Research Contribution

Third-Order Semi-Implicit Runge-Kutta Methods for Time-Dependent Index-One Differential-Algebraic Equations

TOSHIYUKI KOTO*

We study the order of accuracy of implicit Runge-Kutta (IRK) methods applied to time-dependent differential algebraic equations (DAEs) of index 1. A sufficient condition is derived for an IRK method to attain third-order accuracy for DAEs. Furthermore, a third-order semi-implicit RK method is constructed on the basis of the condition. Numerical experiments confirm the results.

1. Introduction

We discuss the accuracy of implicit Runge-Kutta (IRK) methods applied to the system of differential-algebraic equations (DAEs) of index 1

$$0=F(t, u(t), u'(t)), \quad t\in[a, b],$$

where F and u are m-dimentional vectors. In particular, we are interested in the order of semi-implicit RK methods for a DAE system of the form

$$A(t)u'(t)+B(t)u(t)=g(t), \tag{1.1.a}$$

where $A(t)$ and $B(t)$ satisfy

$$P(t)A(t)Q(t)=\begin{pmatrix} I_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad P(t)B(t)Q(t)=\begin{pmatrix} C(t) & 0 \\ 0 & I_2 \end{pmatrix} \tag{1.1.b}$$

for some non-singular matrices $P(t)$ and $Q(t)$, and I_1 and I_2 are identity matrices of order m_1 and m_2 ($m=m_1+m_2$), respectively.

In order to integrate DAE systems numerically, a certain stability condition is required for IRK methods. Furthermore, it is often observed that even a stable IRK method does not attain the same order for DAE systems as for purely differential systems [e.g. 2, 7]. Such reduction of order has been analyzed for an system of the form

$$y'(t)=f(y(t), z(t)), \quad 0=g(y(t), z(t)),$$
$$(\partial g/\partial z)^{-1} \text{ exists and is bounded.} \tag{1.2}$$

A complete characterization of the order of IRK methods has been obtained by Roche [9]. However, the system (1.2) can be transformed into a purely differential system by the Implicit Function Theorem. Thus it is also possible to integrate it with no reduction of order. It is therefore important to study the system (1.1), in which it is difficult to separate the differential part and the algebraic part of the system.

Regarding the DAE system (1.1), Petzold [8] gives a sufficient condition for an IRK method to attain a certain order. However, as far as a semi-implicit RK method is concerned, the condition guarantees only second-order accuracy at most. Although many semi-implicit RK methods indeed show second-order accuracy for (1.1), some numerical experiments suggest that third-order semi-implicit methods exist.

Our main purpose in this paper is to clarify the existence of such semi-implicit methods. After a description of preliminaries in Section 2, we present in Section 3 a theorem that gives a sufficient condition for an IRK method to attain third-order accuracy for (1.1). On the basis of the condition we construct a third-order semi-implicit RK method, for which experimental verification is shown in the same section. The proof of the theorem is given in Section 4. In the final section, we make some comments on the results.

2. Preliminaries

Let (A, B) be a regular pencil of $m\times m$ matrices, that is, let A and B be $m\times m$ matrices such that $\det(A+\lambda B)$ is not identically zero. For the pair (A, B) there exist non-singular matrices P and Q such that

$$PAQ=\begin{pmatrix} I_1 & 0 \\ 0 & E \end{pmatrix}, \quad PBQ=\begin{pmatrix} C & 0 \\ 0 & I_2 \end{pmatrix}$$

[4, 10]. Here E is a nilpotent matrix of index k, that is, $E^k=0$ and $E^{k-1}\neq 0$ (If $E=0$, we consider $k=1$). The integer k is called the index of the pencil (A, B).

The system (1.1.a) is said to be (global) index 1 if the

*International Institute for Advanced Study of Social Information Science, Fujitsu Limited. Currently, University of Electro-Communications, Tokyo, Japan.

pencil $(A(t), B(t))$ has index 1 for every $t\in[a, b]$. This condition is equivalent to the existence of $P(t)$ and $Q(t)$ such that (1.1.b) holds. In this paper, we assume that $A(t)$, $B(t)$, $g(t)$, $P(t)$, $Q(t)$, and $Q(t)^{-1}$ are sufficiently smooth functions.

