac{1}{z^2}e^z - \frac{1}{z}b^Te^{cz} + b^TA\mathbb{1}_s$
	$-\frac{1+z+\frac{1}{2}z^2}{z^3}$	$-\frac{1}{z^3}e^z + \frac{1}{2}b^Tc^2$
	$\frac{1+z}{z^3}$	$\frac{1}{z^3}e^z - \frac{1}{z}b^T(c\cdot e^{cz}) + b^T\big(c\cdot(A\mathbb{1}_s)\big)$
	$-\frac{1}{2z^3}$	$-\frac{1}{2z^3}e^z + \frac{1}{2z^2}b^Te^{2cz} - \frac{1}{z}b^T\big(e^{cz}\cdot(A\mathbb{1}_s)\big)+\frac{1}{2}b^T\big((A\mathbb{1}_s)\cdot(A\mathbb{1}_s)\big)$
	$\frac{2+z}{z^3}$	$\frac{2}{z^3}e^z - \frac{1}{z^2}b^Te^{cz} + b^TAc$
	$-\frac{1}{z^3}$	$-\frac{1}{z^3}e^z + \frac{1}{z^2}b^Te^{cz} - \frac{1}{z}b^TAe^{cz} + b^TA^2\mathbb{1}_s$

TABLE 4. Weight functions when consistent initial values have been used.

where $\psi_s(\tau)(z)$ is the weight function $\psi(\tau)$ with all exponential functions set to 0. Usually, $\psi_s(\tau)(z) \neq 0$ for at least some $\tau \in T\backslash T_s$, in which case the numerical smooth solution depends on the initial value y_0. Fortunately the number of terms that has to be considered can be reduced in two common cases:

a) If $c_i > 0$, $i = 1, \cdots, s$ then $\Psi_s(\star) = e^{cz}|_{\mathrm{Re}(z)\to-\infty} = [0, 0, \cdots, 0]^T$, thus $\psi_s(\tau)(z) = 0$ for all $\tau \in T\backslash T_s$. In this case

$$y_{s,1} = \sum_{\tau\in T_s} \psi_s(\tau)(z)h^{|\tau|}F(\tau)(t_0, 0; \cdot),$$

the numerical solution do not depend on the initial values, and the comparision with (19) is straightforward.

b) If the method is explicit and $c_i > 0$, $i = 2, \cdots, s$, then $c_1 = 0$ and $\Psi_s(\star) = e^{cz}|_{\mathrm{Re}(z)\to-\infty} = [1, 0, \cdots, 0]^T$. Since the method is explicit, $A(z)$ is lower triangular, and $\Psi_s(\tau)(z) = [0, \times, \cdots, \times]^T$ for all $\tau \in T\backslash\{\star\}$. Consider trees of the form

$$\tilde{\tau} = [\bullet^k, \tau_1, \cdots, \tau_l, \star^m] \text{ with } \tau_i \in T\backslash\{\star\}, \quad i = 1, \cdots, l, \text{ and } k + l \geq 1, \quad m \geq 1.$$

For such trees we get

$$\Psi_s(\tilde{\tau}) = A_s(z)\left(c^k \cdot \prod_{i=1}^{l} \Psi_s(\tau_i)\cdot(\Psi_s(\star))^m\right) = [0, 0, \cdots, 0]^T$$

and similar $\psi_s(\tilde{\tau})(z) = 0$. Thus for all $\tau \in T$ that is or has a subtree of the form $\tilde{\tau}$, $\psi_s(\tau)(z) = 0$.

The weight functions of the first order "bushy" trees , , $\cdots$ are in this case given by

$$\psi_s([\star^k])(z) = \frac{1}{k!}b^T(z)\big(\Psi_s(\star)\big)^k = \frac{1}{k!}b_{s,1}(z).$$

These terms together give the following contribution to the error

$$\sum_{k=1}^{\infty} \psi_s([\star^k]) hF([\star^k])(Q_0) = hb_{s,1}(z) \sum_{k=1}^{\infty} \frac{1}{k!} f_{ky}(t_0, 0) y_0^k$$
$$= hb_{s,1}(z)\big(f(t_0, y_0) - f(t_0, 0)\big).$$

Radau IIA satisfies the condition of case *a)*, CM3 and CMO3 those of case *b)*.

Error behaviour over the first few steps. The influence of initial values are reduced as the integration goes on. In this section we will explain why.

Assume that n steps are taken. We allow for variable step size, thus $h_k = t_{k+1} - t_k$ and $z_k = \lambda h_k$, $k = 0, \cdots, n$, However, each stepsize is assumed to be large enough to damp out the transients. The numerical solution after one step is given by (23), or

$$y_{s,n+1} = \sum_{\tau \in T} \frac{1}{\lambda^{|\tau|}} \mathcal{P}_N(\tau)(z_n) F(\tau)(t_n, 0; y_{s,n}),$$

where $\mathcal{P}_N(\tau)(z_n) = z_n^{|\tau|} \psi_s(\tau)(z_n)$. If $n \geq 1$, each $y_{s,n}$ in $F(\tau)(t_n, 0; y_{s,n})$ can be replaced by its series:

$$y_{s,n} = \sum_{\tau \in T} \frac{1}{\lambda^{|\tau|}} \mathcal{P}_N(\tau)(z_{n-1}) F(\tau)(t_{n-1}, 0; y_{s,n-1})$$
$$= \sum_{\tau \in T} \frac{1}{\lambda^{|\tau|}} \mathcal{P}_N(\tau)(z_{n-1}) \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{-z_{n-1}}{\lambda} \right)^k \frac{\partial^k F}{\partial t^k}(t_n, 0; y_{s,n-1}),$$

where the elementary differential $F(\tau)(t_{n-1}, 0, y_{n-1})$ have been replaced by its taylor expansion around t_n, using $h_{n-1} = z_{n-1}/\lambda$. For simplicity, we will in this subsection use the notation $y_{s,0} = y_0$ even if y_0 is not necessarily consistent.

