Numerical Algorithms 16 (1997) 209–230

Nordsieck representation of DIMSIMs

J.C. Butcher [a,*], P. Chartier [b] and Z. Jackiewicz [c,**]

[a] *Department of Mathematics, The University of Auckland, Private Bag 92019, Auckland, New Zealand*
[b] *INRIA, Campus Universitaire de Beaulieu, F-35042 Rennes Cedex, France*
[c] *Department of Mathematics, Arizona State University, Tempe, AZ 85287, USA*

Received 17 October 1997
Communicated by C. Brezinski

A new representation for diagonally implicit multistage integration methods (DIMSIMs) is derived in which the vector of external stages directly approximates the Nordsieck vector. The methods in this formulation are zero-stable for any choice of variable mesh. They are also easy to implement since changing step-size corresponds to a simple rescaling of the vector of external approximations.The paper contains an analysis of local truncation error and of error accumulation in a variable step-size situation.

Keywords: DIMSIM methods, Nordsieck representation, variable step-size, error accumulation, error estimation

1. Introduction

Diagonally implicit multistage integration methods (DIMSIMs) for systems of ordinary differential equations (ODEs)

$$\begin{cases} y'(x) = f(y(x)), & x \in [x_0, X], \\ y(x_0) = y_0, \end{cases} \tag{1.1}$$

$f : \mathbb{R}^m \to \mathbb{R}^m$, have the form[1]

$$\begin{cases} Y^{[n]} = h(A \otimes I_m)F(Y^{[n]}) + (U \otimes I_m)y^{[n-1]}, \\ y^{[n]} = h(B \otimes I_m)F(Y^{[n]}) + (V \otimes I_m)y^{[n-1]}, \end{cases} \tag{1.2}$$

* The work of this author was assisted by the Marsden Fund of New Zealand.
** The work of this author was partially supported by the National Science Foundation under grant NSF DMS-9208048.

[1] The tensor product $\Delta \otimes \Lambda$ of two matrices $\Delta \in \mathbb{R}^{m_1,m_2}$ and $\Lambda \in \mathbb{R}^{n_1,n_2}$ is the block-matrix

$$\begin{bmatrix} \delta_{1,1}\Lambda & \dots & \delta_{1,m_2}\Lambda \\ \vdots & & \vdots \\ \delta_{1,1}\Lambda & \dots & \delta_{1,m_2}\Lambda \end{bmatrix}.$$

$n = 1, 2, \ldots, N$, $Nh = X - x_0$. Here, $Y^{[n]} \in \mathbb{R}^{ms}$ is an approximation to the vector $Y(x_{n-1})$ with components $y(x_{n-1} + c_i h) \in \mathbb{R}^s$, $i = 1, 2, \ldots, m$,

$$F(Y^{[n]}) = \begin{bmatrix} f(Y_1^{[n]}) \\ f(Y_2^{[n]}) \\ \vdots \\ f(Y_s^{[n]}) \end{bmatrix},$$

and $y^{[n]} \in \mathbb{R}^{mr}$ are external approximations which propagate to the next step. These methods were introduced by Butcher [2] and further investigated in [3,5–9].

It will always be assumed, unless explicitly mentioned, that V is a rank one matrix of the form $V = uv^T$, $u, v \in \mathbb{R}^r$, such that $v^T u = 1$. This condition guarantees that (1.2) is zero-stable. It will also be assumed, throughout the paper, that $c_i \neq c_j$, if $i \neq j$. Without this assumption, many details become impossibly complicated and, for example, theorem 2 would not hold.

The method (1.2) has order p and stage order q if

$$y^{[n-1]} = \sum_{k=0}^{p} h^k \left(\alpha_k \otimes y^{(k)}(x_{n-1})\right) + \mathrm{O}\left(h^{p+1}\right) \tag{1.3}$$

implies that

$$Y^{[n]} = \sum_{k=0}^{q} h^k \left(\frac{c^k}{k!} \otimes y^{(k)}(x_{n-1})\right) + \mathrm{O}\left(h^{q+1}\right) \tag{1.4}$$

and

$$y^{[n]} = \sum_{k=0}^{p} h^k \left(\alpha_k \otimes y^{(k)}(x_n)\right) + \mathrm{O}\left(h^{p+1}\right), \tag{1.5}$$

for some vectors $\alpha_0, \alpha_1, \ldots, \alpha_p \in \mathbb{R}^r$ associated with the method. Here, c^k denotes component-wise powers of c. The conditions which guarantee that (1.2) has order p and stage order $q = p$ were derived by Butcher in [2]. We will collect the vectors α_k in the matrix W defined by

$$W = [\,\alpha_0 \quad \alpha_1 \quad \ldots \quad \alpha_p\,].$$

Theorem 1. The vectors α_k, $k = 0, 1, \ldots, p$, in (1.3) and the matrices A, B, U, and V satisfy (1.4) and (1.5) with $q = p$ if and only if

$$\mathrm{e}^{cz} = zA\mathrm{e}^{cz} + Uw + \mathrm{O}\left(z^{p+1}\right), \tag{1.6}$$

$$\mathrm{e}^{z} w = zB\mathrm{e}^{cz} + Vw + \mathrm{O}\left(z^{p+1}\right), \tag{1.7}$$

where $\mathrm{e}^{cz} = [\mathrm{e}^{c_1 z}, \ldots, \mathrm{e}^{c_s z}]^T$ and $w = \sum_{k=0}^{p} \alpha_k z^k$.

If $U = I_s$, it is easy to verify by expanding both sides of (1.6) into Taylor series around $z = 0$ and comparing the coefficients of z^k in the resulting expressions that the vectors α_k appearing in (1.3) and (1.5) have the form

$$\alpha_0 = \mathrm{e}, \qquad \alpha_k = \frac{c^k}{k!} - \frac{Ac^{k-1}}{(k-1)!}, \quad k = 1, 2, \ldots, p.$$

If in addition $p = q = r = s$, there exists a convenient representation formula for the coefficient matrix B in terms of the matrices A, V, and the vector c.

Theorem 2 (Butcher [2]). Let $p = q = r = s$ and $U = I_s$. Then DIMSIM (1.2) is of order p and stage order $q = p$ if and only if

$$B = B_0 - AB_1 - VB_2 + VA,$$

where the (i, j) elements of B_0, B_1, and B_2 are given by

$$\int_0^{1+c_i} \phi_j(x)\,\mathrm{d}x/\phi_j(c_j), \quad \phi_j(1 + c_i)/\phi_j(c_j), \quad \text{and} \quad \int_0^{c_i} \phi_j(x)\,\mathrm{d}x/\phi_j(c_j),$$

respectively, and $\phi_j(x) = \prod_{k \neq j}(x - c_k)$.

This formula for B considerably simplifies the construction of DIMSIMs. Many examples of such methods were derived in [2,3,5,6] using symbolic manipulation packages and the approach based on homotopy method, and in [7,9] using the approach based on least squares minimization and a variant of the Fourier series method. The implementation issues for (1.2) such as changing step-size using the Nordsieck technique, local error estimation and construction of continuous interpolants were addressed in [8].

Although the $U = I_s$ seems to be a special case, it can, for many methods, be regarded as a convenient choice amongst various possible representations. Since the purpose of the r vectors making up the input data $y^{[n-1]}$ and the corresponding output data $y^{[n]}$ is merely to carry information from step to step, it is possible to rearrange this data by taking r independent linear combinations of the r sub-vectors and using the resulting combinations instead of the sub-vectors themselves. This is equivalent to choosing a non-singular matrix T and replacing the partitioned matrix that characterizes the method

$$\begin{bmatrix} A & U \\ B & V \end{bmatrix},$$

by the "transformed method"

$$\begin{bmatrix} A & UT \\ T^{-1}B & T^{-1}VT \end{bmatrix}.$$

If U is square and non-singular, then transforming it to the case $U = I_s$ is easily achieved by the choice $T = U^{-1}$.