We now introduce the concept of solvability of (1.1.a) following Gear and Petzold [5]: The system (1.1.a) is said to be *solvable* if for any $g(t)$, there exist solutions to the DAEs defined on $[a, b]$, and solutions that have the same initial value are identical.

Let $v(t)=Q(t)^{-1}u(t)$. Using (1.1.b), the system (1.1.a) is rewritten as

$$\begin{pmatrix} I_1 & 0 \\ 0 & 0 \end{pmatrix} v'(t)+\begin{pmatrix} C(t) & 0 \\ 0 & I_2 \end{pmatrix} v(t) + \begin{pmatrix} I_1 & 0 \\ 0 & 0 \end{pmatrix} Q(t)^{-1}Q'(t)v(t)=P(t)g(t). \quad (2.1)$$

This system has a solution $v(t)$ determined by an initial value v_0 constrained by $v_{0(k)}=(P(a)g(a))_{(k)}$ for $k=m_1+1$, $m_1+2, \ldots, m$. Here $v_{0(k)}$ represents the kth component of the vector v_0. Therefore the system (1.1) is solvable, and a solution $u(t)$ is determined by an initial value u_0 satisfying $(Q(a)^{-1}u_0)_{(k)}=(P(a)g(a))_{(k)}$ for $k=m_1+1$, $m_1+2, \ldots, m$. By the assumption that $A(t)$ and $B(t)$ are sufficiently smooth, the solution $u(t)$ is also sufficiently smooth.

Let $h(>0)$ be a step size and let

$$a=t_0<t_1<\cdots<t_n<\cdots<t_N=b,\ t_n=t_0+nh$$

be a partition of $[a, b]$. Then an s-stage IRK method for (1.1) is formulated as follows [8]:

$$\begin{aligned} &A(t_{n,i})U_i'+B(t_{n,i})U_i=g(t_{n,i}),\\ &U_i=u_n+h\sum_{j=1}^{s}a_{ij}U_j',\ t_{n,i}=t_n+c_ih,\ i=1, 2, \ldots, s,\\ &u_{n+1}=u_n+h\sum_{i=1}^{s}b_iU_i', \end{aligned} \quad (2.2)$$

where u_0 is given by the initial value $u(a)$ of a true solution $u(t)$ of (1.1), and a_{ij}, b_i, and c_i are the parameters of the IRK method. As usual, it is assumed that $c_i=\sum_{j=1}^{s}a_{ij}$.

For notational convenience, we define the matrix $\mathbf{A}$ and the vectors $\mathbf{b}$ and $\mathbf{c}$ by

$$\mathbf{A}=(a_{ij})(1\leqq i, j\leqq s),\quad \mathbf{b}=(b_1, b_2, \ldots, b_s)^T,$$
$$\mathbf{c}=(c_1, c_2, \ldots, c_s)^T,$$

respectively. When the matrix $\mathbf{A}$ is lower triangular, the method is said to be semi-implicit.

Let $V_i'=Q(t_{n,i})^{-1}U_i$. Then the equations of (2.2) yield

$$\begin{pmatrix} I_1 & 0 \\ 0 & 0 \end{pmatrix} V_i'+\begin{pmatrix} C(t_{n,i}) & 0 \\ 0 & I_2 \end{pmatrix} Q(t_{n,i})^{-1}u_n + h\sum_{j=1}^{s}a_{ij}\begin{pmatrix} C(t_{n,i}) & 0 \\ 0 & I_2 \end{pmatrix} Q(t_{n,i})^{-1}Q(t_{n,j})V_j' =P(t_{n,i})g(t_{n,i}),\quad i=1, 2, \ldots, s. \quad (2.3)$$

Since $Q(t_{n,i})^{-1}Q(t_{n,j})=I+O(h)$, V_i''s (or U_i''s) are determined from u_n if the matrix $\mathbf{A}$ is non-singular. On the other hand, when $Q(t)$ is a constant matrix, $\mathbf{A}$ must be non-singular for those values to be obtained. Thus, we can compute u_{n+1} from u_n by (2.2) for any system of the form (1.2) if and only if $\mathbf{A}$ is non-singular. Hereafter $\mathbf{A}$ will be assumed to be non-singular.