Example 5. Consider the term in the GB-series of $y_{s,n+1}$ given by

$$\frac{1}{\lambda} \mathcal{P}_N([\star])(z_n) f_y y_n = \frac{1}{\lambda} \mathcal{P}_N([\star])(z_n) \Bigg(\frac{1}{\lambda} P_N(\circ)(z_{n-1}) f_y \left(f - \frac{z_{n-1}}{\lambda} f_t + \cdots \right)$$
$$+ \frac{1}{\lambda} \mathcal{P}_N([\star])(z_{n-1}) f_y f_y y_{n-1} + \frac{1}{\lambda^2} \mathcal{P}_N([\bullet])(z_{n-1}) f_y f_t$$
$$+ \frac{1}{\lambda^2} \mathcal{P}_N([\bullet])(z_{n-1}) f_y f_t + \frac{1}{\lambda^2} \mathcal{P}_N([\circ])(z_{n-1}) f_y f_y f + \cdots \Bigg).$$

The term corresponding to $[\star]$ thus contributes to terms of the trees

$$[\circ], \quad [[\bullet]], \quad [[\star]], \quad [[\circ]], \quad \cdots$$

If $n \geq 2$ then $y_{s,n-1}$ can be expressed by its series, and so on. By this procedure it is possible to write the GB-series for $y_{s,n+1}$ as

$$y_{s,n+1} = \sum_{\tau \in T} \frac{1}{\lambda^{|\tau|}} \mathcal{P}_{N,i}(z_n, \cdots, z_{n-i}) F(\tau)(t_n, 0; y_{s,n-i})$$
$$= \sum_{\tau \in T} \psi_{s,i}(\tau)(z_n, \cdots, z_{n-i}) h_n^{|\tau|} F(\tau)(t_n, 0; y_{s,n-i}), \qquad i \leq n,$$

$\star$	0
$\circ$	$\psi_s(\circ)(z_n)$
$[\star]$	0
$[\bullet]$	$\psi_s([\bullet])(z_n)$
$[\circ]$	$\psi_s([\circ])(z_n) + \frac{z_{n-1}}{z_n}\psi_s([\star])(z_n)\psi_s(\circ)(z_{n-1})$
$[\star,\star]$	0
$[\star,\bullet]$	0
$[\star,\circ]$	0
$[[\star]]$	$\frac{z_{n-1}}{z_n}\psi_s([\star])(z_n)\psi_s([\star])(z_{n-1})$ if $i=1$, 0 otherwise.
$[\bullet,\bullet]$	$\psi_s([\bullet,\bullet])(z_n)$
$[\bullet,\circ]$	$\psi_s([\bullet,\circ])(z_n) + \frac{z_{n-1}}{z_n}\psi_s([\bullet,\star])\psi_s(\circ)(z_{n-1})$
$[\circ,\circ]$	$\psi_s([\circ,\circ])(z_n) + \frac{z_{n-1}}{z_n}\psi_s([\star,\circ])(z_n)\psi_s(\circ)(z_{n-1}) + \frac{z_{n-1}^2}{z_n^2}\psi_s([\star,\star])(z_n)\psi_s(\circ)(z_{n-1})^2$
$[[\bullet]]$	$\psi_s([[\bullet]]) + \frac{z_{n-1}^2}{z_n^2}\psi_s([\star])(z_n)\psi_s([\bullet])(z_{n-1}) - \frac{z_{n-1}^2}{z_n^2}\psi_s([\star])(z_n)\psi_s(\circ)(z_{n-1})$
$[[\circ]]$	$\psi_s([[\circ]])(z_n) + \frac{z_{n-1}}{z_n}\psi_s([[\star]])(z_n)\psi_s(\circ)(z_{n-1}) + \frac{z_{n-1}^2}{z_n^2}\psi_s([\star])(z_n)\psi_s([\circ])(z_{n-1})$ if $i=1$
	$+\frac{z_{n-1}z_{n-2}}{z_n^2}\psi_s([\star])(z_n)\psi_s([\star])(z_{n-1})\psi_s(\circ)(z_{n-2})$ if $i\geq 2$

TABLE 5. Weight functions $\psi_{s,i}(\tau)$ for $i \geq 1$

where $z_n^{|\tau|}\psi_{s,i}(\tau) = \mathcal{P}_{N,i}(\tau)$. The expressions for some $\psi_{s,i}$ are given in Table 5, see also Table 9. We observe that the elementary weight functions do not change after a certain number of steps. This number of steps is related to the *height* of the tree, defined by

$$H(\tau) = \begin{cases} 1 & \text{if } \tau \in \{\circ, \star\}, \\ 1 + \max_{j=1,\cdots,m} H(\tau_i) & \text{if } \tau = [\bullet^k, \tau_1, \cdots, \tau_m] \in T. \end{cases}$$

By the construction of $\psi_{s,i}$ we can conclude that

$$\psi_{s,i}(\tau)(z_n, \cdots, z_{n-i}) = \psi_{s,i+1}(\tau)(z_n, \cdots, z_{n-i-1}) \qquad \text{if } i \geq H(\tau) - 1$$

and in particular

$$\psi_{s,i}(\tau) = 0, \quad \text{if} \quad \tau \in T \backslash T_s \quad \text{and} \quad i \geq H(\tau) - 1.$$

Thus, the influence of initial the value is reduced for each step.

