Error propagation of general linear methods for ordinary differential equations

J. C. Butcher [a,1] Z. Jackiewicz [b,2,*] W. M. Wright [c,3]

[a] *Department of Mathematics, The University of Auckland, Private Bag 92019, Auckland, New Zealand.*

[b] *Department of Mathematics, Arizona State University, Tempe, Arizona 85287, USA.*

[c] *Department of Mathematical and Statistical Sciences, La Trobe University, Melbourne, Victoria 3086, Australia.*

Abstract

We discuss error propagation for general linear methods for ordinary differential equations up to terms of order $p+2$, where p is the order of the method. These results are then applied to the estimation of local discretization errors for methods of order p and for the adjacent order $p+1$. The results of numerical experiments confirm the reliability of these estimates. This research has applications in the design of robust stepsize and order changing strategies for algorithms based on general linear methods.

Key words: General linear methods, Nordsieck representation, error propagation, local error estimation for methods of adjacent orders, adaptive stepsize selection, stability analysis
1991 MSC: 65L05, 65L06

* Corresponding author.
Email addresses: butcher@math.auckland.ac.nz (J. C. Butcher), jackiewi@math.la.asu.edu (Z. Jackiewicz), w.wright@latrobe.edu.au (W. M. Wright).
URLs: http://www.math.auckland.ac.nz/ butcher/ (J. C. Butcher), http://math.la.asu.edu/ jackiewi/zdzislaw.html (Z. Jackiewicz), http://www.latrobe.edu.au/mathstats/staff/wright/ (W. M. Wright).

[1] The research of this author was supported by the New Zealand Marsden Fund.
[2] The research of this author was supported by the National Science Foundation under grant DMS–0509597.
[3] The research of this author was supported by a New Zealand Science and Technology Postdoctoral Fellowship.

1 Introduction

This paper is concerned with the propagation of errors and the estimation of errors when a general linear method (GLM) is used for the numerical solution of an initial value problem

$$y'(x) = f(y(x)), \tag{1.1a}$$
$$y(x_0) = y_0, \tag{1.1b}$$

where $f : \mathbb{R}^m \to \mathbb{R}^m$ is given. We will consider general linear methods with s stages and with $r = p+1$ values passed from step to step, where p is the order of the method. We will also assume that the information output at the end of step n, with stepsize $h_n = h$ is supposed to approximate, to within $O(h^{p+1})$, the Nordsieck vector made up from subvectors $y(x_n)$, $hy'(x_n)$, $h^2y''(x_n)$, ..., $h^p y^{(p)}(x_n)$. The $p+1$ components of this output approximation will be written for convenience as y_n, for the first component, and $y^{[n]}$ for the remaining p components, with the full $p+1$ subvector output written as $\mathbf{y}^{[n]}$. That is,

$$\mathbf{y}^{[n]} = \begin{bmatrix} y_n \\ y^{[n]} \end{bmatrix}.$$

Taking the abscissa vector as $c = [c_1, \ldots, c_s]^T$, we can write the partitioned coefficient matrix in the form

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{U} \\ \hline \mathbf{B} & \mathbf{V} \end{array}\right] = \left[\begin{array}{c|cc} A & e & U \\ \hline b^T & 1 & v^T \\ B & 0 & V \end{array}\right], \tag{1.2}$$

where $A \in \mathbb{R}^{s\times s}$, $U \in \mathbb{R}^{s\times p}$, $B \in \mathbb{R}^{p\times s}$, $V \in \mathbb{R}^{p\times p}$, $b \in \mathbb{R}^s$ and $v \in \mathbb{R}^p$. With this terminology, the process of computing y_n and $y^{[n]}$ from input approximations y_{n-1} and $y^{[n-1]}$, takes the form

$$\begin{aligned} Y &= ey_{n-1} + AhF + Uy^{[n-1]}, \\ F_i &= f(Y_i), \qquad\qquad i = 1, 2, \ldots, s \\ y_n &= y_{n-1} + b^T hF + v^T y^{[n-1]}, \\ y^{[n]} &= BhF + Vy^{[n-1]}. \end{aligned} \tag{1.3}$$