It is the purpose of this paper to derive a more convenient representation for (1.2) which will be easy to implement in a variable step-size environment. In this representation the vector of external stages will directly approximate the Nordsieck vector

$$z(x_n) = \begin{bmatrix} y(x_n) \\ hy'(x_n) \\ \vdots \\ h^p y^{(p)}(x_n) \end{bmatrix},$$

where y is the solution to (1.1). This representation will be zero-stable for any step-size pattern and changing step-size will be accomplished by a simple rescaling of the vector of external approximations.

The paper is organized with the discussion of Nordsieck representation in section 2 and an analysis of local truncation error, for methods designed in this way, in section 3. The accumulation of truncation error, in a variable step-size situation, is considered in section 4 and the estimation of local truncation error in section 5. Examples of methods, as transformed to Nordsieck form, are given in section 6 and the results of some numerical tests are presented in section 7.

2. Nordsieck representation of DIMSIMs

Our starting point is the method (1.2) such that $p = q = r = s$ and with $U = I_s$. Let us consider, in addition to $y^{[n]}$, an approximation $\eta^{[n]} \in \mathbb{R}^m$ of order p to $\sum_{k=0}^{p} h^k t_k y^{(k)}(x_n)$ of the form

$$\eta^{[n]} = h\big(b^T \otimes I_m\big) F\big(Y^{[n]}\big) + \big(v^T \otimes I_m\big) y^{[n-1]}.$$

To avoid the possibility that $\eta^{[n]}$ can be written as a linear combination of the output quantities, $y_i^{[n]}$, $i = 1, 2, \ldots, r$, we assume that the matrix

$$\begin{bmatrix} B & e \\ b^T & 1 \end{bmatrix}$$

is non-singular.

Define the vector $t = [t_0, t_1, \ldots, t_p]^T$ and the matrix

$$\widetilde{W} = \begin{bmatrix} W \\ t^T \end{bmatrix} = \begin{bmatrix} \alpha_0 & \alpha_1 & \ldots & \alpha_p \\ t_0 & t_1 & \ldots & t_p \end{bmatrix}.$$

The independence of the vectors $y_i^{[n]}$, $i = 1, 2, \ldots, s$, and $\eta^{[n]}$ guarantees that $\widetilde{W}$ is non-singular. The relationship between t and b is given by the following result.

Theorem 3. The first component of t is given by $t_0 = 1$ and for given $t_1, t_2, \ldots, t_p$ the vector b is equal to

$$b^T = l^T C^{-1}, \tag{2.1}$$

where C is the Vandermonde matrix

$$C = \begin{bmatrix} e & c & \dots & c^{p-1} \end{bmatrix}$$

and the vector l with components l_k is defined by

$$l_k = (k-1)!\left(\sum_{j=0}^{k} \frac{t_{k-j}}{j!} - v^T\alpha_k\right), \quad k = 1, 2, \dots, p.$$

Proof. It follows from theorem 1 (formula (1.7)) that

$$\mathrm{e}^z\xi = zb^T\mathrm{e}^{cz} + v^T w + \mathrm{O}\big(z^{p+1}\big),$$

where $\xi = \sum_{k=0}^{p} t_k z^k$. Expanding both sides of this equation into Taylor series around $z = 0$ and comparing the corresponding coefficients of z^k we obtain

$$v^T\alpha_0 = t_0 \tag{2.2}$$

and

$$\frac{b^T c^k}{k!} = \sum_{j=0}^{k+1} \frac{t_{k+1-j}}{j!} - v^T\alpha_{k+1}, \quad k = 0, 1, \dots, p-1. \tag{2.3}$$

It follows from (2.2) that $t_0 = 1$ and equation (2.3) is equivalent to (2.1). □

Put $\widetilde{y}^{[n]} = [y^{[n]^T}, \eta^{[n]^T}]^T$ and consider the method

$$\begin{cases} Y^{[n]} = h(A \otimes I_m)F(Y^{[n]}) + (\widetilde{U} \otimes I_m)\widetilde{y}^{[n-1]}, \\ \widetilde{y}^{[n]} = h(\widetilde{B} \otimes I_m)F(Y^{[n]}) + (\widetilde{V} \otimes I_m)\widetilde{y}^{[n-1]}, \end{cases} \tag{2.4}$$

where

$$\widetilde{U} = \begin{bmatrix} U & 0 \end{bmatrix}, \qquad \widetilde{B} = \begin{bmatrix} B \\ b^T \end{bmatrix}, \qquad \widetilde{V} = \begin{bmatrix} V & 0 \\ v^T & 0 \end{bmatrix}.$$

Since

$$\widetilde{y}^{[n]} = \big(\widetilde{W} \otimes I_m\big) \begin{bmatrix} y(x_n) \\ hy'(x_n) \\ \vdots \\ h^p y^{(p)}(x_n) \end{bmatrix} + \mathrm{O}\big(h^{p+1}\big)$$

we define the vector $z^{[n]}$ by

$$\widetilde{y}^{[n]} = \big(\widetilde{W} \otimes I_m\big) z^{[n]}. \tag{2.5}$$

Observing that

$$\widetilde{U}\widetilde{W} = \begin{bmatrix} U & 0 \end{bmatrix} \begin{bmatrix} W \\ t^T \end{bmatrix} = UW$$

and substituting (2.5) into (2.4) we obtain

$$\begin{cases} Y^{[n]} = h(A \otimes I_m)F(Y^{[n]}) + (P \otimes I_m)z^{[n-1]}, \\ z^{[n]} = h(G \otimes I_m)F(Y^{[n]}) + (Q \otimes I_m)z^{[n-1]}, \end{cases} \quad n = 1, 2, \ldots, N, \tag{2.6}$$

where $P = UW$, $G = \widetilde{W}^{-1}\widetilde{B}$ and $Q = \widetilde{W}^{-1}\widetilde{V}\widetilde{W}$. Since

$$z^{[n]} = \big(\widetilde{W}^{-1} \otimes I_m\big)\widetilde{y}^{[n]} = \begin{bmatrix} y(x_n) \\ hy'(x_n) \\ \vdots \\ h^p y^{(p)}(x_n) \end{bmatrix} + \mathrm{O}\big(h^{p+1}\big), \tag{2.7}$$

this is the desired Nordsieck representation of DIMSIM (1.2).

To simplify Q observe that $\widetilde{W}[e^T, 1]^T = e_1$, where $e_1 = [1, 0, \ldots, 0]^T \in \mathbb{R}^{s+1}$. It follows that

$$Q = \widetilde{W}^{-1} \begin{bmatrix} e \\ 1 \end{bmatrix} \begin{bmatrix} v^T & 0 \end{bmatrix} \begin{bmatrix} W \\ t \end{bmatrix} = e_1 \begin{bmatrix} 1 & v^T\alpha_1 & \ldots & v^T\alpha_p \end{bmatrix}.$$

The matrix G is characterized by the following theorem.

Theorem 4. The matrix G is determined by the formula

$$G = L\,C^{-1}, \tag{2.8}$$

where C is the Vandermonde matrix defined in theorem 3 and L is the matrix with columns L_k given by

$$L_k = (k-1)!\left(\sum_{j=0}^{k} \frac{e_{j+1}}{(k-j)!} - Qe_{k+1}\right), \quad k = 1, 2, \ldots, p.$$

Here, e_i, $i = 1, 2, \ldots, p+1$, is the canonical basis in $\mathbb{R}^{p+1}$. In particular, the matrix G is independent of $t_1, t_2, \ldots, t_p$.

