The Use of Butcher Series in the Analysis of Newton-Like Iterations in Runge-Kutta Formulas*

K. R. Jackson[†] A. Kværnø[‡] S. P. Nørsett[§]

April 1993.

Dedicated to Professor John C. Butcher
on the occasion of his 60th birthday.

Submitted to Annals of Numerical Mathematics.

Abstract

We consider the numerical solution of initial value problems for both ordinary differential equations and differential-algebraic equations by Runge-Kutta (RK) formulas. We assume that the internal stage values of the RK formula are computed by some iterative scheme for solving nonlinear equations, such as Newton's method. Using Butcher series and rooted trees, we give a complete characterization of the local error in the RK formula after k iterations of the scheme. Results are given for three specific schemes: simple iteration, the modified Newton iteration, and the full Newton iteration. The ideas developed in this paper, however, are easily applied to other iterative schemes of this kind.

*This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, the Information Technology Research Centre of Ontario, and the Norwegian Research Council for Science and Humanities.

[†]Computer Science Department, University of Toronto, Toronto, Ontario, Canada, M5S 1A4. E-mail: krj@cs.toronto.edu.

[‡]Division of Mathematical Sciences, Norwegian Institute of Technology, N–7034 Trondheim, Norway. E-mail: anne@imf.unit.no.

[§]Division of Mathematical Sciences, Norwegian Institute of Technology, N–7034 Trondheim, Norway. E-mail: norsett@imf.unit.no.

1 Introduction.

Consider the scalar initial value problem (IVP) for ordinary differential equations (ODEs)

$$y'(x) = f(y(x)), \qquad y(x_0) = y_0. \tag{1}$$

This problem can be solved by an s-stage p^{th}-order Runge-Kutta (RK) formula

$$\begin{aligned} Y &= y_0\varepsilon_s + hAf(Y), & (2)\\ y_1 &= y_0 + hb^T f(Y) & (3) \end{aligned}$$

where $A \in \mathbb{R}^{s\times s}$, $b \in \mathbb{R}^s$, $\varepsilon_s = [1, 1, \ldots, 1]^T \in \mathbb{R}^s$, the vector of stage values $Y = [Y_1, \ldots, Y_s]^T \in \mathbb{R}^s$ and $f(Y) = [f(Y_1), \ldots, f(Y_s)]^T$. Although we restrict our discussion to scalar autonomous problems to simplify the notation, all of our results apply to systems of nonautonomous problems.

If (1) is a nonlinear IVP and (2)–(3) is an implicit RK formula, then (2) is a system of nonlinear equations. Usually some variant of Newton's method of the form

$$\begin{aligned} (I_s - hAJ^{(k)})(Y^{(k+1)} - Y^{(k)}) &= y_0\varepsilon_s + hAf(Y^{(k)}) - Y^{(k)}, \qquad k = 0, 1, 2, \ldots, & (4)\\ y_1^{(k+1)} &= y_0\varepsilon_s + hb^T f(Y^{(k)}) & (5) \end{aligned}$$

is used to solve (2). Here $J^{(k)} = 0$ for simple iteration and $J^{(k)}$ is a diagonal matrix with some approximation to the Jacobian f_y on the diagonal for the Newton variants. The starting values $\{Y_i^{(0)}\}$ for the iteration (4) can be obtained by interpolating values from the previous step, by an explicit RK formula, or simply by taking $Y_i^{(0)} = y_0$ for $i = 1, \ldots, s$, called the *trivial predictor*. We refer to (4)–(5) together with the predictor for $Y^{(0)}$ as the *iterated RK method* associated with the underlying RK formula (2)–(3).

It is well known that the number of iterations of (4) used to "solve" (2) might affect the order of the approximate solution $y_1^{(k)}$. For a variety of formulas and problems, several authors have recently studied the maximum number of iterations needed to preserve the order of the underlying formula (2)–(3). See, for example, Sugiura and Torii [12] for nonstiff problems; Alexander [1], and van Dorsselaer and Spijker [3] for stiff systems; and Hairer, Lubich and Roche [5] for differential-algebraic equations (DAEs) of index 1, 2 and 3. See also the earlier work of Liniger [9]. These results are all based on the classical analysis of iterative schemes for systems of nonlinear equations.

In §2 we present some convergence results of this kind, together with some numerical experiments illustrating the shortcomings of this approach. In this way, we hope to motivate the use of Butcher series, commonly called B-series, in the analysis of iterative schemes in RK formulas. We provide a thorough discussion of the B-series approach in [10]. We summarize these results in §3, concentrating on how B-series can be used to explain the order behaviour observed in §2.

One possible criticism of the results in §2 and §3 is that the analysis appears to apply to nonstiff problems only. To address this potential deficiency, Kværnø [8] extended the B-series analysis of [10] to semi-explicit DAEs of index 1. In addition to being important in their own right, DAEs may be viewed as the limiting case of a large class of stiff ODEs (as explained in [4]). We summarize in §4 some of the results of [8]. We conclude in §5 with a short discussion of the strengths and weaknesses of using B-series to study Newton-like iterations in RK formulas.

2 Why Use Butcher Series?

Although many iterative schemes for solving (2) are of interest, we will concentrate on three specific schemes of the form (4):

- simple iteration, for which $J^{(k)} = 0$;
- the modified Newton iteration, for which $J^{(k)} = \mathrm{diag}(f_y(y_0), \ldots, f_y(y_0))$ is updated at the beginning of each step only; and
- the full Newton iteration, for which $J^{(k)} = \mathrm{diag}(f_y(Y_1^{(k)}), \ldots, f_y(Y_s^{(k)}))$ is updated each iteration.

Let

$$E_k = Y^{(k)} - Y = \mathcal{O}(h^{\rho_k + 1})$$

where Y is the exact solution of (2). Then, for h sufficiently small, it is straightforward to prove that

$$\begin{array}{rclcll} \|E_{k+1}\| & = & \mathcal{O}(h)\,\|E_k\| & \Rightarrow & \rho_{k+1} = \rho_k + 1 & \text{for simple iteration,} \\ \|E_{k+1}\| & = & \mathcal{O}(h^2)\,\|E_k\| & \Rightarrow & \rho_{k+1} = \rho_k + 2 & \text{for the modified Newton iteration,} \\ \|E_{k+1}\| & = & \mathcal{O}(h)\,\|E_k\|^2 & \Rightarrow & \rho_{k+1} = 2(\rho_k + 1) & \text{for the full Newton iteration,} \end{array}$$

which leads to the following result.