We have chosen, for ease of notation, not to include the Kronecker products in the formulation of method or in the remainder of the paper. The vector $e = [1, \ldots, 1]^T \in \mathbb{R}^s$, the stepsize $h = x_n - x_{n-1}$, the numerical solutions y_{n-1}, y_n are approximations of order p to $y(x_{n-1})$ and $y(x_n)$; the internal stages Y_i and the stage derivatives F_i are approximations of stage order q to

$y(x_{n-1}+c_ih_n)$, and $y'(x_{n-1}+c_ih_n)$ for $i=1,2,\ldots,s$. Furthermore

$$Y=\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_s\end{bmatrix},\quad F:=\begin{bmatrix}F_1\\F_2\\\vdots\\F_s\end{bmatrix}.$$

Even though the stepsize will vary from step to step, it will sometimes be convenient to write simply h, rather than h_n, for the stepsize in step number n. It is also assumed that the vectors $y^{[n-1]}$ and $y^{[n]}$ are approximations of order p to the Nordsieck vectors $z(x_{n-1},h)$ and $z(x_n,h)$, respectively, where $z(x,h)$ is defined by

$$z(x,h):=\begin{bmatrix}hy'(x)\\h^2y''(x)\\\vdots\\h^py^{(p)}(x)\end{bmatrix}.$$

Although we are trying to approximate the Nordsieck vector at the end of step number n to within $O(h^{p+1})$, we need to obtain more detailed information about $y^{[n]}$. This is because we want to use components selected from the input data in each step, combined with the stage derivatives, to estimate $h^{p+1}y^{(p+1)}(x_n)$, $h^{p+2}y^{(p+2)}(x_n)$ and $h^{p+2}\frac{\partial f}{\partial y}y^{(p+1)}(x_n)$. These tasks have to be carried out in a variable stepsize environment but we will need to consider first the constant stepsize case. This will lead to an approximation

$$\begin{aligned}y_i^{[n]}&=h^iy^{(i)}(x_n)-\alpha_ih^{p+1}y^{(p+1)}(x_n)-\beta_ih^{p+2}y^{(p+2)}(x_n)\\&\quad-(\gamma_i+\epsilon\,\delta_i)h^{p+2}\frac{\partial f}{\partial y}y^{(p+1)}(x_n)+O(h^{p+3}),\end{aligned}\tag{1.4}$$

where $\delta_i=1$ if $i=1$ and $\delta_i=0$ if $i\neq 1$, and $\frac{\partial f}{\partial y}$ is interpreted to mean $\frac{\partial f}{\partial y}(y(x_n))$. In the remainder of the paper the argument will be clear from the context and is therefore omitted. When the stepsize is allowed to vary from step to step, our aim will be to apply a scale and modify procedure to the output from the step so that in the following step (1.4) can be regarded as being true. Further details of how this can be achieved will be discussed in Section 2.

Even though we have limited our scope to methods which pass a Nordsieck vector from step to step, we further refine the class of methods by requiring that the internal stage approximations are of the same order as the solution approximation. In this situation, the stage order and order conditions take the simple form described in the following theorem, compare, for example, [10].

Theorem 1 *A general linear method in Nordsieck form has order and stage*

order p if and only if

$$\begin{aligned} \varphi_1(cz) &= A\exp(cz) + UZ + O(z^p), \\ \varphi_1(z) &= b^T\exp(cz) + v^TZ + O(z^p), \\ \exp(z)Z &= B\exp(cz) + VZ + O(z^p), \end{aligned} \tag{1.5}$$

where z is a complex parameter and the basis vector $Z = [1, z, \ldots, z^{p-1}]^T$. The rational function $\varphi_1(z) = \frac{\exp(z)-1}{z}$ and both the exp *and φ_1 functions are applied component-wise to a vector.*