Proof. It follows from (2.7) and theorem 1 (formula (1.7)) that

$$\mathrm{e}^z\widetilde{w} = zG\mathrm{e}^{cz} + Q\widetilde{w} + \mathrm{O}\big(z^{p+1}\big), \tag{2.9}$$

where $\widetilde{w} = \sum_{k=0}^{p} e_{k+1}z^k$. Expanding both sides of (2.9) into Taylor series around $z = 0$ and comparing the corresponding coefficients of z^k we obtain

$$Qe_1 = e_1 \tag{2.10}$$

and

$$\frac{Gc^k}{k!} = \sum_{j=0}^{k+1} \frac{e_{j+1}}{(k+1-j)!} - Qe_{k+2}, \quad k = 0, 1, \ldots, p-1. \tag{2.11}$$

It follows from the form of the matrix Q that the condition (2.10) is automatically satisfied and (2.11) is equivalent to (2.8). □

On a nonuniform grid $x_0 < x_1 < \cdots < x_N$, $x_N \geqslant X$, the method (2.6) takes the form

$$\begin{cases} Y^{[n]} = h_n(A \otimes I_m)F(Y^{[n]}) + (P \otimes I_m)\widehat{z}^{[n-1]}, \\ z^{[n]} = h_n(G \otimes I_m)F(Y^{[n]}) + (Q \otimes I_m)\widehat{z}^{[n-1]}, \end{cases} \quad n = 1, 2, \ldots, N, \tag{2.12}$$

where $h_n = x_n - x_{n-1}$, and $\widehat{z}^{[n-1]}$ is an approximation of order p to

$$\widehat{z}(x_{n-1}) = \begin{bmatrix} y(x_{n-1}) \\ h_n y'(x_{n-1}) \\ \vdots \\ h_n^p y^{(p)}(x_{n-1}) \end{bmatrix}.$$

Since

$$\begin{bmatrix} y(x_{n-1}) \\ h_n y'(x_{n-1}) \\ \vdots \\ h_n^p y^{(p)}(x_{n-1}) \end{bmatrix} = \begin{bmatrix} I_m & 0 & \ldots & 0 \\ 0 & \delta_n I_m & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \delta_n^p I_m \end{bmatrix} \begin{bmatrix} y(x_{n-1}) \\ h_{n-1} y'(x_{n-1}) \\ \vdots \\ h_{n-1}^p y^{(p)}(x_{n-1}) \end{bmatrix},$$

$\delta_n = h_n/h_{n-1}$, $\widehat{z}^{[n-1]}$ appearing in (2.12) is defined by the formula

$$\widehat{z}^{[n-1]} = \big(D(\delta_n) \otimes I_m\big) z^{[n-1]}, \tag{2.13}$$

where we have used the notation

$$D(\delta_n) = \operatorname{diag}\big(1, \delta_n, \ldots, \delta_n^p\big).$$

It follows from (2.13) that zero-stability properties of the method (2.12) are determined by the eigenvalues of the matrix $QD(\delta_n) \otimes I_m$. Since the matrix $QD(\delta_n)$ has simple eigenvalue equal to one and eigenvalue zero of multiplicity p for any ratio δ_n, it follows that (2.12) is zero-stable for any variable step-size pattern. This is in contrast to the strategy proposed in [8] for DIMSIMs of the form (1.2) where desirable zero-stability properties were enforced by a suitable choice of some free parameters associated with a matrix which affects the computation of rescaled quantities $\widehat{y}^{[n-1]}$ corresponding to $y^{[n-1]}$.

More generally, if $M(\zeta) = V + \zeta B(I - \zeta A)^{-1} U$ denotes the stability matrix of method (1.2), then the stability matrix of the extended method (2.6) is

$$\widetilde{M}(\zeta) = \begin{bmatrix} V & 0 \\ v^T & 0 \end{bmatrix} + \zeta \begin{bmatrix} B \\ b^T \end{bmatrix} (I - \zeta A)^{-1} \begin{bmatrix} U & 0 \end{bmatrix} = \begin{bmatrix} M(\zeta) & 0 \\ v^T + \zeta b^T (I - \zeta A)^{-1} U & 0 \end{bmatrix}.$$

As a consequence, $\widetilde{M}$ and M have the same eigenvalues and the linear stability properties of the method remain unchanged for constant step-sizes through the augmentation process.

3. Local discretization error of DIMSIM in Nordsieck form

From now on we will assume that the DIMSIM has been brought to its Nordsieck form and is thus defined in its variable step-size version by

$$\begin{cases} Y^{[n]} = h_n(A \otimes I_m)F(Y^{[n]}) + (PD(\delta_n) \otimes I_m)z^{[n-1]}, \\ z^{[n]} = h_n(G \otimes I_m)F(Y^{[n]}) + (QD(\delta_n) \otimes I_m)z^{[n-1]}, \end{cases} \tag{3.1}$$

where $A \in \mathbb{R}^{p\times p}$, $P \in \mathbb{R}^{p\times(p+1)}$, $G \in \mathbb{R}^{(p+1)\times p}$, $Q \in \mathbb{R}^{(p+1)\times(p+1)}$, and $\delta_n = h_n/h_{n-1}$.

Definition 5. The local discretization error at the point x_n for DIMSIM (3.1) is given by

$$\Gamma(x_n) = z(x_n) - h_n(G \otimes I_m)F\big(\Upsilon(x_n)\big) - (Q \otimes I_m)\widehat{z}(x_{n-1}), \tag{3.2}$$

where the vector $\Upsilon(x_n)$ is defined by

$$\Upsilon(x_n) = h_n(A \otimes I_m)F\big(\Upsilon(x_n)\big) + (P \otimes I_m)\widehat{z}(x_{n-1}). \tag{3.3}$$

We recall that

$$z(x_n) = \begin{bmatrix} y(x_n) \\ h_n y'(x_n) \\ \vdots \\ h_n^p y^{(p)}(x_n) \end{bmatrix} \quad \text{and} \quad \widehat{z}(x_{n-1}) = \begin{bmatrix} y(x_{n-1}) \\ h_n y'(x_{n-1}) \\ \vdots \\ h_n^p y^{(p)}(x_{n-1}) \end{bmatrix}.$$

Assuming that the solution y to (1.1) is sufficiently smooth, and that $p = q$, the vector $\Upsilon(x_n)$ can be written as

$$\Upsilon(x_n) = Y(x_{n-1}) + h_n^{p+1}\gamma(x_{n-1}) + \mathrm{O}\big(h_n^{p+2}\big), \tag{3.4}$$

where $h_n^{p+1}\gamma(x_{n-1})$ is the principal part of the error of the internal stages $\Upsilon(x_n)$. This function can be computed by substituting (3.4) into (3.3). This leads to

$$\begin{aligned} &\big(C_p \otimes I_m - (A \otimes I_m)(C_p K \otimes I_m) - P \otimes I_m\big)\widehat{z}(x_{n-1}) \\ &\quad + h_n^{p+1}\big(\gamma(x_{n-1}) + \alpha_{p+1} \otimes y^{(p+1)}(x_{n-1})\big) = \mathrm{O}\big(h_n^{p+2}\big), \end{aligned} \tag{3.5}$$

where C_p and K are defined by

$$C_p = \begin{bmatrix} e & c & \dots & \dfrac{c^p}{p!} \end{bmatrix}, \qquad K = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 1 \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix} \in \mathbb{R}^{p\times p},$$

and where

$$\alpha_{p+1} = \frac{c^{p+1}}{(p+1)!} - \frac{Ac^p}{p!}.$$

Using the properties of the Kronecker product [11] we write (3.5) as

$$\begin{aligned} &\big((C_p - AC_pK - P)\otimes I_m\big)\widehat{z}(x_{n-1}) \\ &\quad + h_n^{p+1}\big(\gamma(x_{n-1}) + \alpha_{p+1}\otimes y^{(p+1)}(x_{n-1})\big) = \mathrm{O}\big(h_n^{p+2}\big). \end{aligned} \tag{3.6}$$

The first term in (3.6) vanishes in view of relation (1.6) in theorem 1 and equating to zero the second term we obtain

$$\gamma(x_{n-1}) = -\alpha_{p+1}\otimes y^{(p+1)}(x_{n-1}).$$

Put

$$a = \begin{bmatrix} \dfrac{1}{(p+1)!} & \dfrac{1}{p!} & \dots & 1 \end{bmatrix}^T.$$

We have the following theorem.

Theorem 6. The local discretization error of the method (3.1) at the point x_n is given by

$$\Gamma(x_n) = h_n^{p+1}\big(\varphi_p\otimes y^{(p+1)}(x_{n-1})\big) + \mathrm{O}\big(h_n^{p+2}\big), \tag{3.7}$$

where

$$\varphi_p = a - \frac{Gc^p}{p!}.$$

Remark. Similar estimates have been derived in [4,8].