Theorem 1 *If*

$$Y^{(0)} - Y = \mathcal{O}(h^{r+1}),$$

then for $k \geq 1$

$$\begin{aligned} Y^{(k)} - Y &= \mathcal{O}(h^{\rho_k + 1}) \\ y_1^{(k)} - y(x_0 + h) &= \mathcal{O}(h^{\min(r_k, p) + 1}) \end{aligned}$$

where

$$\begin{aligned} \rho_k &= \begin{cases} r + k & \textit{for simple iteration,} \\ r + 2k & \textit{for the modified Newton iteration,} \\ 2^k(r+2) - 2 & \textit{for the full Newton iteration,} \end{cases} \\ r_k &= \rho_{k-1} + 1, \end{aligned}$$

and p *is the order of the underlying RK formula (2)–(3).*

See Sugiura and Torii [12] for a proof of the modified and full Newton cases.

The starting values $\{Y_i^{(0)}\}$ for the iteration (4) are frequently obtained by interpolating values from the previous step. If the interpolant is of order $q+1$ and $q \le p$, then

$$Y_i^{(0)} - y(x_0 + c_i h) = \mathcal{O}(h^{q+1}).$$

The *stage order* of the RK formula (2)–(3) is the largest integer η for which the simplifying assumption

$$C(\eta): \quad \sum_{j=1}^{s} a_{ij} c_j^{l-1} = \frac{1}{l} c_i^l, \qquad i = 1, \ldots, s, \qquad l = 1, \ldots, \eta,$$

holds. Therefore, if the underlying formula (2)–(3) has stage order η, then

$$Y^{(0)} - Y = \mathcal{O}(h^{r+1}).$$

where $r = \min(q, \eta)$. Thus, Theorem 1 suggests that, for formulas with low stage order, the use of a high order interpolant will be of little help in achieving high order in $y_1^{(k)}$ with k small. But, as the following example shows, this is not the case.

Example 1 We solved the problem

$$\begin{aligned} y_1' &= 4xy_1 + 2\sqrt{y_1}\,y_2, & y_1(.1) &= e^{.02}\sin^2(.1), \\ y_2' &= 2xy_2 + y_1 + y_2^2 - e^{x^2}(\sin(x) + e^{x^2}), & y_2(.1) &= e^{.01}\cos(.1), \end{aligned} \tag{6}$$

on the interval $[0.1, 0.5]$ with the iterated RK method (4)–(5) based on the underlying eighth-order explicit RK (ERK) formula [2, p. 208] together with a third-order ERK formula to compute the starting values $\{Y_i^{(0)}\}$ for the iteration. Thus, the starting values of the iteration satisfy $Y_i^{(0)} - y(x_n + c_i h) = \mathcal{O}(h^4)$, but $Y^{(0)} - Y = \mathcal{O}(h^2)$, since the stage order of an ERK formula can not exceed 1. Hence, $r = 1$ in Theorem 1.

We used a fixed stepsize h and the same number of iterations k on each step throughout any one numerical integration of (6), but we varied h and k from integration to integration. The global errors for several integrations with the same k but different h's were used to compute an *observed order* for the iterated RK method. These observed orders are given in Table 1 together with the orders predicted by Theorem 1. Note that the predicted order

Simple Iteration

k	$p_p^{(k)}$	$p_o^{(k)}$
1	2	4
2	3	5
3	4	6

Modified Newton Iteration

k	$p_p^{(k)}$	$p_o^{(k)}$
1	2	4
2	4	6
3	6	8

Full Newton Iteration

k	$p_p^{(k)}$	$p_o^{(k)}$
1	2	4
2	5	8
3	8	8

Table 1: Predicted order $p_p^{(k)} = \min(r_k, p)$ and observed order $p_o^{(k)}$ when k iterations are used in each step.

$p_p^{(k)} = \min(r_k, p)$ is always less than the observed order $p_o^{(k)}$ except for the full Newton iteration with $k = 3$, where $p_p^{(3)} = p_o^{(3)} = 8$, the order of the underlying RK formula. □

Example 1 illustrates that the order guaranteed by Theorem 1 may be significantly lower than the order actually obtained. In addition, this example shows that an iterated RK method can benefit from a high order predictor even if the stage order of the underlying RK formula is low, a result which is not implied by Theorem 1. A possible criticism of the conclusions above is that, for the RK formula used in Example 1, $b_i = 0$ for the stages responsible for the low stage order, giving a higher effective stage order. This is of course true, but several experiments with other RK formulas and other predictors give results similar to those in Example 1, confirming the benefit of high order predictors. In addition, these experimental observations are supported by the following theoretical result which we prove using B-series in [10].

Theorem 2 *If*

$$Y_i^{(0)} = y(x_0 + c_i h) + \mathcal{O}(h^{q+1}), \quad i = 1, \ldots, s,$$

with $q \geq 0$, then for $k \geq 1$

$$\begin{aligned} Y^{(k)} &= Y + \mathcal{O}(h^{\rho_k+1}), && (7)\\ y_1^{(k)} &= y(x_0 + h) + \mathcal{O}(h^{\min(p,q_k)+1}), && (8) \end{aligned}$$

where ρ_k is given in Theorem 1,

$$q_k = \begin{cases} q + k & \textit{for simple iteration,} \\ q + 2k - 1 & \textit{for the modified Newton iteration,} \\ 2^{k-1}(r+2) + q - r - 1 & \textit{for the full Newton iteration,} \end{cases}$$

$r = \min(q, \eta)$ and η is the stage order of the underlying RK formula.

The difference between the order of $y_1^{(k)}$ guaranteed by Theorems 1 and 2 is $q_k - r_k = q - r$ for $q_k \leq p$. It is difficult to see how this difference in orders can be explained without a detailed knowledge of the error $Y^{(k)} - Y$, which the B-series analysis provides, but the classical analysis does not.

Note that the observed orders $p_o^{(k)}$ in Table 1 always agree exactly with the orders guaranteed by Theorem 2 except for the full Newton iteration with $k = 2$, where the order ensured by Theorem 2 is 7 while the observed order $p_o^{(2)}$ is 8. This *order increase* also occurs in the next example. In both cases the phenomenon can be explained easily by a more careful examination of the B-series associated with the particular iterated RK method.

Example 2 Each member of the family of iterated RK methods considered in this example employs simple iteration and is based on the three-stage Kuntzmann-Butcher formula of order six [6, p. 204]. The methods differ only in their choice of predictor selected from the one-parameter family

$$Y_i^{(0)} = l_\omega(x_n + c_i h), \qquad i = 1, 2, 3, \tag{9}$$

where

$$l_\omega(x) = l_1(x) + \omega[l_1(x) - l_2(x)],$$

$$l_1(x) = y_n + (x - x_n) f(y_n) \quad \text{and} \quad l_2(x) = y_{n-1} + (x - x_{n-1}) f(y_{n-1}).$$

It is easy to show that $Y^{(0)} - Y = \mathcal{O}(h^2)$ for any ω and that there is no ω for which $Y^{(0)} - Y = \mathcal{O}(h^3)$. Therefore, $q = r = 1$ for this family of iterated RK methods.