As a direct consequence of the above theorem the matrices U, v^T and V can be chosen so that the order and stage order are guaranteed to equal p. The equations (1.5) are equivalent to

$$\begin{aligned} U &= D - AC, \\ v^T &= P - b^TC, \\ V &= E - BC, \end{aligned} \tag{1.6}$$

where the Vandermonde matrices C and D are

$$C = \left[e\ c\ \tfrac{1}{2!}c^2\ \ldots\ \tfrac{1}{(p-1)!}c^{p-1}\right], \quad D = \left[c\ \tfrac{1}{2!}c^2\ \tfrac{1}{3!}c^3\ \ldots\ \tfrac{1}{p!}c^p\right],$$

the vector P and Toeplitz matrix E are

$$P = \left[1\ \tfrac{1}{2!}\ \tfrac{1}{3!}\ \ldots\ \tfrac{1}{p!}\right], \quad E = \exp(K), \quad K = \left[0\ e_1\ e_2\ \ldots\ e_{p-2}\right].$$

This can be seen by noting that $\varphi_1(cz) = DZ + O(z^p)$, $\varphi_1(z) = PZ + O(z^p)$, $\exp(cz) = CZ + O(z^p)$ and $\exp(z)Z = EZ + O(z^p)$.

The results of this paper will only apply to methods where the stage order is equal to the order and an approximation of $O(h^{p+1})$ to a Nordsieck vector is passed from step to step. Even with these restrictions in place, several well known classes of methods satisfy these criteria and we list some of them here. As proposed by Nordsieck in [20] an efficient implementation of linear multistep methods, reinterprets the information passed from step to step to approximate a Nordsieck vector. Since linear multistep methods can be represented as one stage general linear methods, with stage order equal to the order, these methods satisfy our criteria. As do composite linear multistep methods, where a method or a selection of methods are used over a series of smaller steps and interpreted as one step of a larger method. The Predict Evaluate Correct (PEC) or Predict Evaluate Correct Evaluate (PECE) or variants, see [19], can also be represented in this framework, provided that the data passed from step to step is transformed into Nordsieck form and the method has stage order equal to the order, the reader is referred to [2], for further details. The Nordsieck representation of DIMSIMs, was introduced in [3] (compare also

[4,16]) corresponds to the case when $s = p$, $V = 0$ and the stability function has only one non-zero eigenvalue. This representation was inspired by the classical paper [20]. Results concerning the construction and implementation of DIMSIMs are discussed in [11,17]. General linear methods with inherent Runge–Kutta stability (IRKS), investigated in [10,23], correspond to methods where $s = p + 1$, and the stability function has only one non-zero eigenvalue. These methods have many attractive properties (compare [6,7]) and their utilization as building blocks of powerful new algorithms for both nonstiff and stiff differential systems is a subject of recent work [8].

Assessing the relative efficiences of several alternative methods, especially the method already in operation in comparison with a contending method of one higher order, is central to the design of variable order algorithms. In the present paper, in Section 3, we show how to estimate both $h^{p+1}y^{(p+1)}$ and $h^{p+2}y^{(p+2)}$ thus allowing for a reliable assessment of the relative advantages of retaining order p or increasing the order to $p + 1$. The approach we use in this paper is not the only way of making this comparison dynamically, and we draw attention to a recent paper [9] which provides an alternative approach.

In Section 2 we will discuss starting methods and the underlying one-step method, providing the motivation for studying the error propagation of methods. Section 3 investigates error propagation, error estimation and the scale and modify process which ensures that (1.4) can be regarded as true, even when the stepsize is varied. In Section 4 a zero stability analysis is provided. A selection of methods of order 2 and 3 along with error estimates and regions of zero stability are included in Section 5. Several numerical experiments will be given in Section 6, which will validate the aims of this paper.

2 Starting methods and the underlying one step method

To understand the scale and modify procedure, that forms the basis of this paper, we focus attention on the relationship between y_n, the approximation to $y(x_n)$, and $y^{[n]}$, the vector made up from the remaining components of the output $\mathbf{y}^{[n]}$ at the end of step number n. We will assume that we are attempting to approximate an idealised quantity which we will write as $\mathcal{S}y_n$. This assumes that we have a suitable starting procedure $\mathcal{S}$ defined so that

$$\mathcal{S} \circ \mathcal{R} = \mathcal{M} \circ \mathcal{S}, \tag{2.1}$$

where $\mathcal{M}$ denotes the action of applying the method to the input data available at the end of the step and $\mathcal{R}$ is the underlying one-step method.

