SIAM J. NUMER. ANAL.
Vol. 27, No. 2, pp. 447–456, April 1990

009

ON ORDER REDUCTION FOR RUNGE–KUTTA METHODS APPLIED TO DIFFERENTIAL/ALGEBRAIC SYSTEMS AND TO STIFF SYSTEMS OF ODES*

KEVIN BURRAGE† AND LINDA PETZOLD‡

Abstract. In this short note, the order reduction results of Petzold [SIAM *J. Numer. Anal.* 23(1986), pp. 837–852] for implicit Runge–Kutta methods applied to index 1 differential/algebraic systems are extended to include a larger class of methods. The relationship between the order reduction results for differential/algebraic systems and recent results for stiff systems of ordinary differential equations is explained.

Key words. differential/algebraic systems, stiff equations

AMS(MOS) subject classification. 65L05

1. Introduction.

Recently some attention has been paid to the behavior of certain classes of methods when applied to systems of differential algebraic equations (DAEs) of the form

$$F(t, y, y') = 0 \tag{1.1}$$

where the initial values of y are given at $t = 0$. In particular, Petzold [11] has considered the effect on the order and stability of Runge–Kutta methods of the form

$$\begin{aligned} &F\left(t_{n-1} + c_i h, y_{n-1} + h\sum_{j=1}^{M} a_{ij} Y_j', Y_i'\right) = 0, \quad i = 1, 2, \cdots, M, \\ &y_n = y_{n-1} + h\sum_{i=1}^{M} b_i Y_i', \end{aligned} \tag{1.2}$$

when applied to such problems of index 1. Under the assumption that the matrix $\mathcal{A}$ of coefficients of the Runge–Kutta method is nonsingular and $|r| < 1$, where the *stability constant* r is given by $r = 1 - b^T \mathcal{A}^{-1} e$, and $e = (1, 1, \cdots, 1)^T$, sufficient conditions for a Runge–Kutta method to attain a given order are given in [11].

The global order results in [11] are very similar to those predicted by Prothero and Robinson [12] for the model problem

$$y'(t) = \lambda(y(t) - g(t)) + g'(t) \tag{1.3}$$

and by Frank, Schneid, and Ueberhuber [8] for nonlinear ODEs which satisfy a one-sided Lipschitz condition of the form

$$\langle f(t, y) - f(t, z), y - z\rangle \leq \gamma ||y - z||^2 \qquad \forall t \in \Re,\ \forall y, z \in \Re^m. \tag{1.4}$$

* Received by the editors February 10, 1988; accepted for publication (in revised form) May 16, 1989.

† Department of Computer Science, University of Auckland, Auckland, New Zealand.

‡ Computing and Mathematics Research Division, L-316, Lawrence Livermore National Laboratory, P. O. Box 808, Livermore, California 94550. The second author's work was partially supported by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U. S. Department of Energy by Lawrence Livermore National Laboratory under contract W-7405-Eng-48.

These stiff problems are very closely related to index 1 DAEs. However, the orders predicted for DAEs usually appear to be higher for DAEs than those predicted for stiff ODEs by the theory of Prothero and Robinson [12] and Frank et al. [8]. Burrage, Hundsdorfer, and Verwer [5], by considering the order of convergence of Runge–Kutta methods when applied to stiff semilinear systems of the form

$$y'(t) = Qy(t) + g(t, y(t)) \tag{1.5}$$

have shown that in many cases the global order associated with the B-convergence theory is one higher than the stage order. This analysis has been extended to nonlinear problems that satisfy (1.4) by Burrage and Hundsdorfer [4].

The aim of the present paper is first to extend the results of Petzold [11] to the case $|r| = 1$ and second to show that the DAE order results can be obtained from the theory in [5] by letting the magnitude of the stiffness go to infinity. In the process of extending the results of [11] to the case $|r| = 1$, we will also rederive the order results for $|r| < 1$ schemes from Petzold [11], correcting an error in the analysis that does not affect the conclusions of that paper.