Note that $Y_i^{(0)} = l_\omega(x_n + c_i h)$ is not applicable on the first step from x_0 to x_1, since y_{-1} is not defined. Therefore, we iterated the underlying RK formula to convergence on the first step and then switched to the iterated RK method for all remaining steps.

We solved problem (6) many times with this family of iterated RK methods, using a fixed stepsize h, the same number of iterations k and a constant ω on each step throughout any one numerical integration of (6), but varied h, k and ω from integration to integration. The global errors for several integrations with the same k and ω but different h's were used to compute an *observed order* for the iterated RK method. These observed orders are listed in Table 2 together with the orders guaranteed by Theorems 1 and 2. (Since $r = q$ here, the orders ensured by the two theorems are identical.) For most choices of ω, the observed

Simple Iteration

k	$p_p^{(k)}$	Observed Order $p_o^{(k)}$			
		$\omega = 0$	$\omega = \frac{1}{6}$	$\omega = \frac{1}{10}$	$\omega = \frac{1}{15}$
1	2	2	3	2	2
2	3	3	3	4	3
3	4	4	4	4	5
4	5	5	5	5	5

Table 2: Predicted order $p_p^k = \min(r_k, p)$ and observed order $p_o^{(k)}$ for the iterated RK method described in Example 2.

order equals that guaranteed by the theorems. However, for particular choices of ω and k, we see that the observed order $p_o^{(k)}$ is 1 more than expected. In these cases, we also see that $p_o^{(k)} = p_o^{(k+1)}$. This *order increase* phenomenon is explained in the next section. $\square$

3 How to Use Butcher Series.

A B-series is

$$B(a, y_0) = \sum_{t \in LT} a(t) F(t)(y_0) \frac{h^{\rho(t)}}{\rho(t)!},$$

where LT is the set of all labelled trees; $\rho(t)$, the order of the tree t, is the number of vertices in t; $F(t)(y_0)$ is the elementary differential for the problem (1) evaluated at y_0 associated with the tree t; and $a(t)$ is a possibly vector-valued coefficient associated with the tree t. $\emptyset$ and $\tau = [\emptyset]$ are the unique trees of order 0 and 1, respectively, while $t = [t_1, \ldots, t_m]$ is the tree formed by joining the root of each tree t_i, $i = 1, \ldots, m$, to a new common root for t. See [2] or [6] for a comprehensive introduction to B-series and trees.

It is well known that the exact solution of (1) and the numerical solution (2)–(3) can be written as B-series:

$$y(x_0+h) = B(1, y_0), \qquad Y = B(\phi, y_0), \qquad y_1 = B(\psi, y_0),$$

where $1(t) = 1$ for all t,

$$\begin{array}{lll} \phi(\emptyset) = \varepsilon_s, & \phi(\tau) = A\varepsilon_s = c, & \phi(t = [t_1, \ldots, t_m]) = \rho(t) A \prod_{i=1}^{m} \phi(t_i), \\ \psi(\emptyset) = 1, & \psi(\tau) = b^T \varepsilon_s, & \psi(t = [t_1, \ldots, t_m]) = \rho(t) b^T \prod_{i=1}^{m} \phi(t_i). \end{array} \tag{10}$$

We have used the standard convention in (10) that vector-vector multiplications are performed componentwise: that is,

$$\left(\prod_{i=1}^{m} \phi(t_i) \right)_j = \prod_{i=1}^{m} \phi_j(t_i), \qquad j = 1, \ldots, s.$$

Thus, the RK formula (2)–(3) is of order p if and only if

$$\psi(t) = 1 \quad \forall\, t \in LT \text{ s.t. } \rho(t) \leq p.$$

The iterated RK method (4)–(5) can also be written as a B-series:

$$Y^{(k)} = B(\phi^{(k)}, y_0), \qquad y_1^{(k)} = B(\psi^{(k)}, y_0). \tag{11}$$

Thus, to prove Theorem 2, it is sufficient to show that

$$\phi^{(k)}(t) = \phi(t) \quad \forall\, t \in LT \text{ s.t. } \rho(t) \leq \rho_k, \tag{12}$$

$$\psi^{(k)}(t) = 1 \quad \forall\, t \in LT \text{ s.t. } \rho(t) \leq \min(q_k, p). \tag{13}$$

To this end, we substitute the B-series (11) into the iterative scheme (4)–(5) to obtain a recurrence in both k and t for $\phi^{(k)}(t)$ and $\psi^{(k)}(t)$. Using this approach, we prove Theorem 2 and several related results in [10]. In this paper, though, we restrict ourselves to providing an example of the procedure outlined above for simple iteration with some comments on the more general results in [10].

For simple iteration, (4)–(5) reduces to

$$Y^{(k+1)} = y_0 \varepsilon_s + hAf(Y^{(k)}), \tag{14}$$

$$y_1^{(k+1)} = y_0 + hb^T f(Y^{(k)}). \tag{15}$$

Assuming $Y^{(k)}$, $Y^{(k+1)}$ and $y_1^{(k+1)}$ have B-series (which is the case if $Y^{(0)}$ does), then (14)–(15) can be rewritten as

$$B(\phi^{(k+1)}, y_0) = y_0 \varepsilon_s + hAf(B(\phi^{(k)}, y_0)), \tag{16}$$

$$B(\psi^{(k+1)}, y_0) = y_0 + hb^T f(B(\phi^{(k)}, y_0)). \tag{17}$$

Using the well-known key result [6, p. 246] for B-series

$$hf(B(a,y)) = B(a',y)$$

provided $a(\emptyset) = \varepsilon_l$ (for an l of the appropriate size, either $l=1$ or $l=s$ here), where

$$\begin{aligned} a'(\emptyset) &= 0, \\ a'(\tau) &= \varepsilon_l, \\ a'(t=[t_1,\ldots,t_m]) &= \rho(t)\prod_{i=1}^{m} a(t_i) \quad \text{for } \rho(t)\geq 2, \end{aligned}$$

it is straightforward to show that

$$\begin{array}{lll} \phi^{(k+1)}(\emptyset) = \varepsilon_s, & \psi^{(k+1)}(\emptyset) = 1, & \\ \phi^{(k+1)}(\tau) = A\varepsilon_s = c, & \psi^{(k+1)}(\tau) = b^T\varepsilon_s, & \text{if } \phi^{(k)}(\emptyset) = \varepsilon_s, \\ \phi^{(k+1)}(t) = \rho(t)A\prod_{i=1}^{m}\phi^{(k)}(t_i), & \psi^{(k+1)}(t) = \rho(t)b^T\prod_{i=1}^{m}\phi^{(k)}(t_i), & \text{for } t=[t_1,\ldots,t_m]. \end{array} \tag{18}$$

Therefore, if we know the B-series for $Y^{(0)}$,

$$Y^{(0)} = B(\phi^{(0)}, y_0),$$

we can compute the B-series for $Y^{(k+1)}$ and $y_1^{(k+1)}$ for any $k\geq 0$ from the recurrences (18).