An order of the IRK method for the DAE system (1.1) is defined as follows:

Definition. The *differential-algebraic order* of the IRK method is at least r if

$$u(t)-u_n=O(h^r), \text{ for any fixed } t(=t_n)\in[a, b]. \quad (2.4)$$

In order to characterize the differential-algebraic order of IRK methods, other preliminaries are required. First, we define the *stage order* of an IRK method. For an IRK method, let $q(i)$, $i=1, 2, \ldots, s$, be integers such that

$$\left| w(t_{n+1})-w(t_n)-\sum_{j=1}^{s}a_{ij}w'(t_{n,j})\right| =O(h^{q(i)+1}) \quad (2.5)$$

for any sufficiently smooth function $w(t)$. From the Taylor expansion of $w(t)$, (2.5) yields

$$\sum_{j=1}^{s}a_{ij}c_j^{k-1}=c_i^k/k,\quad k=1, 2, \ldots, q(i). \quad (2.6)$$

The stage order of the IRK method is defined by $q=\min\{q(i), 1\leqq i\leqq s\}$.

Let $R(z)$ denote the stability function of the IRK method, that is,

$$R(z)=1+z\mathbf{b}^T(I-z\mathbf{A})^{-1}\mathbf{e},\quad \mathbf{e}=(1, 1, \ldots, 1)^T. \quad (2.7)$$

It follows from (2.7) that

$$\gamma\equiv\lim_{z\to\infty}R(z)=1-\mathbf{b}^T\mathbf{A}^{-1}\mathbf{e}. \quad (2.8)$$

Using the above notation, Petzold [8] (see also Brenan et al. [2]) shows that if $|\gamma|<1$, then there exists an integer r such that (2.4) holds, and r is greater than or equal to

$$\min\{p, q+1\}, \quad (2.9)$$

where p is the order of the IRK method applied to purely differential systems.

For any semi-implicit method, the stage order q is equal to 1, because $q(1)$ is equal to 1. Thus (2.9) does not exceed 2 for any semi-implicit method. However, it should be noted that some semi-implicit methods show third-order accuracy for (1.1) in numerical experiments.

3. Statement of Results

In this section, we present a sufficient condition for the differential-algebraic order of an IRK method to be greater than or equal to 3. We also construct a third-order semi-implicit RK method.

Theorem. *Assume that an IRK method satisfies*

(i) $p=3$, (ii) $\mathbf{b}^T\mathbf{A}^{-1}\mathbf{c}^2=1$ (iii) $(\mathbf{bc})^T\mathbf{A}^{-1}\mathbf{c}^2=2/3$ *and* (iv) $|\gamma|<1$,

where $\mathbf{c}^2=(c_1^2, c_2^2, \ldots, c_s^2)^T$ *and* $\mathbf{bc}=(b_1c_1, b_2c_2, \ldots, b_sc_s)^T$. *Then, the differential algebraic order of the method is greater than or equal to 3.*

Remark 1. Let $\delta=\mathbf{Ac}-\mathbf{c}^2/2$. Since condition (i) implies that $\mathbf{b}^T\mathbf{c}=1/2$ and $\mathbf{b}^T\mathbf{c}^2=1/3$, conditions (i), (ii), and (iii) are equivalent to the conditions

(i) $p=3$, (ii)$'$ $\mathbf{b}^T\mathbf{A}^{-1}\delta=0$, (iii)$'$ $(\mathbf{bc})^T\mathbf{A}^{-1}\delta=0$.