6. Numerical experiments and discussion

In this section we discuss the results of the previous sections, and how they can be used to explain observable phenomenas of the error behaviour. The discussion is illustrated by the following simple example:

Example 6. Consider the equation

$$y' = \lambda y + (y + \sin(t)), \qquad y(t_0) = y_0, \qquad t_0 = \pi/4. \tag{24}$$

thus $f(t, y) = y + \sin(t)$. The exact solution to this problem is

$$y(t) = Ce^{(\lambda+1)t} + \frac{\cos(t) - (\lambda + 1)\sin(t)}{(\lambda + 1)^2 + 1}.$$

Two different initial values are considered,

$$y_0 = 1, \qquad \text{(non-consistent)},$$

$$y_0 = \frac{\cos(t_0) - (\lambda + 1)\sin(t_0)}{(\lambda + 1)^2 + 1}, \qquad \text{(consistent)}.$$

We will use $\lambda = -10^2, -10^4$ to represent the rapid decay case, and $\lambda = 10^2 i, 10^4 i$ for the highly oscillatory case.

The problem is solved by the methods CM3, Radau IIA and CMO3. The error after one step using different stepsizes is measured and presented in Figure 3 and 4. In the rapid decay case, also the error after two and three steps are presented, using the same stepsize for all the steps.

For the error functions of the three methods, we refer to the appendix.

Nonstiff case. We assume z to be sufficiently small to study the weight functions by their series expansions in z. For a given $\tau \in T$ let $q(\tau)$ be

$$\varphi_e(\tau)(z) - \psi(\tau)(z) = \mathcal{O}\big(z^{q(\tau)+1}\big). \tag{25}$$

The error after one step will be

$$y(t_0 + h) - y_1 = \sum_{\tau\in T} \big(\varphi_e(\tau)(z) - \psi(\tau))(z)\big) h^{|\tau|} F(\tau)(t_0, 0; y_0) = \mathcal{O}(\lambda^{r+1} h^{p+1})$$

where

$$p = \min_{\tau\in T}(q(\tau) + |\tau|) \quad \text{and} \quad r = \max_{\substack{\tau\in T \\ q(\tau)+|\tau|=p}} q(\tau).$$

To find p by computing $q(\tau)$ for all $\tau \in T$ is a formidable task because there is an infinite number of even low order trees, that is the trees with $\star$-vertices. In this situation it is advantageous to take the GB-series around (t_0, y_0).

$$y(t_0 + h) = y_0 + \sum_{\tau\in T} \bar{\varphi}_e(\tau)(z) h^{|\tau|} F(\tau)(t_0, y_0; y_0)$$

$$Y = y_0 \mathbb{1}_s + \sum_{\tau\in T} \bar{\Psi}(\tau)(z) h^{|\tau|} F(\tau)(t_0, y_0; y_0)$$

$$y_1 = y_0 + \sum_{\tau\in T} \bar{\psi}(\tau)(z) h^{|\tau|} F(\tau)(t_0, y_0; y_0)$$

where $\bar{\varphi}_e$, $\bar{\Psi}$ and $\bar{\psi}$ are given by the reccurence formulas of Theorem 1 and Theorem 2 with the modifications described in the respective remarks. Because

$$\bar{\varphi}_e(\star) = e^z - 1 = \mathcal{O}(z), \quad \bar{\Psi}(\star)(z) = e^{cz} - \mathbb{1}_s = \mathcal{O}(z) \text{ and } \bar{\psi}(\star)(z) = \mathrm{e}^z - 1 = \mathcal{O}(z),$$

the contributions to the weight functions corresponding to each $\star$ is $\mathcal{O}(h)$, thus also trees in $T \backslash T_s$ results in terms of order equal to the number of nodes. We get the following lemma:

Lemma 3. *The local truncation error satisfies*

$$y(t_0 + h) - y_1 = \mathcal{O}(\lambda^{r+1} h^{p+1})$$

if

$$\bar{\varphi}_e(\tau)(z) - \bar{\psi}(\tau)(z) = \mathcal{O}\big(z^{q(\tau)+1}\big) \text{ with } q(\tau) \geq p - |\tau|, \quad \forall \tau \in T \text{ with less than } p+1 \text{ vertices,}$$

and $r = \max q(\tau)$, the maximum taken over all τ such that $q(\tau) + |\tau| = p$.

If the initial value is consistent, the condition can be replaced by

$$\varphi_s(\tau)(z) - \psi_s(\tau)(z) = \mathcal{O}\big(z^{q(\tau)+1}\big) \text{ with } q(\tau) \geq p - |\tau|, \qquad \forall \tau \in T_s, \quad |\tau| \leq p.$$

The statement for a consistent initial value is given by (19) and Lemma 2. Applying these results to the example methods give

$$y(t_0 + h) - y_1 = \begin{cases} \mathcal{O}(h^4 + \lambda^4 h^5) & \text{for CM3,} \\ \mathcal{O}(\lambda^3 h^4) & \text{for Radau IIA,} \\ \mathcal{O}(\lambda^3 h^4) & \text{for CMO3.} \end{cases}$$

For CM3 the order 4 term do not depend on λ, the order 5 term might dominate the error and is therefore included. Using consistent inital value we get:

$$y(t_0 + h) - y_1 = \begin{cases} \mathcal{O}(h^4 + \lambda^2 h^5) & \text{for CM3,} \\ \mathcal{O}(\lambda h^4) & \text{for Radau IIA,} \\ \mathcal{O}(\lambda^2 h^4) & \text{for CMO3.} \end{cases}$$

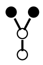
These results are in perfect agreements with the results of Figure 3 and 4.