To see how this works, consider an iterated RK method based on the simple iteration (16)–(17) that employs the trivial predictor $Y^{(0)} = y_0\varepsilon_s$. Thus,

$$\phi^{(0)}(\emptyset) = \varepsilon_s = \phi(\emptyset) \quad \text{and} \quad \phi^{(0)}(t) = 0 \quad \forall\, t\in LT \text{ s.t. } \rho(t)\geq 1. \tag{19}$$

Using (18) and (19) together with (10), we find

$$\begin{array}{ll} \phi^{(1)}(\emptyset) = \varepsilon_s = \phi(\emptyset), & \psi^{(1)}(\emptyset) = 1 = \psi(\emptyset), \\ \phi^{(1)}(\tau) = A\varepsilon_s = c = \phi(\tau), & \psi^{(1)}(\tau) = b^T\varepsilon_s = \psi(\tau), \\ \phi^{(1)}(t) = 0, & \psi^{(1)}(t) = 0, \qquad \text{for all other } t. \end{array} \tag{20}$$

and, continuing in a similar way using (18) and (20) together with (10), we also find

$$\begin{array}{lll} \phi^{(2)}(\emptyset) = \varepsilon_s = \phi(\emptyset), & \psi^{(2)}(\emptyset) = 1 = \psi(\emptyset), & \\ \phi^{(2)}(\tau) = A\varepsilon_s = c = \phi(\tau), & \psi^{(2)}(\tau) = b^T\varepsilon_s = \psi(\tau), & \\ \phi^{(2)}(t) = \rho(t)Ac^{\rho(t)-1} = \phi(t), & \psi^{(2)}(t) = \rho(t)b^Tc^{\rho(t)-1} = \psi(t), & \forall\, t=[\tau,\ldots,\tau] \\ \phi^{(2)}(t) = 0, & \psi^{(2)}(t) = 0, \qquad \text{for all other } t. & \end{array}$$

Similarly,

$$\begin{array}{lll} \phi^{(3)}(t) = \phi(t), & \psi^{(3)}(t) = \psi(t), & \forall\, t\in LT_3, \\ \phi^{(3)}(t) = 0, & \psi^{(3)}(t) = 0, & \forall\, t\notin LT_3, \end{array}$$

where

$$LT_3 = \{\emptyset,\, \tau,\, [\tau,\dots,\tau],\, [t_1,\dots,t_m] : \text{ each } t_i \text{ is of the form } [\tau,\dots,\tau]\}.$$

Now, to make the emerging pattern more clear, we introduce the *height function*, $H(t)$, defined by Butcher [2] to be 1 plus the length of the longest path in the tree t, or, more formally,

$$H(\emptyset) = 0, \quad H(\tau) = 1, \quad H(t = [t_1,\dots,t_m]) = 1 + \max_{i=1,\dots,m} H(t_i).$$

It follows easily by induction on k that

$$\begin{array}{rclrclr} \phi^{(k)}(t) & = & \phi(t), & \psi^{(k)}(t) & = & \psi(t), & \forall\, t \in LT \text{ s.t. } H(t) \le k, \\ \phi^{(k)}(t) & = & 0, & \psi^{(k)}(t) & = & 0, & \forall\, t \in LT \text{ s.t. } H(t) > k. \end{array} \tag{21}$$

Thus, the height of the tree t divides LT into two subsets consisting of (1) those trees t for which $\phi^{(k)}(t) = \phi(t)$, and (2) those trees t for which $\phi^{(k)}(t) = 0$. (These two subsets are not necessarily disjoint, but often they are.)

Because $H(t) \le \rho(t)$ for all $t \in LT$, (21) implies that

$$\phi^{(k)}(t) = \phi(t) \quad \text{and} \quad \psi^{(k)}(t) = \psi(t) \quad \forall\, t \in LT \text{ s.t. } \rho(t) \le k,$$

whence

$$Y^{(k)} = Y + \mathcal{O}(h^{k+1}) \quad \text{and} \quad y_1^{(k)} = y_1 + \mathcal{O}(h^{k+1})$$

in exact accordance with Theorem 2, since $q = r = 0$ for the trivial predictor $Y^{(0)} = y_0\varepsilon_s$. Now note that the only tree t for which $\rho(t) = k+1$ and $H(t) > k$ is $t_k = [_k\tau]_k$, the branchless tree of height $k+1$. (The notation $[_k$ and $]_k$ stands for k occurrences of [and], respectively.) Thus, if it happens that $\phi(t_k) = 0$, then $Y^{(k)} = Y + \mathcal{O}(h^{k+2})$. Similarly, if $\psi(t_k) = 0$, then $y_1^{(k)} = y_1 + \mathcal{O}(h^{k+2})$ and the iterated RK method will exhibit an "unexpected" order increase if $k+1 \le p$. Alternatively, if the problem (1) is such that the elementary differential $F(t_k)(y) = 0$ throughout the integration, then the observed order for this problem will be at least $k+1$, rather than k, assuming $k+1 \le p$. In this sense, the order of this iterated RK method is close to $k+1$ for $k+1 \le p$.

If the predictor $Y^{(0)}$ satisfies

$$\phi^{(0)}(t) = \phi(t) \qquad \forall\, t \in LT \text{ s.t. } \rho(t) \le r,$$

then it follows from the recurrence (18) that $\phi^{(1)}(t) = \phi(t)$ and $\psi^{(1)}(t) = \psi(t)$ for all trees of the form $t = [t_1,\dots,t_m]$ for which $\rho(t_i) \le r$, $i = 1,\dots,m$. Similarly, $\phi^{(2)}(t) = \phi(t)$ and $\psi^{(2)}(t) = \psi(t)$ for all trees t composed from such trees, etc. This motivates our definition [10] of $H(t,r)$, the *r-lifted height* of the tree t:

$$\begin{array}{rcll} H(t,r) & = & 0 & \forall\, t \in LT \text{ s.t. } \rho(t) \le r, \\ H(t,r) & = & 1 & \forall\, t \in LT \text{ s.t. } \rho(t) = r+1, \\ H(t,r) & = & 1 + \max_i H(t_i,r) & \forall\, t \in LT \text{ s.t. } \rho(t) \ge r+2, \quad t = [t_1,\dots,t_m]. \end{array}$$