Actually it will be sufficient if (2.1) holds only to within $O(h^{p+3})$ accuracy because our aim will be to look for estimates of quantities which behave like

h^{p+1} and h^{p+2}. To actually calculate $\mathcal{S}y_{n-1}$ at the start of step n we need to have some way of calculating the Nordsieck vector, with appropriate modifications of the h^{p+1} and h^{p+2} terms. Let $y(x)$ denote the solution to (1.1a), restricted to the interval $[x_{n-1}, x_n) = [x_{n-1}, x_{n-1} + h)$, and with initial value given by $y(x_{n-1}) = y_{n-1}$. To actually evaluate $\mathcal{S}y_{n-1}$, we need to calculate the various high derivatives of $y(x)$ at $x = x_{n-1}$. In practice we will only want to work to within $O(h^{p+3})$ so that elementary differentials up to order $p + 2$ are needed, but not for a higher order.

Define $\mathcal{F}$ as the mapping which selects from $\mathbf{y}^{[n]}$ just the first subvector y_n. Thus, $\mathcal{F} \circ \mathcal{S} = \mathrm{id}$. In Section 3 we show that the components of $y^{[n]}$ produced by $\mathcal{S}y_n$ are given by

$$\begin{aligned} y_i^{[n]} &= h^i y^{(i)}(x_n) - \alpha_i h^{p+1} y^{(p+1)}(x_n) - \beta_i h^{p+2} y^{(p+2)}(x_n) \\ &\quad - (\gamma_i + \epsilon\, \delta_i) h^{p+2} \frac{\partial f}{\partial y} y^{(p+1)}(x_n) + O(h^{p+3}), \end{aligned}$$

for certain known vectors α, β and γ and constant ϵ. In Fig. 1, $\mathcal{R}$ is the underlying one-step method and T is the local truncation error.

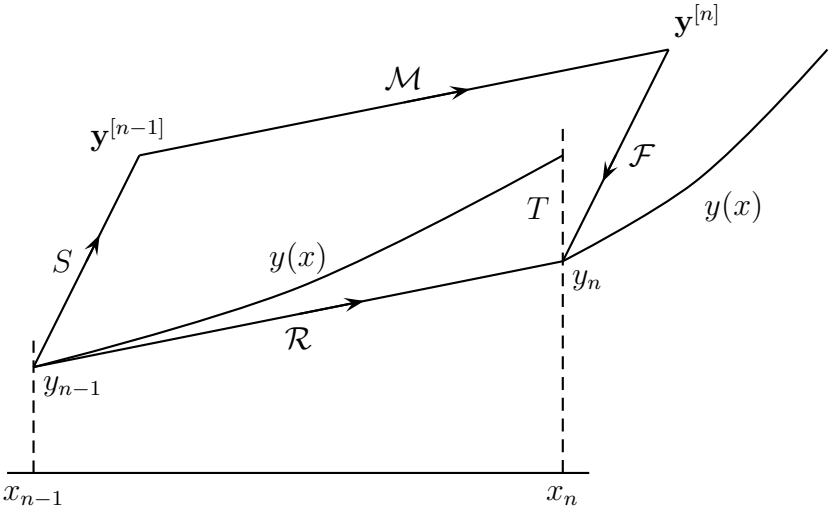

Fig. 1. A single step of the method $\mathcal{M}$, with starting method $\mathcal{S}$ and finishing method $\mathcal{F}$. The underlying one-step method $\mathcal{R}$ commutes to within $O(h^{p+3})$ and the local truncation error is represented by T.