Before proceeding to the main results, we note that our extended order results have a close relationship to several results that have recently appeared in the literature. In particular, Kværno [9] has derived a complete set of order conditions for the local truncation error for nonlinear index 1 systems, and a lower bound on the order of the global error. These results explain the higher than expected orders achieved by certain methods with $|r| < 1$ in numerical experiments reported in Petzold [11]. In this paper, we derive bounds on the global error for methods with $|r| = 1$ which explain the results of numerical experiments in Kværno [9] for these methods. In Ascher [1], a convergence result is outlined and order conditions are given for Gaussian collocation methods applied directly to fully implicit linear index 1 systems. The latter result is closely related to the $|r| = 1$ order results presented here, as Gaussian collocation methods are an important subset of the implicit Runge–Kutta methods that we consider. Our order results are in agreement with Ascher's for Gaussian collocation methods. Our analysis is in effect a generalization of the results in Petzold [11], while Ascher focuses on Gaussian collocation methods.

2. Extended order results for index 1 DAEs. In this section we extend the results of Petzold [11] to the case $|r| = 1$. Before proving the main result we will first define some terminology.

Following [11], we will say that a nonlinear DAE system (1.1) is *uniform index* 1 if the index of the local constant coefficient system $Aw'(t) + Bw(t) = g(t)$, where $A = F_{y'}(\hat{t}, \hat{y}, \hat{y}')$ and $B = F_y(\hat{t}, \hat{y}, \hat{y}')$ is 1 for all $(\hat{t}, \hat{y}, \hat{y}')$ in a neighborhood of the graph of the solution, and if the partial derivatives of A with respect to t, y, y' exist and are bounded, and the rank of A is constant in a neighborhood of the solution.

For uniform index 1 systems, there exist nonsingular matrices $P(t, y(t), y'(t))$, $Q(t, y(t), y'(t))$ which decouple the system [3], so that

$$PAQ = \begin{pmatrix} I_{m_1} & 0 \\ 0 & 0 \end{pmatrix},$$

$$PBQ = \begin{pmatrix} C(t, y(t), y'(t)) & 0 \\ 0 & I_{m_2} \end{pmatrix}$$

where $m = m_1 + m_2$. The matrices Q and C satisfy:

1. $Q(t, y(t), y'(t))$ and $Q^{-1}(t, y(t), y'(t))$ exist and are bounded for all $(t, y(t), y'(t))$ solving (1.1),
2. $Q^{-1}(t_1, y(t_1), y'(t_1))Q(t_2, y(t_2), y'(t_2)) = I_m + O(t_2 - t_1)$,
3. $C(t_1, y(t_1), y'(t_1)) = C(t_2, y(t_2), y'(t_2)) + O(t_2 - t_1)$.

Defining the algebraic conditions $C(q)$, $B(q)$, $A_1(q)$ by

$$C(q): \sum_{j=1}^{s} a_{ij} c_j^{k-1} = c_i^k/k, \qquad i = 1, \cdots, s, \quad k = 1, \cdots, q$$

$$B(q): \sum_{j=1}^{s} b_j c_j^{k-1} = 1/k, \qquad k = 1, \cdots, q$$

$$A_1(q): \; b^T \mathcal{A}^{-1} c^k = 1 \; , \quad k = 1, \cdots, q,$$

a Runge–Kutta method is said to have *internal stage order* k_I if and only if $C(k_I)$ and $B(k_I)$ hold. Furthermore, if $C(k_d)$ and $B(k_d + 1)$ hold, then the *differential order* (nonstiff ODE order) is at least $k_d + 1$. The definition of $A_1(q)$ corresponds to the order conditions for index 1 constant coefficient systems. That is, the *algebraic order* $k_{a,1}$ equals q if and only if $A_1(q)$. In general, for implicit Runge–Kutta methods with nonsingular $\mathcal{A}$ matrices, $B(q)$ and $C(q)$ implies $A_1(q)$.

We now state the main result.