It is now easy to prove by induction on k (see [10]) that

$$\phi^{(k)}(t) = \phi(t) \quad \text{and} \quad \psi^{(k)}(t) = \psi(t) \quad \forall\, t \in LT \text{ s.t. } H(t,r) \le k. \tag{22}$$

The set of trees satisfying $H(t,r) \leq k$ obviously includes all trees for which $\rho(t) \leq r+k$, giving

$$Y^{(k)} = Y + \mathcal{O}(h^{\rho_k+1}) \tag{23}$$

for $\rho_k = r+k$, establishing (7) for the simple iteration case of Theorem 2. In addition, it is worth noting that the only trees satisfying both $\rho(t) = r+k+1$ and $H(t,r) > k$ are the *minimal order trees* of the form $t = [_k t_r]_k$ where t_r is any tree of order $r+1$. Since these trees form a small fraction of all trees of order $r+k+1$ for $k>1$, the order of this iterated RK method is close to $r+k+1$ for $r+k+1 \leq p$.

The minimal order trees can be used to explain the order increase phenomenon illustrated in Example 2. The B-series for the predictor (9) is

$$\begin{aligned} \phi^{(0)}(\emptyset) &= \varepsilon_s, \\ \phi^{(0)}(\tau) &= c, \\ \phi^{(0)}(t) &= (-1)^{\rho(t)}\omega[(\rho(t)-1)\varepsilon_s + \rho(t)c] \qquad \forall\, t \in LT \text{ s.t. } \rho(t) \geq 2. \end{aligned} \tag{24}$$

This shows not only that $Y = y(x_0 + ch) + \mathcal{O}(h^2)$ for any choice of ω but also that there is no ω for which $Y_i^{(0)} = y(x_0 + c_i h) + \mathcal{O}(h^3)$ for all three c_i's associated with the Kuntzmann-Butcher formula. Therefore, the initial guess $Y^{(0)}$ is of order $q = r = 1$. Thus, on the first iteration, the only minimal order tree is $t = [[\tau]]$ and, from (10), (18) and (24), we find

$$\begin{aligned} \phi^{(1)}([[\tau]]) &= 3\omega(A\varepsilon_s + 2Ac) &&= 3\omega(\phi(\tau) + \phi([\tau])) &&= 3\omega(c + c^2), \\ \psi^{(1)}([[\tau]]) &= 3\omega(b^T\varepsilon_s + 2b^Tc) &&= 3\omega(\psi(\tau) + \psi([\tau])) &&= 6\omega, \end{aligned}$$

since the three-stage Kuntzmann-Butcher formula is of order 6 with stage order 3. Thus $\psi^{(1)}([[\tau]]) = 1$ if and only if $\omega = 1/6$. Thus, the iterated RK method of Example 2 with $k = 1$ is of order 3 for $\omega = 1/6$ and of order 2 for all other choices of ω. However, even if $y_1^{(1)} - y_1 = \mathcal{O}(h^4)$ for $\omega = 1/6$, we still have $Y^{(1)} - Y = \mathcal{O}(h^3)$ and not $Y^{(1)} - Y = \mathcal{O}(h^4)$, since there is no ω for which $\phi_i^{(1)}([[\tau]]) = \phi_i([[\tau]])$ for $i = 1,2,3$. Similarly, we have $\psi^{(2)}([_3\tau]_3) = 1$ if and only if $\omega = 1/10$ and $\psi^{(3)}([_4\tau]_4) = 1$ if and only if $\omega = 1/15$, which explains the order increase phenomenon illustrated in Table 2.

Example 2 and the discussion above illustrates that the output $y_1^{(k)}$ of an iterated RK method may obtain a higher order than the internal stage values $Y^{(k)}$. Theorem 2 improves upon Theorem 1 in this respect. To see how this comes about, assume the predictor satisfies $Y_i^{(0)} = y(x_0 + c_i h) + \mathcal{O}(h^{q+1})$, whence the associated B-series satisfies

$$\phi^{(0)}(t) = c^{\rho(t)} \quad \forall\, t \in LT \text{ s.t. } \rho(t) \leq q. \tag{25}$$

Therefore, if $t = [t_1, \ldots, t_m]$ with $\rho(t_i) \leq q$ for $i = 1, \ldots, m$, then $H(t,q) \leq 1$ and, on the first iteration, we get

$$\begin{aligned} \phi^{(1)}(t) &= \rho(t)A\prod_{i=1}^{m}\phi^{(0)}(t_i) &&= \rho(t)A\prod_{i=1}^{m}c^{\rho(t_i)} &&= \rho(t)Ac^{\rho(t)-1} &&= \phi([\tau^{\rho(t)-1}]), \\ \psi^{(1)}(t) &= \rho(t)b^T\prod_{i=1}^{m}\phi^{(0)}(t_i) &&= \rho(t)b^T\prod_{i=1}^{m}c^{\rho(t_i)} &&= \rho(t)b^Tc^{\rho(t)-1} &&= \psi([\tau^{\rho(t)-1}]), \end{aligned}$$

where $[\tau^m] = [\tau, \ldots, \tau]$ is the *bushy* tree with τ repeated m times of height 2 and order $m+1$. So, even though $\phi^{(1)}(t)$ may not equal $\phi(t)$ for all trees satisfying $H(t,q) \leq 1$, $\phi^{(1)}(t)$ does equal the stage weight of some tree of the same order. To generalize this observation, let

$$\Phi_\kappa = \{\phi(t) : \rho(t) = \kappa\} \quad \text{and} \quad \Psi_\kappa = \{\psi(t) : \rho(t) = \kappa\}.$$

Then, under assumption (25), we can easily prove by induction on $k \geq 1$ that

$$\phi^{(k)}(t) \in \Phi_{\rho(t)} \quad \text{and} \quad \psi^{(k)}(t) \in \Psi_{\rho(t)} \quad \forall\, t \in LT \text{ s.t. } H(t,q) \leq k. \tag{26}$$

Since $\psi(t) = 1$ for $\rho(t) \leq p$, $\psi^{(k)}(t) = 1$ if $\rho(t) \leq p$ and $H(t,q) \leq k$, whence $\psi^{(k)}(t) = 1$ if $\rho(t) \leq \min(p, q_k)$ for $q_k = q + k$, since $\rho(t) \leq q + k$ implies $H(t,q) \leq k$. As this implies that (8) holds for the simple iteration case of Theorem 2 and we have already seen from (23) that (7) holds in this case too, we have proved Theorem 2 for simple iteration.