Thus if $q=2$ for an IRK method, then conditions (i), (ii), and (iii) are automatically satisfied (cf. (2.6)). This agrees with Petzold's results.

Remark 2. In the case of the DAE system (1.2), conditions (i), (ii), and (iv) form a sufficient condition for an IRK method to be convergent of order 3 [9]. Thus, an IRK method satisfying the conditions of the theorem is third-order for (1.2). On the other hand, numerical experiments show that condition (iii) cannot be omitted to attain third-order for the system (1.1).

When $s=2$, although several IRK methods (for example, Radau IIA) satisfy the conditions of the theorem, there is no such semi-implicit method. This can be easily verified by computation. We here consider 3-stage diagonally implicit Runge-Kutta (DIRK) methods, represented by the array

$$\begin{array}{c|ccc} c_1 & \alpha & & \\ c_2 & a_{21} & \alpha & \\ c_3 & a_{31} & a_{32} & \alpha \\ \hline & b_1 & b_2 & b_3 \end{array}. \tag{3.1}$$

The condition (i) implies that

$$b_1+b_2+b_3=1, \quad b_1c_1+b_2c_2+b_3c_3=1/2,$$
$$b_1c_1^2+b_2c_2^2+b_3c_3^2=1/3,$$
$$b_1\alpha c_1+b_2(a_{21}c_1+\alpha c_2)+b_3(a_{31}c_1+a_{32}c_2+\alpha c_3)=1/6. \tag{3.2}$$

From (3.2), on the assumption that $(c_2-\alpha)(c_3-\alpha)(c_2-c_3)\neq 0$, b_1, b_2, b_3, a_{21}, a_{32} and a_{31} are represented by $\alpha(=c_1)$, c_2 and c_3, that is,

$$b_1=\frac{1/3-(c_2+c_3)/2+c_2c_3}{(\alpha-c_2)(\alpha-c_3)}, \quad b_2=\frac{1/3-(c_3+\alpha)/2+c_3\alpha}{(c_2-c_3)(c_2-\alpha)},$$
$$b_3=\frac{1/3-(\alpha+c_2)/2+\alpha c_2}{(c_3-\alpha)(c_3-c_2)},$$
$$a_{21}=c_2-\alpha, \quad a_{32}=\frac{\alpha^2-\alpha+1/6}{b_3(c_2-\alpha)}, \quad a_{31}=c_3-\alpha-a_{32}. \tag{3.3}$$

Using these relations, conditions (ii) and (iii) can be reduced to

$$(\alpha^2-\alpha+1/6)c_2=(\alpha^3-3\alpha^2/2+\alpha/3) \tag{3.4a}$$

and

Table 1 Parameters of DIDA3.

$a=$	4.358665215084590D−1
$a_{21}=$	2.820667392457705D−1
$a_{31}=$	4.838154663299224D−2
$a_{32}=$	7.988541035008544D−2
$b_1=$	2.689623426019636D+0
$b_2=$	1.826116589129458D+0
$b_3=$	−3.515740015114909D+0
$c_1=$	4.358665215084590D−1
$c_2=$	7.179332607542295D−1
$c_3=$	5.641334784915367D−1

$$(\alpha-1/3)c_2c_3-(\alpha^2-2\alpha/3)(c_2+c_3)+(\alpha^3-4\alpha^2/3)=0, \tag{3.4b}$$

respectively.

Method (3.1) with $p=3$ is L-stable if and only if α is equal to the reciprocal of the second zero of the 3-degree Laguerre polynomial (see Butcher [3], p. 248). Noticing that an L-stable method satisfies $\gamma=0$, we choose α as the value, that is, $\alpha=0.435866\ldots$. Then (3.4) possesses the unique solution $c_2=(1+\alpha)/2$, $c_3=1-\alpha$, which, together with (3.3), determines a DIRK method satisfying the conditions of the theorem. In Table 1 we present the method, called "DIDA3".