THEOREM 1. *Suppose that* (1.1) *is uniform index* 1, *the Runge–Kutta method satisfies the stability condition* $|r| \leq 1$, *the errors in the initial conditions are* $O(h^G)$, *and the errors in terminating the Newton iterations are* $O(h^{G+\delta})$ *where* $\delta = 1$ *if* $|r| = 1$ *and* $\delta = 0$ *otherwise, and* $G \geq 2$. *Then the global errors satisfy* $||e_n|| = O(h^G)$ *where*

$$G = \begin{cases} q & \text{if } C(q) \text{ and } B(q); \\ q+1 & \text{if } C(q),\ B(q+1) \text{ and } -1 \leq r < 1; \\ q+1 & \text{if } C(q),\ B(q+1),\ A_1(q+1) \text{ and } r = 1. \end{cases}$$

Proof. Let $\delta_i^{(n)}$ be the ith internal local truncation error at the nth step of an s-stage Runge–Kutta method and let $\delta_{i+1}^{(n)}$ be the local error of the updating step.

As in Petzold [11] we introduce the following definitions

$$\begin{aligned} e_n &= y_n - y(t_n) \\ E_i &= Y_i - y(t_{n-1} + c_i h), \qquad i = 1, \cdots, s \\ E_i' &= Y_i' - y'(t_{n-1} + c_i h), \qquad i = 1, \cdots, s \\ \bar{e}_n &= Q^{-1} e_n \\ \bar{E}_i &= Q^{-1} E_i, \qquad i = 1, \cdots, s \\ \bar{E}_i' &= Q^{-1} E_i', \qquad i = 1, \cdots, s \\ \bar{\delta}_i^{(n)} &= Q^{-1} \delta_i^{(n)}, \qquad i = 1, \cdots, s \\ \bar{\eta}_i &= P_i \eta_i, \qquad i = 1, \cdots, s \end{aligned}$$

where η_i is the sum of the residuals from the Newton iteration in the ith stage and higher-order terms in e_{n-1} and E_i' and where matrices without subscripts or superscripts are evaluated at $(t_n, y(t_n), y'(t_n))$. The matrices P_i and Q_i represent P and Q evaluated at $(t_{n-1} + c_i h, y(t_{n-1} + c_i h), y'(t_{n-1} + c_i h))$. In addition, we now partition as in [11] any vector x, $x \in \Re^m$, into two parts of dimension m_1 (corresponding to the

differential part of the DAE) and m_2 (corresponding to the algebraic part), so that $x = (x^{(1)}, x^{(2)})$. We write

$$\begin{aligned}\bar{\delta}^{(n)} &= (\delta_1^{(1)}, \cdots, \delta_s^{(1)}, \delta_1^{(2)}, \cdots, \delta_s^{(2)})^T = (\bar{\delta}^{(1)}, \bar{\delta}^{(2)})^T \\ \bar{\eta}^{(n)} &= (\eta_1^{(1)}, \cdots, \eta_s^{(1)}, \eta_1^{(2)}, \cdots, \eta_s^{(2)})^T = (\bar{\eta}^{(1)}, \bar{\eta}^{(2)})^T\end{aligned}$$

and let

$$\begin{aligned}\hat{T}_1 &= I_s \otimes I_{m_1} + h\mathcal{A} \otimes C \\ \hat{S}_1 &= I_s \otimes C \\ \hat{T}_4 &= \mathcal{A} \otimes I_{m_2} \\ T_n^{-1} &= \begin{pmatrix} \hat{T}_1^{-1} + O(h) & O(h) \\ O(h) & \hat{T}_4^{-1}/h + O(1) \end{pmatrix} \\ K &= \begin{pmatrix} I_{m_1} & 0 \\ 0 & rI_{m_2} \end{pmatrix} \\ U_n &= \hat{U}_n + \begin{pmatrix} O(h) & O(h) \\ O(1) & O(1) \end{pmatrix} \\ b_1^T &= b^T \otimes I_{m_1} \\ b_2^T &= b^T \otimes I_{m_2} \\ \bar{B} &= \begin{pmatrix} b_1^T & 0 \\ 0 & b_2^T \end{pmatrix} \\ \hat{U}_n &= \begin{pmatrix} b_1^T \hat{T}_1^{-1} \hat{S}_1 & 0 \\ 0 & b_2^T \hat{T}_4^{-1}/h \end{pmatrix}\end{aligned}$$

where 0 represents a matrix of zeros with appropriate dimension to make the above matrices consistent.