A similar analysis is given in [10] to establish Theorem 2 for the modified and full Newton iterations also. The key to the proof for these and other iteration schemes is to find a recurrence for the B-series for $Y^{(k)}$ similar to (18) and to find relations similar to (22) and (26). In particular, we must replace the height function $H(t,r)$ by another *growth function* $\mathcal{G}(t,r)$ appropriate for the iteration under consideration. The essential property that we require of $\mathcal{G}(t,r)$ is that $\mathcal{G}(t,r) \leq k$ if and only if $\phi^{(k)}(t)$ depends only on predictor stage weights $\phi^{(0)}(t_i)$ for trees t_i satisfying $\rho(t_i) \leq r$. We show in [10] that, for the modified Newton iteration, the growth function is $R(t,r)$, which is roughly speaking equal to 1 plus the maximum number of ramifications along any path from the root of the tree to the first subtree of order $r+1$ or less, and, for the full Newton iteration, the growth function is $D(t,r)$, which is again roughly speaking equal to 1 plus the maximum number of times the number of branches is doubled along any path from the root of the tree to the first subtree of order $r+1$ or less. Formal definitions of $R(t,r)$ and $D(t,r)$ are given in [10].

The growth function associated with an iterated RK method gives a much more complete description of the convergence of the method than the order alone. For example, given an appropriate growth function for an iterated RK method, it immediately follows that, if

$$\phi^{(0)}(t) = \phi(t) \quad \forall\, t \in LT \text{ s.t. } \rho(t) \leq r,$$

then

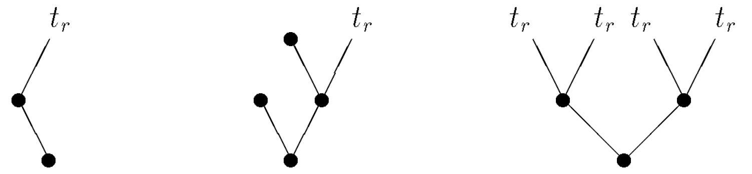
$$\phi^{(k)}(t) = \phi(t) \quad \forall\, t \in LT \text{ s.t. } \mathcal{G}(t,r) \leq k, \tag{27}$$

which implies that

$$Y^{(k)} = Y + \mathcal{O}(h^{\rho_k+1}), \tag{28}$$

for $\rho_k = \rho(t_{\mathcal{G},k+1}) - 1$, where $t_{\mathcal{G},k+1}$ is a tree of lowest possible order for which $\mathcal{G}(t,r) = k+1$. But (28) does not imply (27), except in trivial cases. Thus, (27) gives much more information about the iterated RK method than (28) does. In particular, we may see from (27) that there are relatively few minimal order trees $t_{\mathcal{G},k+1}$ in the collection of all trees of order $\rho_k + 1$. Hence, $Y^{(k)} - Y$ may be "close" to order $\rho_k + 1$ or more, as was explained for the examples discussed earlier. This could have serious implications for error estimation.

Minimal order trees for the growth functions $H(t,r)$, $R(t,r)$ and $D(t,r)$ are shown in Figure 1, where each occurrence of t_r stands for a possibly different subtree of order $r+1$.



$t_{H,3} = [_2 t_r]_2$ $\quad t_{R,3} = [\tau[\tau t_r]_2$ $\quad t_{D,3} = [[t_r t_r][t_r t_r]]$

Figure 1: The "minimal order trees" $t_{H,3}$, $t_{R,3}$ and $t_{D,3}$ for simple iteration, the modified Newton iteration and the full Newton iteration, respectively.

To understand the convergence of $y_1^{(k)}$ to $y(x_0 + h)$, we also need results similar to (26). For the modified Newton iteration, we can simply replace $H(t, q)$ by $R(t, q)$ in (26) to obtain the relevant relation for this iteration scheme. However, for the full Newton iteration, the analysis is somewhat more complicated. Roughly speaking, the trees restricting the order of $y_1^{(k)}$ are similar to $t_{D,3}$ shown in Figure 1, but with one subtree t_r replaced by a subtree t_q of order $q + 1$. Expanding upon this outline, we prove Theorem 2 for the modified and full Newton iterations in [10].

However, all the discussion above depends on the classical B-series analysis for ODEs, which is valid in the limit as $h \to 0$. As pointed out by van Dorsselaer and Spijker [3], for instance, this $h \to 0$ assumption may not be appropriate for stiff problems. We address this concern in the next section.

4 Differential-Algebraic Equations of Index 1.

In this section, we give some results for Newton-like iterations used to solve for the internal stage values of RK formulas applied to IVPs for semi-explicit index 1 differential-algebraic equations (DAEs) of the form

$$\begin{array}{rcl} y' &=& f(y, z), \\ 0 &=& g(y, z), \end{array} \qquad \begin{array}{l} y(x_0) = y_0, \\ z(x_0) = z_0, \end{array} \qquad g(y_0, z_0) = 0, \tag{29}$$

where $g_z(y, z)$ is assumed to be nonsingular. (29) is of interest not only because problems of this form arise in many important applications, but also because it can be viewed as the limiting case of the IVP associated with the stiff singularly-perturbed ODE

$$\begin{array}{r} y' = f(y, z), \\ \epsilon z' = g(y, z), \end{array} \qquad \epsilon \to 0,$$

as described in [4]. Thus, by providing results for (29), we hope to meet some of the objections one could make to the application of the theory presented in §3 to stiff ODEs.

The B-series analysis of Newton-type iterations for (29) is given by Kværnø [8]. Here, we briefly mention only the most important and surprising results. To begin, note that an implicit Runge-Kutta formula with nonsingular coefficient matrix A applied to (29) can be written as

$$Y = y_0\varepsilon_s + hAf(Y,Z), \qquad g(Y,Z) = 0 \tag{30}$$

$$y_1 = R(\infty)y_0 + b^T A^{-1}Y, \qquad z_1 = R(\infty)z_0 + b^T A^{-1}Z \tag{31}$$

where $R(\infty) = 1 - b^T A^{-1}\varepsilon_s$. The Newton iteration scheme is given by

$$\begin{bmatrix} I - hAJ_{11}^{(k)} & -hAJ_{12}^{(k)} \\ -J_{21}^{(k)} & -J_{22}^{(k)} \end{bmatrix} \begin{bmatrix} Y^{(k+1)} - Y^{(k)} \\ Z^{(k+1)} - Z^{(k)} \end{bmatrix} = \begin{bmatrix} y_0\varepsilon_s + hAf(Y^{(k)}, Z^{(k)}) - Y^{(k)}, \\ g(Y^{(k)}, Z^{(k)}), \end{bmatrix} \tag{32}$$

and

$$y_1^{(k)} = R(\infty)y_0 + b^T A^{-1}Y^{(k)}, \qquad z_1^{(k)} = R(\infty)z_0 + b^T A^{-1}Z^{(k)}, \tag{33}$$

where $J_{11}^{(k)}$, $J_{12}^{(k)}$, $J_{21}^{(k)}$ and $J_{22}^{(k)}$ in (32) are matrices corresponding to $J^{(k)}$ in (4). Since simple iteration is not appropriate for this application, only the modified and full Newton iterations are considered. $J_{11}^{(k)} = \mathrm{diag}(f_y(y_0,z_0),\ldots,f_y(y_0,z_0))$ for the former scheme and $J_{11}^{(k)} = \mathrm{diag}(f_y(Y_1^{(k)}, Z_1^{(k)}),\ldots,f_y(Y_s^{(k)}, Z_s^{(k)}))$ for the latter. The other $J_{ij}^{(k)}$ are defined similarly.