4. Accumulation of errors for non-stiff problems

We deal here with non-stiff equations so that the function f is supposed to satisfy a Lipschitz condition of the form

$$\big\|f(y_1) - f(y_2)\big\| \leqslant L\|y_1 - y_2\| \tag{4.1}$$

for $y_1, y_2 \in \mathbb{R}^m$. We consider a more general matrix Q of rank possibly higher than one. However, we will assume that Q has only one eigenvalue of modulus one and multiplicity one. The preconsistency condition $Qe_1 = e_1$ then enables us to decompose Q as

$$Q = e_1 q^T + \widetilde{Q}, \tag{4.2}$$

with $\widetilde{Q}e_1 = 0$, $q^T\widetilde{Q} = 0$ and $q^T e_1 = 1$. Furthermore, if the method is zero-stable, then $\rho(\widetilde{Q}) < 1$ and there exists a Euclidean norm $\|\cdot\|_S$ on $\mathbb{R}^r$ (S is a symmetric positive definite matrix of $\mathbb{R}^{r\times r}$ normalized by the relation $e_1^T S e_1 = 1$), two positive reals $\delta_{\min} < 1$ and $\delta_{\max} > 1$ and a real μ, $0 \leqslant \mu < 1$ such that

$$\left\|\widetilde{Q}D(\delta)\right\|_S \leqslant \mu, \quad \text{for all } \delta \in [\delta_{\min}, \delta_{\max}]. \tag{4.3}$$

We also denote

$$\eta = \max_{\delta_{\min} \leqslant \delta \leqslant \delta_{\max}} \left\|D(\delta)q\right\|_{S^{-1}}.$$

By convention, and for any family of square matrices $(X_k)_{k\in\mathbb{N}}$, we will write

$$\prod_{k=i}^{j} X_k = X_j \cdots X_i, \quad i \leqslant j.$$

Lemma 7. Let $(i,j) \in \mathbb{N}\times\mathbb{N}$ with $j \geqslant i$. Then we have

$$\prod_{k=i}^{j} \big(QD(\delta_k)\big) = e_1 \vartheta_{j,i}^T + \prod_{k=i}^{j} \big(\widetilde{Q}D(\delta_k)\big),$$

with $\vartheta_{j,j}^T = q^T D(\delta_j)$ and

$$\vartheta_{j,i-1}^T = q^T D(\delta_{i-1}) + \vartheta_{j,i}^T \widetilde{Q} D(\delta_{i-1}). \tag{4.4}$$

For any $\delta \in [\delta_{\min}, \delta_{\max}]$, we have in addition

$$\|\vartheta_{j,i}\|_{S^{-1}} \leqslant \frac{\eta}{1-\mu}.$$

Proof. The relation

$$QD(\delta_j) = e_1 q^T D(\delta_j) + \widetilde{Q}D(\delta_j)$$

gives immediately $\vartheta_{j,j}^T = q^T D(\delta_j)$ ($\vartheta_{j,j}^T e_1 = q^T e_1 = 1$). Let us now assume that for a given i, $0 \leqslant i+1 \leqslant j$,

$$\prod_{k=i+1}^{j} \big(QD(\delta_k)\big) = e_1 \vartheta_{j,i+1}^T + \prod_{k=i+1}^{j} \big(\widetilde{Q}D(\delta_k)\big),$$

with $\vartheta_{j,i+1}^T e_1 = 1$. Then it follows that

$$\begin{aligned}\prod_{k=i}^{j}\big(QD(\delta_k)\big) &= \left(e_1\vartheta_{j,i+1}^T + \prod_{k=i+1}^{j}\big(\widetilde{Q}D(\delta_k)\big)\right)\big(e_1 q^T D(\delta_i) + \widetilde{Q}D(\delta_i)\big)\\ &= e_1\big(\vartheta_{j,i+1}^T e_1\big)q^T D(\delta_i) + e_1\vartheta_{j,i+1}^T\widetilde{Q}D(\delta_i)\\ &\quad + \left(\prod_{k=i+1}^{j}\big(\widetilde{Q}D(\delta_k)\big)\right)e_1 q^T D(\delta_i) + \prod_{k=i}^{j}\big(\widetilde{Q}D(\delta_k)\big).\end{aligned}$$

Since $\vartheta_{j,i+1}^T e_1 = 1$ and

$$\left(\prod_{k=i+1}^{j}\big(\widetilde{Q}D(\delta_k)\big)\right)e_1 = 0,$$

we obtain

$$\vartheta_{j,i}^T = q^T D(\delta_i) + \vartheta_{j,i+1}^T\widetilde{Q}D(\delta_i).$$

It follows that $\vartheta_{j,i}^T e_1 = q^T e_1 + \vartheta_{j,i+1}^T\widetilde{Q}e_1 = 1$ and relation (4.4) thus holds by induction on i. Besides, we have

$$\|\vartheta_{j,i}\|_{S^{-1}} \leqslant \eta + \mu\|\vartheta_{j,i+1}\|_{S^{-1}},$$

and the bound stated in the lemma can be obtained by induction on i. □

Remark. By convention, we shall define $\vartheta_{j,j+1}^T = q^T D(1) = q^T$.

Remark. If Q is assumed to be of rank one, then $\widetilde{Q} = 0$ and for any $(i,j) \in \mathbb{N}^2$, $i \leqslant j$, we have $\vartheta_{j,i}^T = q^T D(\delta_i)$.

Lemma 8. Let $\Delta z^{[n]} = z^{[n]} - z(x_n)$ denote the error at the nth step, and let $\Delta F^{[n]}$ be the quantity $F(Y^{[n]}) - F(\Upsilon(x_n))$. If $z^{[n]}$ satisfies the recursion (3.1), $q_1 = 1$, then we have

$$\begin{aligned}\Delta z^{[n]} &= \left(\prod_{k=1}^{n}\big(QD(\delta_k)\big)\otimes I_m\right)\Delta z^{[0]}\\ &\quad + \sum_{i=1}^{n-1}\left(\prod_{k=i+1}^{n}\big(QD(\delta_k)\big)\otimes I_m\right)\big(h_i(G\otimes I_m)\Delta F^{[i]} + \Gamma(x_i)\big)\\ &\quad + h_n(G\otimes I_m)\Delta F^{[n]} + \Gamma(x_n).\end{aligned}\tag{4.5}$$

Proof. We have on the one hand

$$\begin{cases} Y^{[n]} = h_n(A \otimes I_m)F(Y^{[n]}) + (PD(\delta_n) \otimes I_m)z^{[n-1]}, \\ z^{[n]} = h_n(G \otimes I_m)F(Y^{[n]}) + (QD(\delta_n) \otimes I_m)z^{[n-1]}, \end{cases} \tag{4.6}$$

and on the other hand

$$\begin{cases} \Upsilon(x_n) = h_n(A \otimes I_m)F(\Upsilon(x_n)) + (PD(\delta_n) \otimes I_m)z(x_{n-1}), \\ z(x_n) = h_n(G \otimes I_m)F(\Upsilon(x_n)) + (QD(\delta_n) \otimes I_m)z(x_{n-1}) - \Gamma(x_n). \end{cases} \tag{4.7}$$

Subtracting (4.7) from (4.6) we thus get

$$\Delta z^{[n]} = h_n(G \otimes I_m)\Delta F^{[n]} + \bigl(QD(\delta_n) \otimes I_m\bigr)\Delta z^{[n-1]} + \Gamma(x_n).$$

Formula (4.5) is then easily obtained by induction. □

Let $\|\cdot\|_2$ denote the 2-norm on $\mathbb{R}^m$. It can be extended to $\mathbb{R}^{rm}$ and $\mathbb{R}^{sm}$, respectively, by the definitions

$$\|g\|^2_{S\otimes I_m} = g^T(S \otimes I_m)g \quad \text{for all } g \in \mathbb{R}^{rm}$$

and

$$\|g\|^2_2 = g^T g \quad \text{for all } g \in \mathbb{R}^{sm}.$$

Note that we will also use the notations $\|\cdot\|_{S\otimes M}$ and $\|\cdot\|_2$ for the subordinated norms corresponding to linear operators.