Rapid decay case. In this case we assume $\mathrm{Re}(\lambda) << 0$ and $|z| >> 1$. Let us first consider consistent initial values, in which case the error behaves as

$$\sum_{\tau \in T_s} (\varphi_s(\tau)(z) - \psi_c(\tau)(z)) h^{|\tau|} F(\tau)(t_0, 0; \cdot).$$

The lowest order error terms are given by (22) ignoring all expontential terms. For the CM3 method there are also significant contributions from the fourth order terms:


$$: \varphi_s(\tau)(z) - \psi_c(\tau)(z)|_{e^z=0} = -\frac{z^2 + 6z + 12}{z^4}$$


$$: \varphi_s(\tau)(z) - \psi_c(\tau)(z))|_{e^z=0} = \frac{z^4 + 5z^3 + 8z^2 - 4z - 16}{4z^6}$$

Using consistent initial value, the error after one step of the methods behaves as

$$y(t_0 + h) - y_1 = \begin{cases} \mathcal{O}\left(\frac{1}{z^3} h^3 + \frac{1}{z^2} h^4\right) & \text{or} \quad \mathcal{O}\left(\frac{1}{\lambda^3} + \frac{1}{\lambda^2} h^2\right) & \text{for CM3} \\ \mathcal{O}\left(\frac{1}{z^2} h^3\right) & \text{or} \quad \mathcal{O}\left(\frac{1}{\lambda^2} h\right) & \text{for RadauIIA} \\ \mathcal{O}\left(\frac{1}{z} h^2\right) & \text{or} \quad \mathcal{O}\left(\frac{1}{\lambda} h\right) & \text{for CMO3.} \end{cases}$$

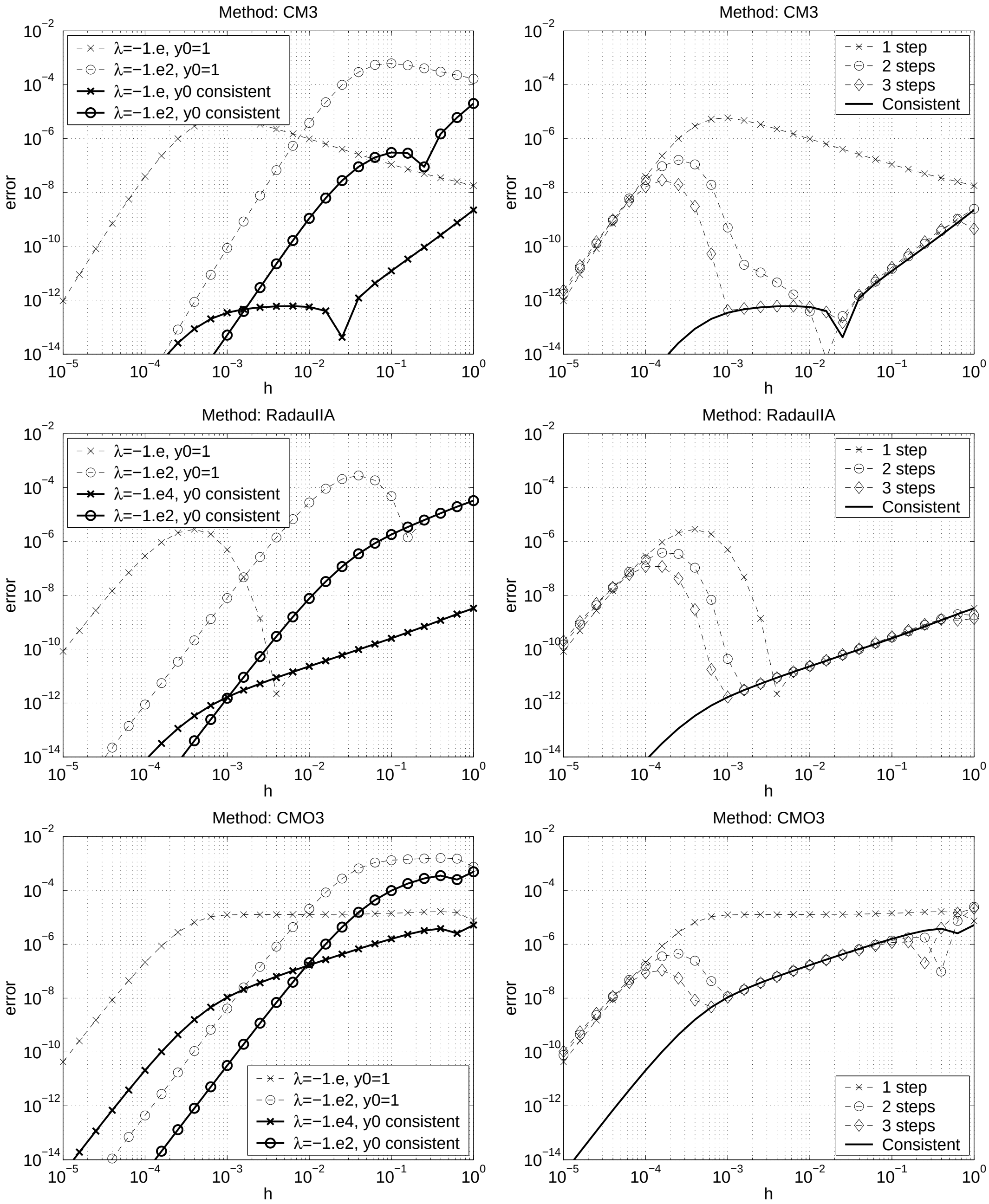


FIGURE 3. Local error: To the left, local error for different values of λ, with and without consistent initial values. To the right, the local error measured for the first three steps.

when $y_0 = y_c(t_0)$, which is exactly what we observe.