Having introduced the notation of [11] we now quote the main result in [11, (3.17), p. 849], which relates the global error e_n to the initial error e_0. Rewriting this equation slightly and correcting an error in the original proof which does not change the conclusions, we obtain

$$\begin{aligned}\bar{e}_n = (K + O(h))^n \bar{e}_0 &+ \sum_{i=1}^{n-1} (K + O(h))^i (h\hat{U}_{n-i}\ \bar{\delta}^{(n-i)} + \bar{\delta}_{s+1}^{(n-i)}) \\ &+ \sum_{i=1}^{n-1} (K + O(h))^i h\bar{B}T_{n-i}^{-1}\ \bar{\eta}^{(n-i)} + \sum_{i=1}^{n-1} (K + O(h))^i h(U_{n-i} - \hat{U}_{n-i})\bar{\delta}^{(n-i)}.\end{aligned} \tag{2.1}$$

We note that $h\hat{U}_{n-i}\ \bar{\delta}^{(n-i)} + \bar{\delta}_{s+1}^{(n-i)}$ is $O(h^{\min(k_d+1, k_I+2)})$ in the differential part and $O(h^{k_{a,1}+1})$ in the algebraic part. Assuming that $Q(t, y(t), y'(t))$ is a continuous function of t and that the solution $y(t)$ to the DAE is sufficiently smooth, so that $\bar{\delta}^{(n-i)}$ and $\bar{\delta}_{s+1}^{(n-i)}$ are continuous, then $h\hat{U}_{n-i}\ \bar{\delta}^{(n-i)} + \bar{\delta}_{s+1}^{(n-i)}$ is also a continuous function of t. Similarly, we will assume $\bar{B}T_{n-i}^{-1}\bar{\eta}^{(n-i)}$ is a continuous function of t, which is true under mild assumptions on the smoothness of y if the Newton errors are neglected. Finally, we know that $||\bar{\delta}^{(n-i)}|| = O(h^{k_I+1})$.

It is shown in [6] that, for $i \leq n \leq 1/h$, if $|r| < 1$ then

$$(K + O(h))^i = \begin{pmatrix} O(1) & O(h) \\ O(h) & |r|^i O(1) + O(h) \end{pmatrix}.$$

This result is due to A. C. Hindmarsh and follows from diagonalization of a matrix which bounds the above matrix. The above result can also be shown to be true for $r = -1$. This follows from considering powers of the matrix $(K + O(h))^2$. For the case $r = 1$, we will use the result $||(K + O(h))^i|| = O(1)$, which follows simply from the fact that $||K + O(h)|| = 1 + O(h)$.

Partition (2.1) into differential and algebraic parts, let $\bar{\eta}_1$, $\bar{\eta}_2$ be the Newton errors, and suppose that the nonlinear higher-order terms satisfy $||\bar{\eta}^{(i)}|| \leq \Delta_i$ (this does not include the Newton errors) and the errors in the initial conditions satisfy $||\bar{e}_0^{(i)}|| = O(\zeta_i)$, $i = 1, 2$. Then we have from (2.1) the following cases:

(i) $-1 < r < 1$:

$$\begin{aligned}
\bar{e}_n^{(1)} &= O(\zeta_1) + O(h\zeta_2) + O(h^{k_d}) + O(h^{k_{a,1}+1}) + O(h^{k_I+1}) \\
&\quad + O(\Delta_1) + O(\Delta_2) + O(\bar{\eta}_1) + O(\bar{\eta}_2); \\
\bar{e}_n^{(2)} &= O(h\zeta_1) + O(h\zeta_2) + O(h^{k_d+1}) + O(h^{k_{a,1}+1}) + O(h^{k_I+2}) \\
&\quad + O(h\Delta_1) + O(\Delta_2) + O(h\bar{\eta}_1) + O(\bar{\eta}_2).
\end{aligned}$$