Lemma 9.

$$\begin{aligned} &\|M \otimes I_m\|_{S\otimes I_m} = \|M\|_S, && M \in \mathcal{L}\bigl(\mathbb{R}^r, \mathbb{R}^r\bigr), \\ &\bigl\|(G \otimes I_m)g\bigr\|_{S\otimes I_m} \leqslant \bigl\|S^{1/2}G\bigr\|_2\|g\|_2, && g \in \mathbb{R}^{sm}, \\ &\bigl\|\bigl(w^T \otimes I_m\bigr)g\bigr\|_2 \leqslant \|w\|_{S^{-1}}\|g\|_{S\otimes I_m}, && w \in \mathbb{R}^r,\ g \in \mathbb{R}^{rm}. \end{aligned}$$

Proof. These inequalities derive straightforwardly from standard properties on norms and Kronecker products. □

Theorem 10. Assume that f satisfies the Lipschitz condition (4.1) and that there exist constants $\Omega > 0$ and $H > 0$ such that

$$\bigl\|\Delta z^{[0]}\bigr\|_{S\otimes I_m} \leqslant \Omega h^{p+1} \quad \text{and} \quad \bigl\|\Gamma(x_n)\bigr\|_{S\otimes I_m} \leqslant \Omega h^{p+1}$$

for $n = 1, 2, \ldots, N$ and for all $h \leqslant H$. Then there exists $\widetilde{h} > 0$ such that for any sequence of step-sizes $\{h_i\}_{i=0}^N$ satisfying

$$\begin{cases} \max\{h_i\colon\ i = 0, \ldots, N\} \leqslant \widetilde{h}, \\ \delta_{\min} \leqslant \delta_i \leqslant \delta_{\max}, \end{cases} \quad 1 \leqslant i \leqslant N, \tag{4.8}$$

there exists a bounded sequence $\{\nu_i\}_{i=0}^N \in \mathbb{R}^{Nm}$ for which the following inequalities hold:

$$\begin{cases} \|\nu_n\|_2 \leqslant K_1 \max_{n=0,\dots,N} \left\| \left(\vartheta_{n,1}^T \otimes I_m\right)\Delta z^{[0]} + \sum_{i=1}^{n} \left(\vartheta_{n,i+1}^T \otimes I_m\right)\Gamma(x_i) \right\|_2 \\ \qquad + K_2 h^{p+1} \leqslant \widetilde{K} h^p, \\ \left\|\Delta z^{[n]} - e_1 \otimes \nu_n\right\|_{S\otimes I_m} \leqslant \widetilde{K} h^{p+1}, \end{cases} \tag{4.9}$$

for $n = 1, 2, \dots, N$, where $h = \max\{h_n\colon\ n = 0, 1, \dots, N\}$ and where K_1, K_2 and $\widetilde{K}$ are some positive constants.

Proof. For $n \geqslant 0$, let

$$\nu_n = \left(\vartheta_{n,1}^T \otimes I_m\right)\Delta z^{[0]} + \sum_{i=1}^{n} \left(\vartheta_{n,i+1}^T \otimes I_m\right)\left(h_i (G \otimes I_m)\Delta F^{[i]} + \Gamma(x_i)\right),$$

and let $a_n = \|\Delta z^{[n]} - e_1 \otimes \nu_n\|_{S\otimes I_m}$, $b_n = \|\nu_n\|_2$. Then for any sequence $\{h_i\}_{i=0}^N$ satisfying (4.8) we obtain, using lemmas 7 and 9,

$$b_n \leqslant \Gamma_\Sigma + k_1 \sum_{i=0}^{n-1} h_{i+1}(a_i + b_i),$$

where

$$\Gamma_\Sigma = \max_{n=0,\dots,N} \left\| \left(\vartheta_{n,1}^T \otimes I_m\right)\Delta z^{[0]} + \sum_{i=1}^{n} \left(\vartheta_{n,i+1}^T \otimes I_m\right)\Gamma(x_i) \right\|_2 .$$

Similarly, we can bound a_n as follows:

$$a_n \leqslant k_2 h^{p+1} + k_3 \sum_{i=0}^{n-1} \mu^{n-1-i} h_{i+1}(a_i + b_i).$$

k_1, k_2 and k_3 are some positive constants depending on L, η, μ, Ω and $\|R^{1/2}G\|_2$. The rest of the proof follows by a standard induction on n. □

Remark. It can be noticed that, by definition, we have

$$\nu_{n+1} = \nu_n + \mathrm{O}\left(h_n^{p+1}\right).$$

This theorem shows that the main term of the global error (in h^p) lies in the direction of e_1. This information will be of interest in the next section for the estimation of local errors. It also shows that the main part of the error Γ_Σ is built up by the accumulation of local errors through the action of the ϑ's. With $\Gamma(x_i)$ of the form stated in theorem 6, Γ_Σ may be estimated as follows:

$$\begin{aligned}\Gamma_\Sigma &= \max_{n=0,\dots,N}\left\| \left(\vartheta_{n,1}^T \otimes I_m\right)\Delta z^{[0]} + \sum_{i=1}^{n}\left(\vartheta_{n,i+1}^T \otimes I_m\right)\left(h_i^{p+1}\varphi_p \otimes y^{(p+1)}(x_{i-1})\right)\right\|_2 \\ &\quad + \mathrm{O}\left(h^{p+1}\right) \\ &\leqslant \max_{n=0,\dots,N}\left(\sum_{i=1}^{n} h_i^{p+1}\left|\vartheta_{n,i+1}^T\varphi_p\right|\left\|y^{(p+1)}(x_{i-1})\right\|_2\right) + \mathrm{O}\left(h^{p+1}\right).\end{aligned}$$

In order to design a safe step-size-control strategy, one should consequently estimate the quantities $h_i^{p+1}|\vartheta_{n,i+1}^T\varphi_p| \cdot \|y^{(p+1)}(x_{i-1})\|_2$. In the case of constant step-size, this is simply the main part of $(q^T \otimes I_m)\Gamma(x_n)$.

We now investigate in detail the main case considered in this paper; that is, we assume that Q is of rank one. We then have $\widetilde{Q} = 0$ and $\vartheta_{n,i+1}^T = q^T D(\delta_i)$. Whenever the step is kept constant ($h_{i+1} = h_i$), the main contribution of the local error $\Gamma(x_i)$ to the global error is $(q^T \otimes I_m)\Gamma(x_i)$, as in the case of constant step-size sequences. Otherwise, it *deviates* from this quantity by a term of the form

$$\left(q^T\left(I - D(\delta_{i+1})\right) \otimes I_m\right)\Gamma(x_i),$$

which can be bounded by $h_i^{p+1} M \max_{x_0 \leqslant x \leqslant X} \|y^{(p+1)}(x)\|_2 + \mathrm{O}(h_i^{p+2})$, where

$$M = \max_{\delta_{\min} \leqslant \delta \leqslant \delta_{\max}} \left|q^T\varphi_p - q^T D(\delta)\varphi_p\right|.$$

Hence,

$$\max_{n=0,\dots,N}\left\| \left(q^T \otimes I_m\right)\Delta z^{[0]} + \sum_{i=1}^{n}\left(q^T \otimes I_m\right)\Gamma(x_i)\right\|_2$$

deviates from Γ_Σ by a term which can be bounded by

$$h^p M\left(\max_{x_0 \leqslant x \leqslant X}\left\|y^{(p+1)}(x)\right\|_2\right)\max_{n=0,\dots,N}\left(\sum_{i=1}^{n-1} h_i|1 - \delta_{i+1}|\right) + \mathrm{O}\left(h^{p+1}\right),$$

and thus remains small compared to Γ_Σ, under appropriate assumptions on the sequence of step-sizes and for small h. Note that if $q = e_1$, then we have exactly

$$\Gamma_\Sigma = \max_{n=0,\dots,N}\left\| \left(q^T \otimes I_m\right)\Delta z^{[0]} + \sum_{i=1}^{n}\left(q^T \otimes I_m\right)\Gamma(x_i)\right\|_2.$$

Similarly, if $\varphi_p = e_1$ or more generally $q(I - D(\delta))\varphi_p = 0$ for any δ, then

$$\Gamma_\Sigma = \max_{n=0,\dots,N}\left\| \left(q^T \otimes I_m\right)\Delta z^{[0]} + \sum_{i=1}^{n}\left(q^T \otimes I_m\right)\Gamma(x_i)\right\|_2 + \mathrm{O}\left(h^{p+1}\right).$$

Remark. Another case of special interest is $\rho(\widetilde{Q}) = 0$. The matrix $\widetilde{Q}$ is then nilpotent of index $s-1$. This means that whenever s consecutive steps have been kept equal the contribution of the local error $\Gamma(x_i)$ to Γ_Σ is once again of the form $(q^T \otimes I_m)\Gamma(x_i)$.