Similarly, letting $c_2=(1+\alpha)/2$ and $c_3=1$ for the same α, we obtain a method satisfying every condition of the theorem except (iii). This is known as Alexander's method, and is characterized by a special stability property (see Alexander [1], Theorem 5, p. 1012). Alexander's DIRK method shows numerically second-order accuracy for (1.1).

Now, we consider the two-dimensional problem

$$A(t)u'(t)+B(t)u(t)=g(t)(t>0), \quad u(0)=(1, 1/2)^T,$$
$$A(t)=\begin{pmatrix}1 & -t\\ 0 & 0\end{pmatrix}, \quad B(t)=\begin{pmatrix}1 & -(1+t)\\ -1/2 & 1+t/2\end{pmatrix},$$
$$g(t)=\begin{pmatrix}0\\ \sin(t)\end{pmatrix}. \tag{3.5}$$

The matrices $A(t)$ and $B(t)$ satisfy

$$P(t)A(t)Q(t)=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \quad P(t)B(t)Q(t)=\begin{pmatrix}1/2 & 0\\ 0 & 1\end{pmatrix},$$

where

$$P(t)=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \quad Q(t)=\begin{pmatrix}1+t/2 & t\\ 1/2 & 1\end{pmatrix}.$$

Therefore (3.5) is an example of DAEs of the form (1.1), which has the true solution

$$u_{(1)}(t)=(1+t/2)e^{-t}+t\sin(t), \quad u_{(2)}(t)=e^{-t}/2+\sin(t). \tag{3.6}$$

To find the accuracy of DIDA3 and Alexander's

Table 2 Numerical Results for Problem (3.5).

method\N	4	8	16	32	64	128	256	512	r_{eff}
DIDA3	3.32	4.24	5.16	6.07	6.97	7.88	8.79	9.70	3.02
Alexander	2.16	2.79	3.40	4.01	4.62	5.22	5.82	6.42	2.02

method, we integrated problem (3.5) from $t=0$ to $t=1$ with the step size $h=1/N$ (for various values of N). More specifically, we measured the accuracy by the number of correct significant digits of the first component of the numerical solution at $t=1$, that is, the value of

$$-\log_{10}|u_{(1)}(1)-u_{(1)N}|. \tag{3.7}$$

For each method, assuming that those values (obtained for various values of h) fall on a straight line for $-\log_{10}(h)$, we calculated the slope by the method of least squares. The slope r_{eff} is considered as an effective differential-algebraic order of the corresponding method.

The results are displayed in Table 2. All the numerical experiments were conducted by double-precision arithmetic on an FM11 AD2+ computer.

4. Proof of Theorem

In this section, we will prove the theorem. For simplicity, we describe the proof in the case $m_1=m_2=1$, which is essentially the same as in the general case. We also assume that h is sufficiently small.

Since (2.2) is regarded as an application of the IRK method to

$$P(t)A(t)u'(t)+P(t)B(t)u(t)=P(t)g(t),$$

we can assume that $P(t)$ is the identity matrix without loss of generality.

For a true solution $u(t)$ of (1.1), we define ϕ_n, $n=0, 1, \ldots,$ by

$$\phi_0=u(a), \quad \Phi_i'=u'(t_{n,i}),$$

$$\Phi_i=\phi_n+h\sum_{j=1}^{s}a_{ij}\Phi_j', \quad \phi_{n+1}=\phi_n+h\sum_{i=1}^{s}b_i\Phi_i'. \tag{4.1}$$