Using an inconsistent initial value, the error of both explicit methods will be dominated by terms represented by the bushy trees, $\cdots$. Since both methods satisfy the conditions of case *b)*, page 17, these terms together gives a contribution to the error of

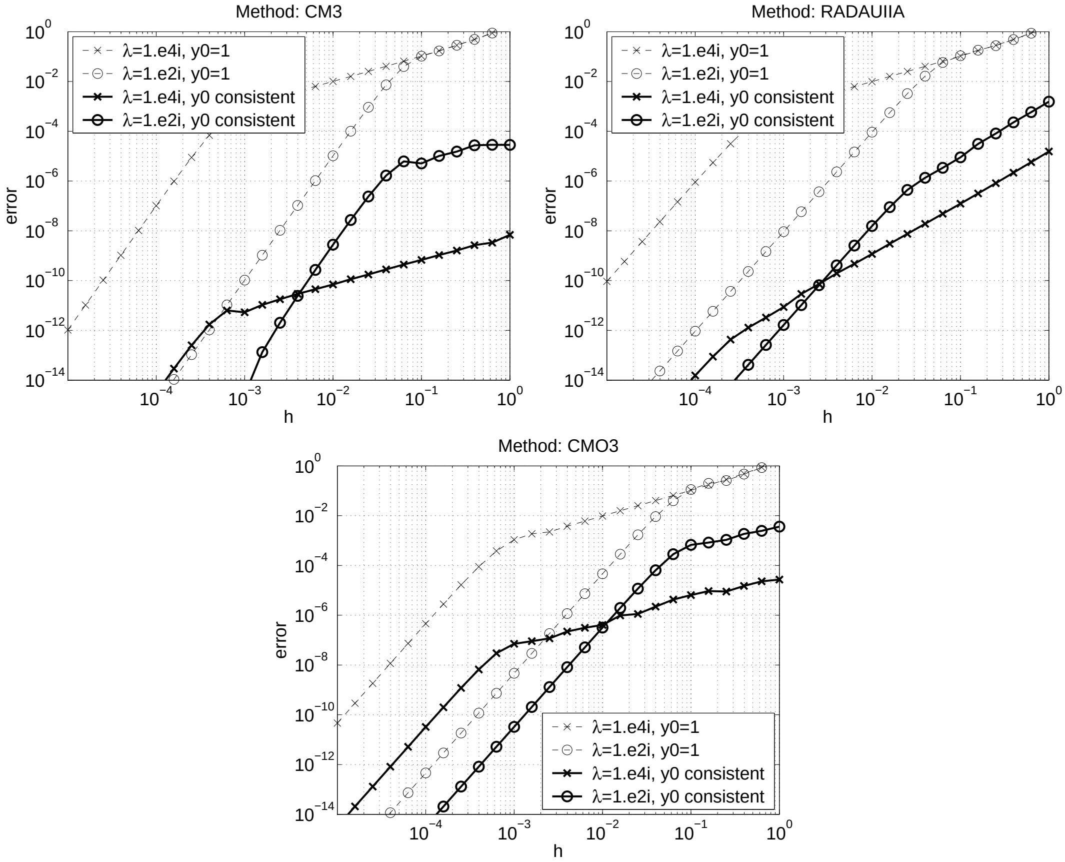


FIGURE 4. Local error, highly oscillatory case.

$b_{s,1}(z)\,h\,(f(t_0,y_0) - f(t_0,0))$. For some arbitrary y_0 the error after one step will behave as

$$y(t_0+h) - y_1 = \begin{cases} \mathcal{O}\left(\frac{1}{z^2}h\right) & \text{or} \quad \mathcal{O}\left(\frac{1}{\lambda h}\right) \quad \text{for CM3} \\ \mathcal{O}\left(\frac{1}{z}h\right) & \text{or} \quad \mathcal{O}\left(\frac{1}{\lambda}\right) \quad \text{for CMO3.} \end{cases}$$

After one more step these terms disappear. The dominating error terms of the CMO3 method is then represented by $\substack{\bullet\\|\\\circ}$, the same as for the consistent inital value case. The term corresponding to $\substack{\star\\|\\\circ\\|\\\circ}$ together with the fourth order terms discussed above dominates the error for the CM3 method. Thus

$$y(t_0+2h) - y_2 = \mathcal{O}\left(\tfrac{1}{z^4}h^2 + \tfrac{1}{z^2}h^4\right) \text{ or } \mathcal{O}\left(\tfrac{1}{\lambda^4 h^2} + \tfrac{1}{\lambda^2}h^2\right) \quad \text{for CM3.}$$

The $1/(\lambda^4 h^2)$ term disappear in the next step and the error for the CM3 method behaves as in the consistent initial value case.

The smooth solution of the Radua IIA method do not depend on the initial values.

Rapid oscillations. At last, we will consider the case of rapid oscillations, that is $\mathrm{Im}(z)$ large, thus all the exponential represents oscillations. In this situation, assume that either y_0 is small, or f linear in y. For an arbitrary initial value y_0 the error for all the methods will be dominated by the order one term represented by $\overset{\star}{\underset{\circ}{|}}$, that is they are all dominated by a term of the size $e^z h$. For $y_0 = y_c(t_0)$ the picture is more diverse. From the tables in Appendix 1 we observe that

$$y(t_0+h) - y_1 \sim \begin{cases} \frac{1}{\lambda^2} h e^z & \text{for CM3} \\ \frac{1}{\lambda} h^2 e^z & \text{for Radau IIA} \\ \frac{1}{\lambda}\left(\frac{3}{4} e^{\frac{2z}{3}} + \frac{1}{4}\right) h & \text{for CMO3} \end{cases}$$

which again is an agreement with the observed results of Figure 4.

7. Conclusion

The main idea of this paper is to express the error of an exponential RK method applied to a single Fourier mode, by use of a modification of B-series. In the modified series the stiffness is isolated from the elementary differentials and included into the coefficients of the series. This makes it possible to study the stiffness dependence on each term separately, and thereby gain better insight into how methods might behave when applied to stiff equations. Obviously, the study of a scalar equation (3) can not replace any of the order and convergence results mentioned in the introduction. However, as we have seen, it gives a quite precise and illustrative description of certain properties of the error which is not covered by previous results, and is therefore a relevant supplement to the present literature. One example is the dependence on the initial value, which can be precisely described using the presented theory, but is not evident from classical error analysis.