5. Local error estimation

According to the analysis undertaken in the previous section, a reliable control of the global error Γ_Σ can be achieved by bounding the local contributions $h_n^{p+1}(q^T\varphi_p)y^{(p+1)}(x_{n-1})$ at each step. In [8] we derived estimates of the local discretization error of DIMSIMs of the form (1.2) which were defined on uniform and nonuniform meshes. These estimates were expressed in terms of internal and external stages $Y^{[n]}$ and $y^{[n-1]}$ for uniform meshes and in terms of $Y^{[n]}$ and $\widehat{y}^{[n-1]}$ for nonuniform meshes, where $\widehat{y}^{[n-1]}$ are rescaled approximations corresponding to $y^{[n-1]}$. Similar estimates could be derived for methods (3.1). We have found, however, that in the case of DIMSIMs expressed in Nordsieck form it is more natural to work with estimates expressed in terms of internal approximations $Y^{[n]}$ and $Y^{[n-1]}$ computed at two consecutive steps. It is the purpose of this section to derive such error estimates for variable step-size method (3.1).

Theorem 11. Assume that the hypotheses of theorem 10 are satisfied and method (3.1) has stage order equal to the order p. Then the principal part of the local discretization error can be estimated by the formula

$$\begin{aligned} &h_n^{p+1}\big(q^T\varphi_p\big)y^{(p+1)}(x_{n-1}) \\ &\quad = h_n\big(\beta^T \otimes I_m\big)F\big(Y^{[n]}\big) + h_{n-1}\big(\gamma^T \otimes I_m\big)F\big(Y^{[n-1]}\big) + \mathrm{O}\big(h^{p+2}\big), \end{aligned} \tag{5.1}$$

where $h = \max_{i=0,\dots,n} h_i$ and where the vectors $\beta = \beta(\delta_n)$ and $\gamma = \gamma(\delta_n)$ satisfy the system of equations

$$\begin{cases} \beta^T C_p K D T + \gamma^T C_p K = 0, \\ \beta^T\left(C_p K D a + \delta_n^{p+1}\dfrac{c^p}{p!}\right) + \gamma^T\dfrac{c^p}{p!} = \delta_n^{p+1} q^T \varphi_p, \end{cases} \tag{5.2}$$

where T is defined by

$$T = \begin{bmatrix} 1 & 1 & \dfrac{1}{2!} & \dots & \dfrac{1}{p!} \\ 0 & 1 & 1 & \dots & \dfrac{1}{(p-1)!} \\ 0 & 0 & 1 & \dots & \dfrac{1}{(p-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}. \tag{5.3}$$

The matrices C_p, K and the vector a are defined in section 3 and the matrix $D = D(\delta_n)$ is defined in section 2.

Proof. Since method (3.1) has order p equal to the stage order q one has

$$Y^{[n]} = Y(x_{n-1}) + \mathrm{O}\big(h^p\big),$$

and more precisely

$$\begin{aligned} Y^{[n]} &= Y(x_{n-1}) + \big(PD(\delta_n) \otimes I_m\big)\Delta z^{[n-1]} + \mathrm{O}\big(h^{p+1}\big) \\ &= Y(x_{n-1}) + e \otimes \nu_{n-1} + \mathrm{O}\big(h^{p+1}\big), \end{aligned}$$

where ν_{n-1} is defined in theorem 10 and where we have used $PD(\delta_n)e_1 = Pe_1 = e$. Then

$$\begin{aligned} h_n F\big(Y^{[n]}\big) &= h_n Y'(x_{n-1}) + h_n\left(I_s \otimes \frac{\partial f}{\partial y}\big(y(x_{n-1})\big)\right)\big(Y^{[n]} - Y(x_{n-1})\big) + \mathrm{O}\big(h^{p+2}\big), \\ &= h_n Y'(x_{n-1}) + h_n\left(e \otimes \frac{\partial f}{\partial y}\big(y(x_{n-1})\big)\nu_{n-1}\right) + \mathrm{O}\big(h^{p+2}\big). \end{aligned}$$

Expanding $Y'(x_{n-1})$ into Taylor series around x_{n-1} we obtain

$$\begin{aligned} h_n F\big(Y^{[n]}\big) &= h_n\big(e \otimes y'(x_{n-1})\big) + h_n^2\big(c \otimes y''(x_{n-1})\big) + \cdots \\ &\quad + h_n^p\left(\frac{c^{p-1}}{(p-1)!} \otimes y^{(p)}(x_{n-1})\right) + h_n^{p+1}\left(\frac{c^p}{p!} \otimes y^{(p+1)}(x_{n-1})\right) \\ &\quad + h_n\left(e \otimes \frac{\partial f}{\partial y}\big(y(x_{n-1})\big)\nu_{n-1}\right) + \mathrm{O}\big(h^{p+2}\big), \end{aligned}$$

which can be written in a more compact form as

$$\begin{aligned} h_n F\big(Y^{[n]}\big) &= (C_p K \otimes I_m)\widehat{z}(x_{n-1}) + h_n^{p+1}\left(\frac{c^p}{p!} \otimes y^{(p+1)}(x_{n-1})\right) \\ &\quad + h_n\left(e \otimes \frac{\partial f}{\partial y}\big(y(x_{n-1})\big)\nu_{n-1}\right) + \mathrm{O}\big(h^{p+2}\big). \qquad (5.4) \end{aligned}$$

We have also

$$\widehat{z}(x_{n-1}) = (DT \otimes I_m)\widehat{z}(x_{n-2}) + h_{n-1}^{p+1}\big(Da \otimes y^{(p+1)}(x_{n-2})\big) + \mathrm{O}\big(h^{p+2}\big). \qquad (5.5)$$

Substituting (5.4) (and a similar relation with n replaced by $n-1$) and (5.5) into (5.1) and using the relations

$$y^{(p+1)}(x_{n-1}) = y^{(p+1)}(x_{n-2}) + \mathrm{O}(h_{n-1}),$$

$$\frac{\partial f}{\partial y}\big(y(x_{n-1})\big) = \frac{\partial f}{\partial y}\big(y(x_{n-2})\big) + \mathrm{O}(h)$$

and $\nu_{n-1} = \nu_{n-2} + \mathrm{O}(h^{p+1})$, we must satisfy the following relation:

$$\begin{aligned}
&\delta_n^{p+1} h_{n-1}^{p+1}\big(q^T\varphi_p\big) y^{(p+1)}(x_{n-2}) \\
&\quad = \big(\big(\beta^T C_p KDT + \gamma^T C_p K\big) \otimes I_m\big)\widehat{z}(x_{n-2}) \\
&\qquad + h_{n-1}^{p+1}\left(\beta^T C_p KDa + \delta_n^{p+1}\beta^T \frac{c^p}{p!} + \gamma^T \frac{c^p}{p!}\right) y^{(p+1)}(x_{n-2}) \\
&\qquad + \big(\delta_n \beta^T e + \gamma^T e\big)\frac{\partial f}{\partial y}\big(y(x_{n-2})\big)\nu_{n-2} + \mathrm{O}\big(h^{p+2}\big).
\end{aligned} \tag{5.6}$$

Comparing the corresponding terms in equation (5.6) and noticing that equation $\delta_n\beta^T e + \gamma^T e = 0$ is implied by $\beta^T C_p KDT + \gamma^T C_p K = 0$ leads to the system (5.2). This completes the proof. □

6. Examples of explicit and implicit DIMSIMs in Nordsieck form

Consider first DIMSIMs with $p = q = r = s = 3$, $c = [0, 1/2, 1]^T$, and coefficient matrices given by

$$\left[\begin{array}{c|c} A & U \\ \hline B & V \end{array}\right] = \left[\begin{array}{ccc|ccc}
0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 \\
\frac{1}{4} & 1 & 0 & 0 & 0 & 1 \\ \hline
\frac{5}{4} & \frac{1}{3} & \frac{1}{6} & -\frac{2}{3} & \frac{4}{3} & \frac{1}{3} \\
\frac{35}{24} & -\frac{1}{3} & \frac{1}{8} & -\frac{2}{3} & \frac{4}{3} & \frac{1}{3} \\
\frac{17}{12} & 0 & \frac{1}{12} & -\frac{2}{3} & \frac{4}{3} & \frac{1}{3}
\end{array}\right]. \tag{6.1}$$

This method, which was first derived in [6], is an example of type 1 DIMSIM (A is lower triangular with zero on the main diagonal) and is appropriate for non-stiff differential systems solved in a sequential computing environment. By construction, method (6.1) has the same stability properties as a 3-stage explicit Runge–Kutta method of order 3.