Roche [11] gives a complete tree theory for RK formulas applied to (29). He defines a *DA-series* as

$$\begin{aligned} DA_y(a,y,z) &= \sum_{t\in LDAT_y} a(t)F(t)(y,z)\frac{h^{\rho(t)}}{\rho(t)!}, \\ DA_z(b,y,z) &= \sum_{u\in LDAT_z} b(u)G(u)(y,z)\frac{h^{\rho(u)}}{\rho(u)!}. \end{aligned}$$

The elementary differentials $F(t)$ and $G(u)$ are associated with the trees t and u, respectively, exactly as is the ODE case. The DA-trees, however, have two different kinds of vertices: *meager* ($\cdot$) and *fat* ($\circ$). The order $\rho(t)$ of a tree t is the number of meager vertices in t. Trees corresponding to the elementary differentials of the y-component have a meager root; those corresponding to z have a fat one. The exact solution $y(x)$ and $z(x)$ of (29), the stage values Y and Z, and the numerical solution y_1 and z_1 of the RK formula (30)–(31) can all be expressed as DA-series:

$$\begin{aligned} y(x_0+h) &= DA_y(1,y_0,z_0), & z(x_0+h) &= DA_z(1,y_0,z_0), \\ Y &= DA_y(\phi_y,y_0,z_0), & Z &= DA_z(\phi_z,y_0,z_0), \\ y_1 &= DA_y(\psi_y,y_0,z_0), & z_1 &= DA_z(\psi_z,y_0,z_0). \end{aligned}$$

The stage values $Y^{(k)}$ and $Z^{(k)}$ and the numerical solution $y_1^{(k)}$ and $z_1^{(k)}$ of the iterated RK method (32)–(33) can also be written as DA-series in a similar way. The RK formula (30)–(31) is said to be of local order (p_y+1, p_z+1) if and only if p_y is the largest integer such that $\psi_y(t) = 1$ for all $t \in LDAT_y$ satisfying $\rho(t) \le p_y$ and p_z is the largest integer such that $\psi_z(u) = 1$ for all $u \in LDAT_z$ satisfying $\rho(u) \le p_z$. From (31) and (33), we see that

$\psi_y^{(k)}(t) = \psi_y(t)$ if $\phi_y^{(k)}(t) = \phi_y(t)$ and similarly that $\psi_z^{(k)}(u) = \psi(u)$ if $\phi_z^{(k)}(u) = \phi_z(u)$. Thus the order of convergence of $y_1^{(k)}$ and $z_1^{(k)}$ to y_1 and z_1, respectively, will be at least as large as the order of convergence of $Y^{(k)}$ and $Z^{(k)}$ to Y and Z, respectively. An analysis similar to the one done for ODEs can be performed. As might be expected, the convergence of $Y^{(k)}$ and $Z^{(k)}$ to Y and Z, respectively, can be expressed in terms of growth functions here too. More surprisingly, though, *the growth functions for DAEs are exactly the same as for ODEs.* Thus we have the following result.

Theorem 3 *If*

$$\phi_y^{(0)}(t) = \phi_y(t) \quad \textit{and} \quad \phi_z^{(0)}(u) = \phi_z(u) \quad \forall\, t,\, u \in LDAT \ \textit{s.t.}\ \rho(t),\, \rho(u) \le r,$$

then

$$\phi_y^{(k)}(t) = \phi_y(t) \quad \textit{and} \quad \phi_z^{(k)}(u) = \phi_z(u) \quad \forall\, t,\, u \in LDAT \ \textit{s.t.}\ \mathcal{G}(t,r),\, \mathcal{G}(u,r) \le k,$$

where $\mathcal{G}(t,r)$ is given by $R(t,r)$ for the modified Newton iteration and by $D(t,r)$ for the full Newton iteration.

Even though $\mathcal{G}(t,r)$ is the same for ODEs and DAEs, whence the minimal order trees are the same too, the order increases more slowly in the DAE case because DAE trees consist of both fat and meager vertices, and only the meager vertices contribute to the order. Thus, a DAE tree with some fat vertices has lower order than the corresponding ODE tree. Furthermore, because the trees corresponding to y-components have meager roots, while the trees corresponding to z-components have fat ones, the minimal order trees are of one order less for z than for y. See Figure 2 for some examples of minimal order DAE trees.

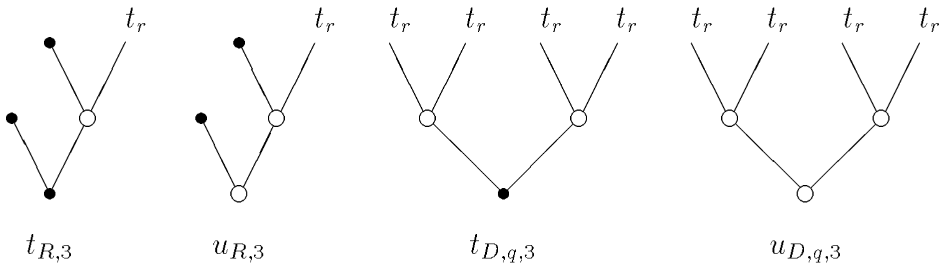


Figure 2: Some minimal order DAE trees.

As in the ODE case, the growth function can be used to characterize the trees t and $\hat{t}$ for which $\phi_y^{(k)}(t) = \phi_y(\hat{t})$ with $\rho(t) = \rho(\hat{t})$ and $\psi_y^{(k)}(t) = b^T A^{-1} \phi_y^{(k)}(t) = b^T A^{-1} \phi_y(\hat{t}) = 1$, and similarly for the z-components. Following this outline, Kværnø [8] proves the following result for DAEs, similar to Theorem 2 for ODEs.