The basic idea of the paper can easily be adopted to other one-step methods, e.g. the implicit-explicit Runge-Kutta methods, see [11]. The extension to the system (1) is straightforward in the linear case. In the semilinear case the theory applies if L and f commute. But at the moment the theory is not applicable for general semilinear systems.

References

[1] H. Berland, B. Owren, and B. Skaflestad. B-series and order conditions for exponential integrators. Preprint, Numerics No. 5/04, 2004.

[2] J. C. Butcher. *Numerical methods for ordinary differential equations*. Wiley, 2003.

[3] E. Celledoni, A. Marthinsen, and B. Owren. Commutator-free Lie group methods. *Future Generation Computer Systems*, 19:341–352, 2003.

[4] J. Certaine. The solution of ordinary differential equations with large time constants. *Mathematical methods for digital computers*, 1960.

[5] S.M. Cox and P.C. Matthews. Exponential time differencing for stiff systems. *J. of Comp. Phys.*, 176:430–455, 2002.

[6] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I, Nonstiff Problems*. Springer, Berlin, Second revised edition, 2000.

[7] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II, Stiff and Differential-Algebraic Problems*. Springer, Berlin, Second revised edition, 2002.

[8] M. Hochbruck and A. Ostermann. Explicit exponential Runge-Kutta for semilinear parabolic equations. to appear in SIAM J. Numer. Anal., 2004.

[9] M. Hochbruck and A. Ostermann. Exponential Runge-Kutta methods for parabolic equations. Preprint, 2004.

[10] S. Koikari. Rooted tree analysis of Runge-Kutta methods with exact treatment of linear terms. *J. Comp. Appl. Math.*, 177(2):427–453, 2005.

[11] A. Kværnø. Time integration methods for coupled equations. Preprint, Numerics No.7/04, 2004.

[12] J.D. Lawson. Generalized Runge-Kutta processes for stable systems with large Lipschitz constants. *SIAM J. Numer. Anal.*, 4:372–380, 1967.

[13] B.V. Minchev and W.M. Wright. A review of exponential integrators for semilinear problems. Preprint, Numerics No.2/05, 2005.
[14] S.P. Norsett. An a-stable modification of the Adams-Bashforth methods. In *Lecture Notes in Mathematics*, number 109, pages 214–219. Springer, Berlin, 1969.
[15] D. A. Pope. An exponential method of numerical integration of ordinary differential equations. *Comm. ACM*, 6(8):491–493, 1963.

Appendix A. Error functions

In this appendix error functions of the methods CM3, RadauIIA and CMO3 are given.

Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim
E-mail address: anne@math.ntnu.no

TABLE 6. The error functions $\varphi_e(\tau)(z) - \psi(\tau)(z)$ and the error functions for consistent initial value $\phi_s(\tau)(z) - \psi_c(\tau)(z)$ for the CM3 method.

τ	$\phi_e(\tau)(z) - \psi(\tau)(z)$	$\phi_s(\tau)(z) - \psi_c(\tau)(z)$
$\star$	0	–
$\circ$	0	0
	$\frac{-4+z}{z^3}e^{2z} + \frac{8-4z}{z^3}e^{\frac{3z}{2}} + \frac{6+z^2}{z^2}e^{z} - \frac{8+4z}{z^3}e^{\frac{z}{2}} + \frac{4+z}{z^3}$	–
	0	0
	$\frac{-4+z}{z^4}e^{2z} + \frac{8-4z}{z^4}e^{\frac{3z}{2}} + \frac{6+z^2}{z^3}e^{z} - \frac{8+4z}{z^4}e^{\frac{z}{2}} + \frac{4+z}{z^4}$	0
	$\frac{-4+z}{2z^3}e^{3z} + \frac{12-z+2z^2}{2z^3}e^{2z} - \frac{12+z+2z^2}{2z^3}e^{z} + \frac{4+z}{2z^3}$	–
	$\frac{-4+z}{z^3}e^{2z} + \frac{4-2z}{z^3}e^{\frac{3z}{2}} + \frac{8+6z+2z^2+z^3}{2z^3}e^{z} - \frac{4+2z}{z^3}e^{\frac{z}{2}}$	–
	$\frac{-4+z}{z^4}e^{3z} + \frac{16-2z+2z^2}{z^4}e^{2z} + \frac{-8+2z}{z^4}e^{\frac{3z}{2}} - \frac{12+7z+2z^2+z^3}{z^4}e^{z} + \frac{8+2z}{z^4}e^{\frac{z}{2}}$	–
	$\frac{-8+2z}{z^4}e^{\frac{5z}{2}} + \frac{-4+z}{z^4}e^{(2z)} + \frac{24+2z^2}{z^4}e^{\frac{3z}{2}} + \frac{-32+4z-2z^2+z^4}{2z^4}e^{z} - \frac{16+10z+2z^2}{z^4}e^{\frac{z}{2}} + \frac{12+7z+z^2}{z^4}$	–
	0	0
	$\frac{-4+z}{z^4}e^{2z} + \frac{4-2z}{z^4}e^{\frac{3z}{2}} + \frac{8+6z+2z^2+z^3}{2z^4}e^{z} - \frac{4+2z}{z^4}e^{\frac{z}{2}}$	0
	$\frac{-4+z}{2z^5}e^{3z} + \frac{20-3z+2z^2}{2z^5}e^{2z} + \frac{-8+4z}{z^5}e^{\frac{3z}{2}} - \frac{12+13z+2z^2+2z^3}{2z^5}e^{z} + \frac{8+4z}{z^5}e^{\frac{z}{2}} - \frac{4+z}{2z^5}$	0
	$\frac{-4+z}{z^4}e^{2z} + \frac{8+2z^2}{z^4}e^{z} - \frac{4+z}{z^4}$	$\frac{4-5z+z^2}{z^5}e^{2z} - \frac{8-4z}{z^5}e^{\frac{3z}{2}} + \frac{2+z^2}{z^4}e^{z} + \frac{8+4z}{z^5}e^{\frac{z}{2}} - \frac{4+5z+z^2}{z^5}$
	$\frac{-8+2z}{z^5}e^{\frac{5z}{2}} + \frac{8-2z}{z^5}e^{2z} + \frac{16+4z+2z^2}{z^5}e^{\frac{3z}{2}} + \frac{-32-8z-2z^2-2z^3+z^4}{2z^5}e^{z} - \frac{8+6z+2z^2}{z^5}e^{\frac{z}{2}} + \frac{8+6z+z^2}{z^5}$	0