It can be verified using the results given in section 2 that the Nordsieck representation of (6.1) takes the form

$$\left[\begin{array}{c|c} A & P \\ \hline G & Q \end{array}\right] = \left[\begin{array}{ccc|cccc} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & -\frac{1}{2} & \frac{1}{8} & \frac{1}{48} \\ \frac{1}{4} & 1 & 0 & 1 & -\frac{1}{4} & 0 & \frac{1}{24} \\ \hline \frac{5}{4} & \frac{1}{3} & \frac{1}{6} & 1 & -\frac{3}{4} & \frac{1}{6} & \frac{1}{24} \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & -4 & 3 & 0 & 0 & 0 & 0 \\ 4 & -8 & 4 & 0 & 0 & 0 & 0 \end{array}\right]. \tag{6.2}$$

For this method the local discretization error $\Gamma(x_n)$ has the form given by formula (3.7) in theorem 6 with $p = 3$ and

$$\varphi_3 = \begin{bmatrix} \frac{1}{144} & 0 & \frac{1}{12} & \frac{1}{2} \end{bmatrix}^T.$$

Since $q^T = [1, -3/4, 1/6, 1/24]$ the principal part of $\Gamma(x_n)$ takes the form

$$\left(q^T \varphi_3\right) h_n^4 y^{(4)}(x_{n-1}) = \frac{1}{24} h_n^4 y^{(4)}(x_{n-1}).$$

Using theorem 11 the above expression can be estimated by

$$\frac{1}{24} h_n^4 y^4(x_{n-1}) = h_n\left(\beta^T \otimes I_m\right) F\left(Y^{[n]}\right) + h_{n-1}\left(\gamma^T \otimes I_m\right) F\left(Y^{[n-1]}\right) + \mathrm{O}\left(h_n^5\right), \tag{6.3}$$

where the vectors $\beta = [\beta_1, \beta_2, \beta_3]^T$ and $\gamma = [\gamma_1, \gamma_2, \gamma_3]^T$ can be computed from (5.2) with $p = 3$. This leads to the system of five equations in six unknowns and choosing β_3 as a free parameter we obtain

$$\begin{cases} \beta_1 = \dfrac{2\beta_3 + 9\beta_3\delta_n - 2\delta_n^2 + 7\beta_3\delta_n^2}{(1+\delta_n)(2+\delta_n)}, \\ \beta_2 = \dfrac{2(\delta_n^2 - 2\beta_3 - 6\beta_3\delta_n - 4\beta_3\delta_n^2)}{(1+\delta_n)(2+\delta_n)}, \\ \gamma_1 = \dfrac{\delta_n^3(\beta_3 - \delta_n + 2\beta_3\delta_n)}{2+\delta_n}, \\ \gamma_2 = \dfrac{2\delta_n^3(\delta_n - 2\beta_3 - 2\beta_3\delta_n)}{1+\delta_n}, \\ \gamma_3 = \dfrac{\delta_n^3(7\beta_3 - 3\delta_n + 9\beta_3\delta_n - \delta_n^2 + 2\beta_3\delta_n^2)}{(1+\delta_n)(2+\delta_n)}. \end{cases} \tag{6.4}$$

Consider next type 2 DIMSIMs (A is lower triangular with $\lambda \neq 0$ on the main diagonal) of the form

$$A = \begin{bmatrix} 0.43586652 & 0 & 0 \\ 0.25051488 & 0.43586652 & 0 \\ -1.2115943 & 1.0012746 & 0.43586652 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.83379073 & 0.64599891 & -0.31582709 \\ 0.60625754 & 1.2869318 & -0.47974168 \\ -0.30841677 & 3.8034216 & -1.1207225 \end{bmatrix}, \tag{6.5}$$

$$v = [\,0.55209096 \quad 0.73485666 \quad -0.28694762\,]^T.$$

This method was derived in [6] and is appropriate for stiff differential systems in a sequential computing environment. By construction (6.5) has the same stability properties as a 3-stage SDIRK method of order 3. As a consequence, this method is A-stable and L-stable (compare [10]).

The Nordsieck representation of (6.5) takes the form

$$A = \begin{bmatrix} 0.43586652 & 0 & 0 \\ 0.25051488 & 0.43586652 & 0 \\ -1.2115943 & 1.0012746 & 0.43586652 \end{bmatrix},$$

$$P = \begin{bmatrix} 1 & -0.43586652 & 0 & 0 \\ 1 & -0.18638114 & -0.09293326 & -0.03364998 \\ 1 & 0.77445317 & -0.43650382 & -0.17642592 \end{bmatrix},$$

$$G = \begin{bmatrix} 0.83379073 & 0.64599891 & 0.12003944 \\ 0 & 0 & 1 \\ 1 & -4 & 3 \\ 4 & -8 & 4 \end{bmatrix}, \tag{6.6}$$

$$q = [\,1 \quad -0.59982908 \quad 0.05696111 \quad 0.02589708\,]^T.$$

The local discretization error $\Gamma(x_n)$ of this method has the form (3.7) with $p = 3$ and

$$\varphi_3 = \begin{bmatrix} 0.00820178 & 0 & \frac{1}{12} & \frac{1}{2} \end{bmatrix}^T.$$

The principal part of $\Gamma(x_n)$ reads

$$\left(q^T \varphi_3\right) h_n^4 y^{(4)}(x_{n-1}) = 0.02589708 h_n^4 y^{(4)}(x_{n-1}).$$

Using theorem 11 the above expression can be estimated by

$$\begin{aligned} &0.02589708 h_n^4 y^4(x_{n-1}) \\ &\quad = h_n\left(\beta^T \otimes I_m\right) F\left(Y^{[n]}\right) + h_{n-1}\left(\gamma^T \otimes I_m\right) F\left(Y^{[n-1]}\right) + \mathrm{O}\left(h_n^5\right), \end{aligned} \tag{6.7}$$

with $\beta = [\beta_1, \beta_2, \beta_3]^T$ and $\gamma = [\gamma_1, \gamma_2, \gamma_3]^T$ given by

$$\begin{cases} \beta_1 = \dfrac{2\beta_3 + 9\beta_3\delta_n - 1.243060006\delta_n^2 + 7\beta_3\delta_n^2}{(1+\delta_n)(2+\delta_n)}, \\ \beta_2 = \dfrac{2(0.62153003\delta_n^2 - 2\beta_3 - 6\beta_3\delta_n - 4\beta_3\delta_n^2)}{(1+\delta_n)(2+\delta_n)}, \\ \gamma_1 = \dfrac{\delta_n^3(\beta_3 - 0.62153003\delta_n + 2\beta_3\delta_n)}{2+\delta_n}, \\ \gamma_2 = \dfrac{2\delta_n^3(0.62153003\delta_n - 2\beta_3 - 2\beta_3\delta_n)}{1+\delta_n}, \\ \gamma_3 = \dfrac{\delta_n^3(7\beta_3 - 1.86459009\delta_n + 9\beta_3\delta_n - 0.62153003\delta_n^2 + 2\beta_3\delta_n^2)}{(1+\delta_n)(2+\delta_n)}, \end{cases} \tag{6.8}$$

where β_3 is a free parameter. The quality of these error estimating formulas will be tested in the next section.