Since ϕ_n's are regarded as numerical solutions of the differential equation

$$\phi'(t)=u'(t), \quad \phi(a)=u(a),$$

from condition (i), we obtain

$$u(t_n)-\phi_n=O(h^3). \tag{4.2}$$

Therefore, in order to prove the theorem, it suffices to show that

$$\phi_n-u_n=O(h^3). \tag{4.3}$$

The estimate (4.2), together with the definition of Φ_i, yields

$$u(t_{n,i})=\Phi_i-h^2\delta_{(i)}u''(t_n)+O(h^3), \tag{4.4}$$

where $\delta_{(i)}$ represents the ith component of the vector $\delta=\mathbf{A}\mathbf{c}-\mathbf{c}^2/2$, that is,

$$\delta_{(i)}=\sum_{j=1}^{s}a_{ij}c_j-c_i^2/2.$$

On the other hand, since $u(t)$ is a true solution of (1.1), we have

$$A(t_{n,i})u'(t_{n,i})+B(t_{n,i})u(t_{n,i})=g(t_{n,i}). \tag{4.5}$$

Substituting (4.4) into (4.5) and noticing that

$$B(t_{n,i})=B(t_n)+O(h),$$

we obtain

$$A(t_{n,i})\Phi_i'+B(t_{n,i})\Phi_i=g(t_{n,i})+h^2\delta_{(i)}B(t_n)u''(t_n)+O(h^3). \tag{4.6}$$

Now, we introduce several variables. First we put

$$\Delta u_n=\phi_n-u_n, \quad \Delta U_i'=\Phi_i'-U_i', \quad \Delta U_i=\Phi_i-U_i.$$

From (2.2), (4.1), and (4.6), they satisfy

$$A(t_{n,i})\Delta U_i'+B(t_{n,i})\Delta U_i=h^2\delta_{(i)}B(t_n)u''(t_n)+O(h^3), \tag{4.7}$$

$$\Delta U_i=\Delta u_n+h\sum_{j=1}^{s}a_{ij}\Delta U_j',$$

$$\Delta u_{n+1}=\Delta u_n+h\sum_{i=1}^{s}b_i\Delta U_i'. \tag{4.8}$$

Furthermore, let

$$(\Delta y_n, \Delta z_n)^T=Q(t_n)^{-1}\Delta u_n, \quad (\Delta Y_i', \Delta Z_i')^T=Q(t_n)^{-1}\Delta U_i'$$
$$(\Delta Y_i, \Delta Z_i)^T=Q(t_n)^{-1}\Delta U_i.$$

Multiplying each equation of (4.8) by $Q(t_n)^{-1}$, we have

$$\Delta Y_i=\Delta y_n+h\sum_{j=1}^{s}a_{ij}\Delta Y_j', \tag{4.9.a}$$

$$\Delta y_{n+1}=\Delta y_n+h\sum_{i=1}^{s}b_i\Delta Y_i', \tag{4.9.b}$$

$$\Delta Z_i=\Delta z_n+h\sum_{j=1}^{s}a_{ij}\Delta Z_j', \tag{4.10.a}$$

$$\Delta z_{n+1}=\Delta z_n+h\sum_{i=1}^{s}b_i\Delta Z_i'. \tag{4.10.b}$$

Moreover, a simple computation shows that

$$Q(t_{n,i})^{-1}Q(t_n)=I-c_ihQ(t_n)^{-1}Q'(t_n)+O(h^2)$$
$$=\begin{pmatrix}1+O(h) & -c_ihR_n+O(h^2)\\ O(h) & 1+O(h)\end{pmatrix},$$

where R_n is the (1, 2)-component of $Q(t_n)^{-1}Q'(t_n)$. Thus, noticing that

$$A(t_{n,i})=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}Q(t_{n,i})^{-1},$$

$$B(t_{n,i})=\begin{pmatrix}C(t_{n,i}) & 0\\ 0 & 1\end{pmatrix}Q(t_{n,i})^{-1},$$

and letting

$$(\eta_n, \zeta_n)^T = B(t_n)u''(t_n),$$

we can rewrite (4.7) in the form

$$(1+O(h))\Delta Y_i' + (-c_i hR_n + O(h^2))\Delta Z_i' + (C(t_n)+O(h))\Delta Y_i + O(h)\Delta Z_i = h^2\delta_{(i)}\eta_n + O(h^3), \quad (4.11)$$

$$O(h)\Delta Y_i + (1+O(h))\Delta Z_i = h^2\delta_{(i)}\zeta_n + O(h^3), \quad (4.12)$$