Hence

$$e_n = O(\zeta) + O(\bar{\eta}) + O(\Delta) + O(h^G)$$

where

$$G = \min(k_d, k_I + 1).$$

(ii) $r = -1$:

$$\begin{aligned}
\bar{e}_n^{(1)} &= O(\zeta_1) + O(\zeta_2) + O(h^{k_d}) + O(h^{k_{a,1}+1}) + O(h^{k_I+1}) \\
&\quad + O(\Delta_1) + O(\Delta_2) + O(\bar{\eta}_1) + O(h\bar{\eta}_2); \\
\bar{e}_n^{(2)} &= O(\zeta_1) + O(\zeta_2) + O(h^{k_d+1}) + O(h^{k_{a,1}+1}) + O(h^{k_I+2}) \\
&\quad + O(\Delta_1) + O(\Delta_2) + O(\bar{\eta}_1) + O(\bar{\eta}_2/h).
\end{aligned}$$

Hence

$$e_n = O(\zeta) + O(\Delta) + O(\bar{\eta}/h) + O(h^G)$$

where

$$G = \min(k_d, k_I + 1).$$

The better than expected results for $\bar{e}_n^{(2)}$ in this case are due to cancellations in the algebraic part in the sums of (2.1) which come about because of alternating signs in the bottom right block of K when $r = -1$, coupled with the assumptions about smoothness. This can be easily seen by grouping the terms in the sums of (2.1) together two at a time, and then bounding the resulting sums. We also note that when $|r| = 1$ the Newton iterations must be solved to $O(h^{k_G+1})$.

(iii) r = 1 :

$$\begin{aligned}
\bar{e}_n^{(1)} &= O(\zeta_1) + O(\zeta_2) + O(h^{k_d}) + O(h^{k_{a,1}}) + O(h^{k_I+1}) \\
&\quad + O(\Delta_1) + O(\Delta_2) + O(\bar{\eta}_1) + O(h\bar{\eta}_2); \\
\bar{e}_n^{(2)} &= O(\zeta_1) + O(\zeta_2) + O(h^{k_{a,1}}) + O(h^{k_I+1}) + O(h^{k_d+1}) \\
&\quad + O(\Delta_1) + O(\Delta_2) + O(\bar{\eta}_1) + O(\bar{\eta}_2/h).
\end{aligned}$$

Hence

$$e_n = O(\zeta) + O(\bar{\eta}/h) + O(\Delta) + O(h^G)$$

where

$$G = \begin{cases} \min(k_d, k_I) & \text{if } k_{a,1} = k_I; \\ \min(k_d, k_I + 1) & \text{if } k_{a,1} \geq k_I + 1. \end{cases}$$

The better than expected terms $O(\Delta_1)$, $O(\Delta_2)$ in $\bar{e}_n^{(2)}$ come about because $b^T \mathcal{A}^{-1} e = 1 - r = 0$ in this case, so that the smoothness of $\bar{\eta}$ implies that the $O(\bar{\eta}^{(2)}/h)$ term in $\bar{B} T_{n-i}^{-1} \bar{\eta}^{(n-i)}$ vanishes. Thus for linear problems ($\Delta = 0$) we obtain the results in the statement of the theorem by noting that

$$C(q), B(q) \Rightarrow k_I \geq q$$
$$C(q), B(q+1) \Rightarrow k_d \geq q + 1.$$

For nonlinear systems the proof is analogous to that given in [11]. Details are given in [3]. □

It should be noted that implicit Runge–Kutta methods with $|r| = 1$ are in some sense unstable. März [10] observes that if the implicit midpoint method is used to solve the algebraic equation $y = 0$, and a perturbation of size $(-1)^n \delta$ is made in each step, we obtain $y_{n+1} = -y_n + (-1)^n \delta$. Thus $|y_{n+1}| = (n+1)|\delta|$. There is a linear instability in these methods. This is why, in our analysis, the Newton iteration must be solved more precisely for $|r| = 1$ than for $|r| < 1$. We think that for most machines and for most problems this difficulty with roundoff errors is not so severe that it would prevent us from considering these methods.

A more critical stability consideration, for $|r| = 1$ schemes applied to fully implicit index 1 DAEs, has been pointed out by Ascher [2]. The stability of the scheme is controlled by the stability of an underlying ODE problem which is not necessarily stable when the original DAE system is stable. The asymptotic order results are technically correct, but the constant can be very large. The problem is most severe when the differential and algebraic parts of the system are tightly coupled together. For BVPs the instability can sometimes be corrected by locating some of the consistency conditions corresponding to algebraic constraints at the correct boundary [2]. However, for IVPs this is not practical, and the use of $|r| < 1$ schemes is advised. The asymptotic order results for $|r| = 1$ schemes applied to DAE IVPs are useful because they imply the order results for DAE BVPs [7].