TABLE 7. The error functions $\varphi_e(\tau)(z) - \psi(\tau)(z)$ and the error functions for consistent initial value $\varphi_s(\tau)(z) - \psi_c(\tau)(z)$ for the Radau IIA method.

τ	$\varphi_e(\tau)(z) - \psi(\tau)(z)$	$\varphi_s(\tau)(z) - \psi_c(\tau)(z)$
$\star$	0	–
$\circ$	0	0
	$\frac{-3+z}{2z^2}e^{2z} + \frac{3-3z}{2z^2}e^{\frac{4z}{3}} + \frac{3+2z+2z^2}{2z^2}e^{z} - \frac{3}{2z^2}e^{\frac{z}{3}}$	–
	0	0
	$\frac{-3+z}{2z^3}e^{2z} + \frac{3-3z}{2z^3}e^{\frac{4z}{3}} + \frac{3+2z+2z^2}{2z^3}e^{z} - \frac{3}{2z^3}e^{\frac{z}{3}}$	0
	$\frac{-3+z}{4z^2}e^{3z} + \frac{3+4z}{4z^2}e^{2z} + \frac{3-3z}{4z^2}e^{\frac{5z}{3}} - \frac{1}{2z}e^{z} - \frac{3}{4z^2}e^{\frac{2z}{3}}$	–
	$\frac{-3+z}{2z^2}e^{2z} + \frac{1-z}{2z^2}e^{\frac{4z}{3}} + \frac{3+2z+z^2}{2z^2}e^{z} - \frac{1}{2z^2}e^{\frac{z}{3}}$	–
	$\frac{-3+z}{2z^2}e^{3z} + \frac{6+3z}{2z^3}e^{2z} + \frac{3-3z}{2z^3}e^{\frac{5z}{3}} + \frac{-3+3z}{2z^3}e^{\frac{4z}{3}}$ $- \frac{3+4z+2z^2}{2z^3}e^{z} - \frac{3}{2z^3}e^{\frac{2z}{3}} + \frac{3}{2z^3}e^{\frac{z}{3}}$	–
	$\frac{-9+6z-z^2}{4z^4}e^{3z} + \frac{9-12z+3z^2}{2z^4}e^{\frac{7z}{3}} + \frac{9+15z-4z^2}{4z^4}e^{2z} + \frac{-9+18z-9z^2}{4z^4}e^{\frac{5z}{3}}$ $+ \frac{-9-3z+6z^2}{2z^4}e^{\frac{4z}{3}} + \frac{-6-2z+z^2}{2z^3}e^{z} + \frac{9-9z}{4z^4}e^{\frac{2z}{3}} + \frac{3}{z^3}e^{\frac{z}{3}}$	–
	$\frac{6-4z+z^2}{6z^3} - \frac{3+z}{3z^2}$	$\frac{z^2-4z+6}{6z^3}e^{z} - \frac{z+3}{3z^3}$
	$\frac{-3+z}{2z^3}e^{2z} - \frac{1-z}{2z^3}e^{\frac{4z}{3}} + \frac{3+2z+z^2}{2z^3}e^{z} - \frac{1}{2z^3}e^{\frac{z}{3}}$	0
	$\frac{-3+z}{4z^4}e^{3z} + \frac{9+2z}{4z^4}e^{2z} + \frac{3-3z}{4z^4}e^{\frac{5z}{3}} + \frac{-3+3z}{2z^4}e^{\frac{4z}{3}}$ $- \frac{3+3z+2z^2}{2z^4}e^{z} - \frac{3}{4z^4}e^{\frac{2z}{3}} + \frac{3}{2z^4}e^{\frac{z}{3}}$	0
	$\frac{-3+z}{2z^4}e^{2z} + \frac{3-3z}{2z^4}e^{\frac{4z}{3}} + \frac{3+2z+2z^2}{2z^4}e^{z} - \frac{3}{2z^4}e^{\frac{z}{3}}$	0
	$\frac{-9+6z-z^2}{4z^5}e^{3z} + \frac{9-12z+3z^2}{2z^5}e^{\frac{7z}{3}} + \frac{9+21z-6z^2}{4z^5}e^{2z} + \frac{-9+18z-9z^2}{4z^5}e^{\frac{5z}{3}}$ $+ \frac{-9-6z+9z^2}{2z^5}e^{\frac{4z}{3}} + \frac{-9-4z-2z^2+z^3}{2z^4}e^{z} + \frac{9-9z}{4z^5}e^{\frac{2z}{3}} + \frac{9}{2z^4}e^{\frac{z}{3}}$	0

TABLE 8. The error functions $\varphi_e(\tau)(z) - \psi(\tau)(z)$ and the error functions for consistent initial value $\varphi_s(\tau)(z) - \psi_c(\tau)(z)$ for the CMO3 method.