In the remainder of this section, we will prove that

$$|\Delta y_{n+1}| \leqq (1+O(h))|\Delta y_n| + O(h)|\Delta z_n| + O(h^4), \quad (4.13)$$

$$|\Delta z_{n+1}| \leqq O(h)|\Delta y_n| + |\gamma||\Delta z_n| + O(h^3). \quad (4.14)$$

Note that $\Delta u_0 = 0$, since u_0 is given by the initial value of the true solution. The proof of the theorem immediately follows from the following lemma (cf. (4.3)).

Lemma. *If* (4.13) *and* (4.14) *are satisfied, then*

$$\|\Delta u_n\| \leqq C\|\Delta u_0\| + O(h^3). \quad (4.15)$$

Proof. By a similar argument to that in part b) of the proof of Theorem 7 in Hairer et al. [6], we obtain

$$|\Delta y_n| \leqq (1+O(h))|\Delta y_0| + O(h)|\Delta z_0| + O(h^3),$$

$$|\Delta z_n| \leqq O(h)|\Delta y_0| + (1+O(h))|\Delta z_0| + O(h^3).$$

Since $\Delta u_n = Q(t_n)(\Delta y_n, \Delta z_n)^T$, these estimates yield (4.15). **Q.E.D.**

Let

$$\Delta Y = (\Delta Y_1, \Delta Y_2, \ldots, \Delta Y_s)^T,$$

$$\Delta Y' = (\Delta Y_1', \Delta Y_2', \ldots, \Delta Y_s')^T,$$

$$\Delta Z = (\Delta Z_1, \Delta Z_2, \ldots, \Delta Z_s)^T,$$

$$\Delta Z' = (\Delta Z_1', \Delta Z_2', \ldots, \Delta Z_s')^T.$$

Since (4.11) yields

$$\Delta Y' = -C(t_n)\Delta Y + O(h),$$

from (4.9.a) we have

$$(I + hC(t_n)\mathbf{A})\Delta Y = \Delta y_n \mathbf{e} + O(h^2).$$

Thus we obtain

$$|\Delta Y_i| \leqq (1+O(h))|\Delta y_n| + O(h^2). \quad (4.16)$$

On the other hand, (4.10.a) and (4.12) are rewritten as

$$h\Delta Z' = \mathbf{A}^{-1}(\Delta Z - \Delta z_n \mathbf{e}) \quad (4.17)$$

and

$$\Delta Z = O(h)\Delta Y + h^2\delta\zeta_n + O(h^3), \quad (4.18)$$

respectively. Substituting (4.18) into (4.17), we have

$$h\Delta Z' = \mathbf{A}^{-1}(O(h)\Delta Y + h^2\delta\zeta_n - \Delta z_n \mathbf{e}) + O(h^3). \quad (4.19)$$

Hence, by (4.10.b), we find

$$\Delta z_{n+1} = \Delta z_n + h\mathbf{b}^T\Delta Z' = (1 - \mathbf{b}^T\mathbf{A}^{-1}\mathbf{e})\Delta z_n + O(h)\Delta Y + h^2\mathbf{b}^T\mathbf{A}^{-1}\delta\zeta_n + O(h^3).$$

Note that $\gamma = 1 - \mathbf{b}^T\mathbf{A}^{-1}\mathbf{e}$ from (2.8). We obtain the estimate (4.14), using (4.16) and condition (ii)'.