We present several numerical experiments on implicit Runge–Kutta methods with $|r| = 1$. The observed results are in agreement with the order predictions of Theorem 1. It should be noted that the initial values for "algebraic" components must be given quite accurately for these methods because, in contrast to L-stable methods such as BDF, these initial errors are not damped out by the method. The experiments below were performed in single precision on a Cray-1 computer with exact initial conditions.

The test problem is the same as in Petzold [11], and is constructed to illustrate the effects of coupling between the differential and algebraic parts of the system. The problem is given by

$$\begin{pmatrix} 1 & -t \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \tilde{Y}_1' \\ \tilde{Y}_2' \end{pmatrix} + \begin{pmatrix} 1 & -(1+t) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \sin(t) \end{pmatrix},$$

with initial values given by

$$\begin{pmatrix} \tilde{Y}_1(0) \\ \tilde{Y}_2(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tilde{Y}_1'(0) \\ \tilde{Y}_2'(0) \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

This problem is index 1 and has the true solution

$$\begin{pmatrix} \tilde{Y}_1(t) \\ \tilde{Y}_2(t) \end{pmatrix} = \begin{pmatrix} \exp(-t) + t\sin(t) \\ \sin(t) \end{pmatrix}.$$

We present the results of numerical experiments on two implicit Runge–Kutta methods with $|r| = 1$. The first is the implicit midpoint method. This method has $r = -1$, $k_I = 1$, $k_d = 2$, $k_{a,1} = 1$. We solved the problem with a decreasing sequence of fixed stepsizes over the interval [0,1]. The order of global error that we observed at the end of the interval was 2, which is in agreement with the lower bound given in Theorem 1.

The second numerical method is the 2-stage A-stable Gauss–Legendre method given by

$$\begin{array}{c|cc} \frac{3-\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\ \frac{3+\sqrt{3}}{6} & \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}.$$

This method has $r = 1$, $k_I = 2$, $k_{a,1} = 2$, $k_d = 4$. The order of the global error that we observed at the end of the interval was 2. This result is in agreement with the lower bound given in Theorem 1.

3. Unified order results for DAEs and stiff IVPs. In this section we will discuss how the results on DAEs can be obtained from the study of the global error recursion schemes given in [5] for general ODE initial value problems by letting the stiffness become infinite. Before doing this it is interesting to note that the algebraic order condition $A_1(q)$ can also be derived by considering the test equation (1.3). If we apply the Runge–Kutta method given in (1.2) to this problem we find after some simplification that

$$y_n = y_{n-1}(1 + h\lambda b^T(I - h\lambda\mathcal{A})^{-1}e) + hb^T(I - h\lambda\mathcal{A})^{-1}(g' - \lambda g)$$

where

$$\begin{aligned} g' &= (g'(t_{n-1} + c_1h), \cdots, g'(t_{n-1} + c_sh))^T \\ g &= (g(t_{n-1} + c_1h), \cdots, g(t_{n-1} + c_sh))^T. \end{aligned}$$

Hence the local error is given by

$$g(t_n) - g(t_{n-1})(1 + h\lambda b^T(I - h\lambda\mathcal{A})^{-1}e) - hb^T(I - h\lambda\mathcal{A})^{-1}(g' - \lambda g)$$

and by letting the stiffness $(\lambda) \to -\infty$ we find by a Taylor series expansion of $g(t_{n-1} + h)$ that the local error is $O(h^{q+1})$ if

$$b^T\mathcal{A}^{-1}c^k = 1, \qquad k = 1, \cdots, q,$$

which is precisely the order condition $A_1(q)$ for linear constant-coefficient index 1 DAEs.

Burrage, Hundsdorfer, and Verwer [5] derived recursion schemes for the global error when Runge–Kutta methods are applied to ODEs. Although they considered the multidimensional case, we will quote their results only for the scalar case.