τ	$\varphi_e(\tau)(z) - \psi(\tau)(z)$	$\varphi_s(\tau)(z) - \psi_c(\tau)(z)$
$\star$	0	–
$\circ$	0	0
	$-\frac{9}{8z}e^{\frac{4z}{3}} + \frac{-1+z}{z}e^{z} + \frac{9}{4z}e^{\frac{2z}{3}} - \frac{1}{8z}$	–
	$\frac{1}{z^2}e^{z} - \frac{3}{4z}e^{\frac{2z}{3}} - \frac{4+z}{4z^2}$	$\frac{1}{z^2}e^{z} - \frac{3}{4z}e^{\frac{2z}{3}} - \frac{4+z}{4z^2}$
	$-\frac{9}{8z^2}e^{\frac{4z}{3}} + \frac{-1+z}{z^2}e^{z} + \frac{9}{4z^2}e^{\frac{2z}{3}} - \frac{1}{8z^2}$	0
	$-\frac{1}{16z}e^{2z} + \frac{9}{16z}e^{\frac{4z}{3}} - \frac{1}{z}e^{z} + \frac{9}{16z}e^{\frac{2z}{3}} - \frac{1}{16z}$	–
	$-\frac{3}{4z}e^{\frac{4z}{3}} + \frac{1}{2}e^{z} + \frac{3}{4z}e^{\frac{2z}{3}}$	–
	$-\frac{1}{8z^2}e^{2z} + \frac{9}{4z^2}e^{\frac{4z}{3}} - \frac{1+z}{z^2}e^{z} - \frac{9}{8z^2}e^{\frac{2z}{3}}$	–
	$-\frac{9}{8z^2}e^{\frac{5z}{3}} + \frac{9+2z^2}{4z^2}e^{z} - \frac{9}{8z^2}e^{\frac{z}{3}}$	–
	$\frac{1}{z^3}e^{z} - \frac{1}{4z}e^{\frac{2z}{3}} - \frac{4+4z+z^2}{4z^3}$	$\frac{1}{z^3}e^{z} - \frac{1}{4z}e^{\frac{2z}{3}} - \frac{4+4z+z^2}{4z^3}$
	$-\frac{3}{4z^2}e^{\frac{4z}{3}} + \frac{-2+z^2}{2z^3}e^{z} + \frac{3}{2z^2}e^{\frac{2z}{3}} + \frac{4+z}{4z^3}$	$-\frac{1}{z^3}e^{z} + \frac{3}{4z^2}e^{\frac{2z}{3}} + \frac{4+z}{4z^3}$
	$-\frac{1}{16z^3}e^{2z} + \frac{27}{16z^3}e^{\frac{4z}{3}} - \frac{1}{z^2}e^{z} - \frac{27}{16z^3}e^{\frac{2z}{3}} + \frac{1}{16z^3}$	0
	$-\frac{3}{8z^2}e^{\frac{4z}{3}} + \frac{-2+z}{z^3}e^{z} + \frac{3}{4z^2}e^{\frac{2z}{3}} + \frac{16+5z}{8z^3}$	$\frac{9-3z}{8z^3}e^{\frac{4z}{3}} - \frac{1}{z^3}e^{z} + \frac{-18+6z}{8z^3}e^{\frac{2z}{3}} + \frac{17+5z}{8z^3}$
	$-\frac{9}{8z^3}e^{\frac{5z}{3}} + \frac{9}{8z^3}e^{\frac{4z}{3}} + \frac{13-4z+2z^2}{4z^3}e^{z} - \frac{9}{4z^3}e^{\frac{2z}{3}} - \frac{9}{8z^3}e^{\frac{z}{3}} + \frac{1}{8z^3}$	0

TABLE 9. The error functions $\varphi_s(\tau)(z) - \psi_{s,i}(\tau)(z)$ for the first three steps.

CM3

τ	$\varphi_s - \psi_{s,0}$	$\varphi_s - \psi_{s,1}$	$\varphi_s - \psi_{s,2}$
	0	0	0
	$\frac{4+z}{z^3}$	0	0
	0	0	0
	$\frac{4+z}{z^4}$	0	0
	$\frac{4+z}{2z^3}$	0	0
	0	0	0
	0	0	0
	$\frac{12+7z+z^2}{z^4}$	$-\frac{(z_1+4)(z_0+4)}{z_1^4 z_0^2}$	0
	0	0	0
	0	0	0
	$-\frac{z+4}{2z^5}$	0	0
	$-\frac{z+4}{z^4}$	$-\frac{4+5z_1+z_1^2}{z_1^5}$	$-\frac{4+5z_2+z_2^2}{z_2^5}$
	$\frac{8+6z+z^2}{z^5}$	$-\frac{(z_1+4)(z_0+4)}{z_1^5 z_0^2}$	0

CMO3

τ	$\varphi_s - \psi_{s,0}$	$\varphi_s - \psi_{s,1}$	$\varphi_s - \psi_{s,2}$
	0	0	0
	$-\frac{1}{8z}$	0	0
	$-\frac{4+z}{4z^2}$	$-\frac{4+z_1}{4z_1^2}$	$-\frac{4+z_2}{4z_2^2}$
	$-\frac{1}{8z^2}$	0	0
	$-\frac{1}{16z}$	0	0
	0	0	0
	0	0	0
	0	$-\frac{1}{64z_1^2}$	0
	$-\frac{4+4z+z^2}{4z^3}$	$-\frac{4+4z_1+z_1^2}{4z_1^3}$	$-\frac{4+4z_2+z_2^2}{4z_2^3}$
	$\frac{z+4}{4z^3}$	$\frac{z+4}{4z^3}$	$\frac{z+4}{4z^3}$
	$\frac{1}{16z^3}$	0	0
	$\frac{5z+16}{8z^3}$	$\frac{64+20z_1-z_0}{32z_1^3}$	$-\frac{4+5z_2-z_2^2}{z^5}$
	$\frac{1}{8z^3}$	$-\frac{1}{64z_1^3}$	0