Theorem 4 *If the RK formula (30)–(31) has stage order η and*

$$Y_i^{(0)} = y(x_0 + c_i h) + \mathcal{O}(h^{q+1}), \qquad Z_i^{(0)} = z(x_0 + c_i h) + \mathcal{O}(h^{q+1}), \qquad i = 1, \cdots, s,$$

with $q \geq 0$*, then for* $k \geq 1$

$$\begin{aligned} Y^{(k)} &= Y + \mathcal{O}(h^{\rho_k+2}) & \qquad Z^{(k)} &= Z + \mathcal{O}(h^{\rho_k+1}) \\ y_1^{(k)} &= y(x_0+h) + \mathcal{O}(h^{\min(p_y+1,q_k+2)}) & \qquad z_1^{(k)} &= z(x_0+h) + \mathcal{O}(h^{\min(p_z+1,q_k+1)}) \end{aligned}$$

where

$$\begin{array}{lll} \rho_k = r + k, & q_k = q + k, & \textit{for the modified Newton iteration,} \\ \rho_k = 2^k(r+1) - 1, & q_k = 2^k(r+1) + q - r - 1, & \textit{for the full Newton iteration,} \end{array}$$

with $r = \min(\eta, q)$.

Van Dorsselaer and Spijker [3] give convergence results for Newton-like iterations for ODEs satisfying the contractivity condition $\|y(x_0+h) - \tilde{y}(x_0+h)\| \leq \|y(x_0) - \tilde{y}(x_0)\|$, where $y(x)$ and $\tilde{y}(x)$ are both solutions to the same ODE. They establish the rate of convergence quoted in Theorem 4 for the stage values $Y^{(k)}$ and $Z^{(k)}$, which implies, as noted above, that $y_1^{(k)}$ and $z_1^{(k)}$ converge at least as fast. They did not, though, prove the stronger convergence result given in Theorem 4 for $y_1^{(k)}$ and $z_1^{(k)}$.

5 Discussion.

We conclude with some remarks about the use of B-series in the analysis of iterated RK methods, including some advantages and disadvantages of this approach.

As we have seen, an iterated RK method, consisting of a predictor, a corrector (the underlying RK formula), and k iterations of a scheme to solve for the internal stage values of the RK formula, can be viewed as a generalized RK method. The central result of this paper is that the internal stage values and output of such a method can be expressed as B-series in terms of the B-series of the predictor. A change in the predictor, the underlying RK formula, the iteration scheme, or even the number of iterations k of the scheme gives a new iterated RK method with (most likely) a different B-series for the numerical solution.

Although we consider in detail only three particular iteration schemes in this paper, we believe that our approach is applicable to other iteration schemes as well. For instance, van der Houwen, Sommeijer and Couzy [7] note that there is a parallel computing advantage to taking the Newton iteration matrix to be $I - hMJ^{(k)}$, for a carefully chosen diagonal matrix M, rather than the usual $I - hAJ^{(k)}$. Their class of iterated RK methods seems to be amenable to analysis by a straightforward modification of the B-series approach outlined in §3.

Theorems 2 and 4 give convergence results that usually agree with the observed order of an iterated RK method, unlike Theorem 1, which is frequently too pessimistic, as illustrated in Example 1. However, even Theorems 2 and 4 can be pessimistic in the sense that the order of a particular iterated RK method, or a method applied to a particular problem, can exhibit a higher than expected order. However, this can be explained by a more careful examination of the growth function associated with the method and the minimal order trees in particular, as illustrated in Example 2 and the following discussion in §3.

Furthermore, knowledge of the B-series for an iterated RK method warns us to be careful in choosing an error estimate for the method. For example, the difference between two subsequent iterations gives an error estimate of the desired order for the terms in the B-series for the error associated with the minimal order trees only. This may result in a poor estimate of the true local error of the method.

The weakness of the B-series approach is that it is based on an asymptotic analysis that is valid in the limit as $h \to 0$, but may fail to reflect accurately the behaviour of a method using large stepsizes, as is routine in solving stiff problems. In particular, we are not even guaranteed that $Y^{(k)} \to Y$ as $k \to \infty$ except for *"sufficiently small h"*. Despite this deficiency shared by the B-series analysis of numerical methods for ODEs, this technique has proven to be an essential analytic tool for the latter class of methods, and we believe the approach outlined in §3 will be useful in gaining a better understanding of iterated RK methods as well. In addition, to address this weakness, we outline in §4 how the B-series analysis for ODEs can be extended to DAEs, which, as explained in §4, can be viewed as the limiting case of a large class of stiff ODEs. More importantly, though, we view the B-series analysis for iterated RK methods as complementing, rather than competing with, other approaches, such as that of van Dorsselaer and Spijker [3], for gaining a better understanding of these methods.

References

[1] R. Alexander, *The modified Newton method in the solution of stiff ordinary differential equations*, Math. Comp., 57 (1991), pp. 673–701.

[2] J. C. Butcher, *The Numerical Analysis of Ordinary Differential Equations: Runge-Kutta and General Linear Methods*, John Wiley & Sons, 1987.

[3] J. L. M. van Dorsselaer and M. N. Spijker, *The error committed by stopping Newton iterations in the numerical solution of stiff initial value problems*, Report nr. TW–92–02, Dept. of Mathematics and Computer Science, Division of Applied Mathematics, Leiden University, 1992.

[4] E. Hairer, Ch. Lubich and M. Roche, *Error of Runge-Kutta methods for stiff problems studied via differential algebraic equations*, BIT, 28 (1988), pp. 678–700.

[5] E. Hairer, Ch. Lubich and M. Roche, *The numerical solution of differential-algebraic systems by Runge-Kutta methods*, Lecture Notes in Math. 1409, Springer Verlag, 1989.

[6] E. Hairer, S. P. Nørsett and G. Wanner, *Solving Ordinary Differential Equations I: Nonstiff Problems*, Springer Verlag, 1987.

[7] P. J. van der Houwen, B. P. Sommeijer and W. Couzy, *Embedded diagonally implicit Runge-Kutta algorithms on parallel computers*, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands, 1989.

[8] A. Kværnø, *The order of Runge-Kutta methods applied to semi-explicit DAEs of index 1, using Newton-type iterations to compute the internal stage values*, Numerics no. 2/1992, Dept. of Math. Sciences, The University of Trondheim, 1992.

[9] W. Liniger, *A stopping criterion for the Newton-Raphson method in implicit multistep integration algorithms for nonlinear systems of ordinary differential equations*, Comm. ACM, 14 (1971), pp. 600–601.

[10] K. R. Jackson, A. Kværnø and S. P. Nørsett, *The order of a Runge-Kutta method incorporating an iterative scheme to compute the internal stage values*, in preparation.

[11] M. Roche, *Implicit Runge-Kutta methods for differential algebraic equations*, SIAM J. Numer. Anal., 26 (1989), pp. 963–975.

[12] H. Sugiura and T. Torii, *A method for constructing generalized Runge-Kutta methods*, J. Comp. Appl. Math., 39 (1991), pp. 300–410.