Defining

$$Z_i^{(n)} = h\int_0^1 f'(t_n + c_i h, Y_i + \theta(Y_i - y(t_n + c_i h)))d\theta, \qquad i = 1, \cdots, s$$

where $f'(t,y)$ is the Jacobian matrix $\partial f(t,y)/\partial y$ in the multidimensional case, Burrage et al. [5] obtained the following recursion scheme for the global error

$$e_{n+1} = (I + b^T Z_n (I - \mathcal{A}Z_n)^{-1} e)\, e_n + b^T Z_n (I - \mathcal{A}Z_n)^{-1}\delta_n + \delta_{s+1} \tag{3.1}$$

where δ_n is the vector of local errors in the internal stages and δ_{s+1} is the local residue error in the final updating stage and $Z^{(n)} = \mathrm{diag}(Z_1^{(n)}, \cdots, Z_s^{(n)})$.

Assuming that the method has stage order q so that $B(q)$ and $C(q)$ hold and defining $k = (k_1, k_2, \cdots, k_s)^T \in \Re^s$ by

$$k = \frac{1}{q!}\left(\frac{1}{q+1}c^{q+1} - \mathcal{A}c^q\right)$$

we have

$$\delta_i^{(n)} = h^{q+1}k_i y^{(q+1)}(t_n) + O(h^{q+2}), \qquad i = 1, \cdots, s$$
$$\delta_{s+1} = h^{q+1}\frac{1}{q!}\left(\frac{1}{q+1} - b^T c^q\right) y^{(q+1)}(t_n) + O(h^{q+2}).$$

By introducing the rational function ψ given by

$$\psi(z) = [b^T(I - \mathcal{A}z)^{-1}e]^{-1}[b^T(I - \mathcal{A}z)^{-1}k], \qquad z \in \mathbb{C} \tag{3.2}$$

and defining

$$v_n = \psi(z)h^{q+1}y^{(q+1)}(t_n),$$

Burrage et al. [5] obtained a perturbed recursion scheme for the global error for semilinear problems of the form (1.5),

$$\hat{e}_{n+1} = (I + b^T Z_n (I - \mathcal{A}Z_n)^{-1} e)\, \hat{e}_n + b^T Z_n (I - \mathcal{A}Z_n)^{-1}\hat{\delta}_n + \hat{\delta}_{s+1} \tag{3.3}$$

where

$$\hat{e}_n = e_n + v_n$$
$$\hat{\delta}_n = \delta_n - kh^{q+1}y^{(q+1)}(t_n)$$
$$\hat{\delta}_{s+1} = \delta_{s+1} + v_{n+1} - v_n.$$

Using these recursion schemes Burrage et al. [5] obtained optimal B-convergence results of the form

$$||e_n|| \le Ch^G, \qquad 0 < h \le \bar{h}$$

where $\bar{h}$ and C are constants independent of the stiffness of the problem. For the semilinear problem (1.5) and assuming the Runge–Kutta method is A-stable and $\mathcal{A}$ is nonsingular (although this latter condition is not necessary) they showed

$$G = \begin{cases} q & \text{if } B(q) \text{ and } C(q) \\ q+1 & \text{if } B(q+1)\ ,\ C(q) \text{ and } \psi \text{ is uniformly bounded on } \mathbb{C}^-. \end{cases} \tag{3.4}$$

By letting the stiffness $\to -\infty$ uniformly we can use the recursion scheme (3.3) to obtain

$$\hat{e}_{n+1} = r\hat{e}_n - b^T\mathcal{A}^{-1}\hat{\delta}_n + \hat{\delta}_{s+1} \tag{3.5}$$

where

$$\begin{aligned} r &= 1 - b^T\mathcal{A}^{-1}e \\ \psi &= (b^T\mathcal{A}^{-1}e)^{-1}(b^T\mathcal{A}^{-1}k). \end{aligned}$$

Noting that if $r = 1$ (so that $b^T\mathcal{A}^{-1}e = 0$) the uniform boundedness of ψ requires

$$\begin{aligned} & b^T\mathcal{A}^{-1}k = 0 \\ \Leftrightarrow\ & b^T\mathcal{A}^{-1}c^{q+1} = (q+1)b^Tc^q = 1 \qquad (\text{if } B(q+1)) \\ \Leftrightarrow\ & A_1(q+1) \qquad (\text{if } B(q+1)) \end{aligned}$$

and we obtain from (3.4) the results of Theorem 1.