We can rewrite (4.11) as

$$\Delta Y' = -(C(t_n)+O(h))\Delta Y + (hR_n\mathbf{C} + O(h^2))\Delta Z' + O(h)\Delta Z + h^2\delta\eta_n + O(h^3). \quad (4.20)$$

where $\mathbf{C} = \mathrm{diag}\,(c_1, c_2, \ldots, c_s)$. Substituting (4.20) into (4.9.b), we obtain

$$\Delta y_{n+1} = \Delta y_n + \mathbf{b}^T h\{-(C(t_n)+O(h))\Delta Y + hR_n\mathbf{C}\Delta Z' + O(h^2)\Delta Z' + O(h)\Delta Z + h^2\delta\eta_n\} + O(h^4), \quad (4.21)$$

From (4.9.a), (4.20), (4.18) and (4.19) it follows that

$$\Delta Y = \Delta y_n + h\mathbf{A}\Delta Y' = \Delta y_n + O(h)\Delta z_n - h(C(t_n)+O(h))\mathbf{A}\Delta Y + O(h^3).$$

This, together with (4.16), yields

$$|\Delta Y_i| \leqq (1+O(h))|\Delta y_n| + O(h)|\Delta z_n| + O(h^3). \quad (4.22)$$

On the other hand, using (4.19) and condition (iii)', we obtain

$$\mathbf{b}^T\mathbf{C}h^2\Delta Z' = h(\mathbf{bc})^T\mathbf{A}^{-1}(O(h)\Delta Y + h^2\delta\zeta_n - \Delta z_n\mathbf{e}) + O(h^4) = h(\mathbf{bc})^T\mathbf{A}^{-1}(O(h)\Delta Y - \Delta z_n\mathbf{e}) + O(h^4). \quad (4.23)$$

Noticing that $\mathbf{b}^T\delta = 0$ from condition (i), we finally obtain the estimate (4.13) from (4.21), (4.22), (4.23), (4.18), and (4.19).

5. Concluding Remarks

In this paper, we deal only with linear DAE systems. However, the present method can be at least theoretically applied to a certain non-linear system by combining a technique of linearization of the system [e.g. 8], though in that case several problems remain to be solved for efficient implementation of the method. The most essential problem is how one can efficiently solve the algebraic equations arising in the evaluation of the implicit formula.

Acknowledgement

We wish to thank Professor Suzuki of Shizuoka Institute of Science and Technology for his careful reading of the manuscript and helpful suggestions. Professor Bao of Nanjing University provided the impetus for the present study.

References

1. ALEXANDER, R. Diagonally Implicit Runge-Kutta Methods for Stiff ODEs, *SIAM J. Numer. Anal.*, **14** (1977), 1006-1021.
2. BRENAN, K. E., CAMPBELL, S. L. and PETZOLD, L. R. *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations, North-Holland*, Amsterdam, 1989.
3. BUTCHER, J. C. *The Numerical Analysis of Ordinary Differential Equations: Runge-Kutta and General Linear Methods, John Wiley & Sons*, Chichester, 1987.
4. GANTMACHER, F. R. *The Theory of Matrices*, **2**, Chapter XII, *Chelsea*, New York, 1964.
5. GEAR, C. W. and PETZOLD, L. R. ODE Methods for the Solution

of Differential/Algebraic Systems, *SIAM J. Numer. Anal.*, **21** (1984), 716-728.

6. HAIRER, E., LUBICH, Ch. and ROCHE, M. Error of Runge-Kutta Methods for Stiff Problems Studied via Differential Algebraic Equations, *BIT*, **28** (1988), 678-700.

7. HAIRER, E., LUBICH, Ch. and ROCHE, M. *The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods*, Lecture Notes in Math., **1409**, *Springer-Verlag*, Berlin, 1989.

8. PETZOLD, L. R. Order Results for Implicit Runge-Kutta Methods Applied to Differential/Algebraic Systems, *SIAM J. Numer. Anal.*, **23** (1986), 837-852.

9. ROCHE, M. Implicit Runge-Kutta Methods for Differential Algebraic Equations, *SIAM J. Numer. Anal.*, **26** (1989), 963-975.

10. SINCOVEC, R. F., ERISMAN, A., YIP, E. L. and EPTON, M. A.: Analysis of Descriptor Systems Using Numerical Algorithms, *IEEE Trans. Automat., Contr.*, **26** (1981), 139-147.

(Received November 29, 1989; revised May 11, 1990)