7. Numerical experiments

We have tested methods (6.2) and (6.6) on many non-stiff and stiff problems, respectively. We will present below the selection of results of numerical experiments which were designed to test the reliability of error estimating formulas (6.3) and (6.7) subject to rapid step changes. We will use the test equation

$$\begin{cases} y'(x) = \lambda(y - \mathrm{e}^{\mu x}) + \mu \mathrm{e}^{\mu x}, \quad x \in [x_0, X], \\ y(x_0) = y_0, \end{cases}$$

where λ and μ are real parameters and y_0 is a given initial value, with the exact solution

$$y(x) = \mathrm{e}^{\mu x} - \left(\mathrm{e}^{\mu x_0} - y_0\right)\mathrm{e}^{\lambda(x-x_0)}.$$

As in [8] the step-size pattern was chosen according to the formula

$$h_{n+1} = r(n)h_n$$

with $h_0 = (X - x_0)/N$, $N = 1000$, and

$$r(n) = \mathrm{RHO}^{(-1)^n \sin(4\pi n/(X-x_0))}$$

for RHO = 1.25, 1.5, 1.75 and 2. The results are displayed in figure 1 for method (6.2) and in figure 2 for method (6.6) for $\lambda = -0.1$ and $\mu = 0.1$. In these figures the inner curves correspond to the local error LE and the outer curves to the estimates EST given by (6.3) and (6.7) with β and γ defined by (6.4) and (6.8), respectively, and $\beta_3 = 1/6$.

As expected, as λ increases the explicit method (6.2) fails to integrate this problem successfully. This happens, for example, for constant step-size $h = 1/1000$ for

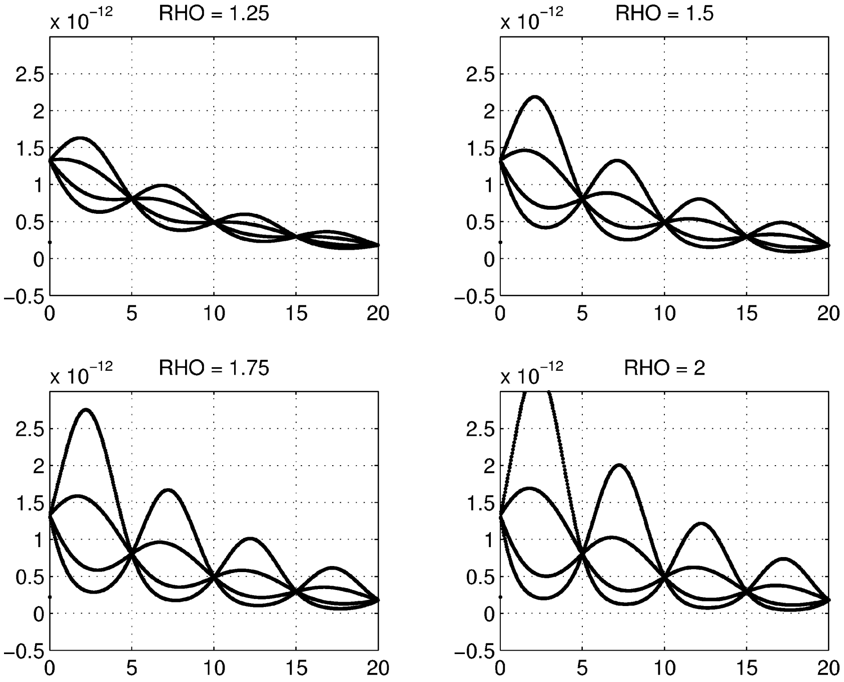

Figure 1. LE versus EST for DIMSIMs of type 1.

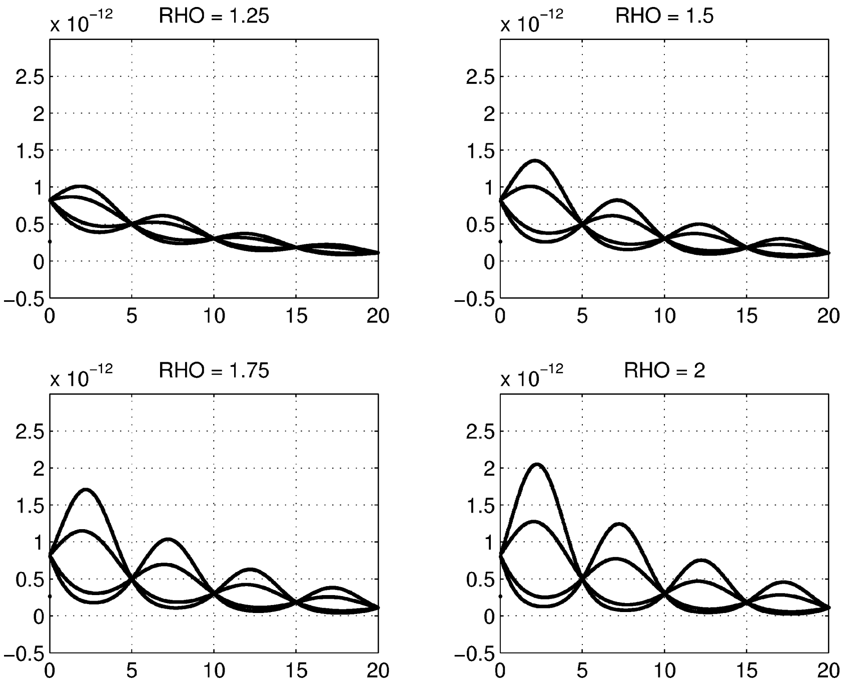

Figure 2. LE versus EST for DIMSIMs of type 2.

$\lambda < -251$ since the stability interval of this method is $(-2.51, 0)$. On the other hand, the implicit method (6.6) can handle this problem for a wide range of the parameter λ. However, the quality of error estimating formula (6.7) deteriorates rapidly as λ becomes larger. This is due to the fact that the higher order term in h_n^5 is becoming dominant and eventually greater than the principal part of the local discretization error

$$h_n^4 \big(q^T \varphi_3\big) y^{(4)}(x_{n-1}).$$

For scalar problems we could improve the quality of the error estimating formula by killing higher order terms. However, this is a very complicated task, even for methods of low orders, and which require different theoretical and numerical tools from that used in this paper. The situation is even more complicated in the vector case. These topics will be addressed in subsequent work.

8. Concluding remarks

Techniques similar to those discussed in this paper have been developed and implemented by Mrs Anjana Singh. Her work also considers a range of alternative error estimates and will be reported in a subsequent paper.

References

[1] J.C. Butcher, *The Numerical Analysis of Ordinary Differential Equations* (Wiley, Chichester, 1987).
[2] J.C. Butcher, Diagonally-implicit multi-stage integration methods, Appl. Numer. Math. 11 (1993) 347–363.
[3] J.C. Butcher, A transformation for the analysis of DIMSIMs, BIT 34 (1994) 25–32.
[4] J.C. Butcher and P. Chartier, The construction of DIMSIMs for stiff ODEs and DAEs, Report Series No. 308 (The University of Auckland, New-Zealand, July 1994).
[5] J.C. Butcher and Z. Jackiewicz, Diagonally implicit general linear methods for ordinary differential equations, BIT 33 (1993) 452–472.
[6] J.C. Butcher and Z. Jackiewicz, Construction of diagonally implicit general linear methods of type 1 and 2 for ordinary differential equations, Appl. Numer. Math. 21 (1996) 385–415.
[7] J.C. Butcher and Z. Jackiewicz, Construction of high order DIMSIMs for ordinary differential equations, submitted.
[8] J.C. Butcher and Z. Jackiewicz, Implementation of diagonally implicit multistage integration methods for ordinary differential equations, to appear in SIAM J. Numer. Anal.
[9] J.C. Butcher, Z. Jackiewicz and H.D. Mittelmann, Nonlinear optimization approach to the construction of general linear methods of high order, J. Comput. Appl. Math. 81 (1997) 181–196.
[10] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems* (Springer, Berlin, 1991).
[11] P. Lancaster and M. Tismenetsky, *The Theory of Matrices* (Academic Press, Orlando, 1985).