For the general nonlinear dissipative problem satisfying (1.4) with $\gamma = 0$, Frank et al. [8] and Burrage and Hundsdorfer [4] showed that if a method is algebraically stable and there exists a diagonal matrix $D > 0$ such that

$$D\mathcal{A} + \mathcal{A}^TD > 0$$

then a Runge–Kutta method is optimally B-convergent with order G, where

$$G = \begin{cases} q & \text{if } B(q) \text{ and } C(q) \\ q+1 & \text{if } B(q+1)\ ,\ C(q)\ ,\ k = te(t \in \Re),\ c_i - c_j \text{ not an integer } (i \neq j). \end{cases}$$

Some methods achieve a higher order of optimal B-convergence for the semilinear problem (1.5) than for the general nonlinear dissipative problem satisfying (1.4). Thus the order of optimal B-convergence for general nonlinear dissipative problems may be one lower in some cases than the order for index 1 DAEs.

We conclude this paper with the remark that if in a DAE system the "differential" part is stiff then the $O(h)$ analysis of Petzold [11] is no longer appropriate since the error term depends on the stiffness of the problem. In this case we must use the stability of the method (A-stability for stiff linear problems and nonlinear stability for dissipative problems) to bound the global errors and in this case k_d is replaced by the appropriate B-convergence order as discussed in [4], [5], and [8].

Acknowledgments. The authors would like to thank E. Griepentrog for pointing out an error in the proof of Theorem (3.2) in Petzold [11] which did not affect the conclusions of that paper, and A. C. Hindmarsh for his assistance in correcting the error. We would also like to thank J. Coyle for pointing out an error in an earlier version of this paper.

REFERENCES

[1] U. ASCHER, *On numerical differential algebraic problems with application to semiconductor device simulation*, SIAM J. Numer. Anal., 26(1989), pp. 517–538.

[2] ———, *Symmetric schemes and differential algebraic equations*, SIAM J. Sci. Statist. Comput., 10(1989), pp. 937–949.

[3] K. E. BRENAN, S. L. CAMPBELL, AND L. R. PETZOLD, *The Numerical Solution of Initial Value Problems in Differential-Algebraic Equations*, Elsevier Science Publishing Co., New York, 1989.

[4] K. BURRAGE AND W. H. HUNDSDORFER, *The order of B-convergence of algebraically stable Runge–Kutta methods,* BIT, 27(1987), pp. 62–71.

[5] K. BURRAGE, W. H. HUNDSDORFER, AND J. A. VERWER, *A study of B-convergence of Runge–Kutta methods,* Computing, 36(1986), pp. 17–34.

[6] K. BURRAGE AND L. R. PETZOLD, *On order reduction for Runge–Kutta methods applied to differential/algebraic systems and to stiff systems of ODEs*, UCRL98046, Lawrence Livermore National Laboratory, Livermore, CA, 1988.

[7] K. CLARK AND L. PETZOLD, *Numerical solution of boundary value problems in differential/algebraic systems*, SIAM J. Sci. Statist. Comput., 10(1989), pp. 915–936.

[8] R. FRANK, J. SCHNEID, AND C. W. UEBERHUBER, *Order results for implicit Runge–Kutta methods applied to stiff systems,* SIAM J. Numer. Anal., 22(1985), pp. 515–534.

[9] A. KVÆRNO, *Order conditions for Runge–Kutta methods applied to differential-algebraic systems of index one*, Report 4/87, Division of Numerical Mathematics, Norwegian Institute of Technology, 1987.

[10] R. MÄRZ, *On difference and shooting methods for boundary value problems in differential/algebraic equations*, Humboldt-Universität zu Berlin, Preprint 24, 1982.

[11] L. PETZOLD, *Order results for implicit Runge–Kutta methods applied to differential/algebraic systems,* SIAM J. Numer. Anal., 23(1986), pp. 837–852.

[12] A. PROTHERO AND A. ROBINSON, *On the stability and convergence of one-step methods for solving stiff systems of ordinary differential equations,* Math. Comp., 28(1974), pp. 145–162.