3 Error propagation for GLMs

Our aim is to investigate the error propagation in a variable stepsize setting and to maintain the form of the approximate Nordsieck vector given by (1.4), given that the stepsize varies. It is not possible to achieve these aim with the current form of the method (1.3) unless the last equation is changed. The change we propose will amount to using the scale and modify process; which

when the stepsize is constant will reduce to the form given in (1.3). We will discuss later in this section the details of the scale and modify process. In the mean time, let the variable stepsize $h_n = x_n - x_{n-1}$ and the vectors $y^{[n-1]}$ and $y^{[n]}$ approximate to order p the Nordsieck vectors $z(x_{n-1}, h_n)$ and $z(x_n, h_{n+1})$. We will assume that the input quantities y_{n-1} and $y^{[n-1]}$ to the current step satisfy the relations

$$\begin{aligned} y_{n-1} &= y(x_{n-1}), \\ y^{[n-1]} &= z(x_{n-1}, h_n) - \alpha\, h_n^{p+1} y^{(p+1)}(x_{n-1}) - \beta\, h_n^{p+2} y^{(p+2)}(x_{n-1}) \\ &\quad - \gamma\, h_n^{p+2} \frac{\partial f}{\partial y} y^{(p+1)}(x_{n-1}) + O(h_n^{p+3}), \end{aligned} \tag{3.1}$$

where $y(x)$ is the solution to (1.1a) and α, β, and γ are constant vectors to be determined. We then compute an error constant ϵ and constants ρ and σ, such that the output quantities y_n and $y^{[n]}$ computed in the step satisfy

$$\begin{aligned} y_n &= y(x_n) - \epsilon\, h_n^{p+1} y^{(p+1)}(x_n) - \rho\, h_n^{p+2} y^{(p+2)}(x_n) \\ &\quad - \sigma\, h_n^{p+2} \frac{\partial f}{\partial y} y^{(p+1)}(x_n) + O(h_n^{p+3}), \\ y^{[n]} &= z(x_n, h_{n+1}) - \alpha\, h_{n+1}^{p+1} y^{(p+1)}(x_n) - \beta\, h_{n+1}^{p+2} y^{(p+2)}(x_n) \\ &\quad - (\gamma + \epsilon\, e_1)\, h_{n+1}^{p+2} \frac{\partial f}{\partial y} y^{(p+1)}(x_n) + O(h_n^{p+3}), \end{aligned} \tag{3.2}$$

where $e_1 = [1, 0, \ldots, 0]^T \in \mathbb{R}^p$ and α, β and γ are the same vectors given in (3.1). The constants ϵ, ρ and σ are the elementary weights of the local truncation error (represented by T in Fig. 1 the difference between the exact solution and the underlying one-step method) corresponding to the elementary differentials of order p+1 and p+2. Note that last term in $y^{[n]}$ has constant term $\gamma + \epsilon\, e_1$ rather than just γ. The extra contribution in the first component of the Nordsieck vector, comes from the solution approximation. Substituting the expression for the exact solution, the first equation of (3.2), into the differential equation (1.1a), gives

$$\begin{aligned} h_{n+1} y'(x_n) &= h_{n+1} f\left(y_n + \epsilon\, h_n^{p+1} y^{(p+1)}(x_n) + O(h_n^{p+2})\right) \\ &= h_{n+1} f(y_n) + \epsilon\, h_n^{p+2} \frac{\partial f}{\partial y} y^{(p+1)}(x_n) + O(h_n^{p+3}). \end{aligned}$$

For the remaining components of the approximate Nordsieck vector this contribution becomes $O(h_n^{p+3})$. Given that we are limiting our attention to general linear methods with high stage order, that is $p = q$, there exists a vector ξ of stage errors, such that

$$Y = y(x_{n-1} + ch_n) - \xi\, h_n^{p+1} y^{(p+1)}(x_{n-1}) + O(h_n^{p+2}), \tag{3.3}$$

which implies that the corresponding stage derivatives satisfy

$$h_n F = h_n y'(x_{n-1} + ch_n) - \xi\, h_n^{p+2} \frac{\partial f}{\partial y} y^{(p+1)}(x_{n-1}) + O(h_n^{p+3}). \tag{3.4}$$

Now that we have the explicit form of each of the quantities computed, in step number n, that is the quantities y_n, $y^{[n]}$, Y and $h_n F$, we are interested in expanding them, using Taylor series, about the point x_{n-1}. We first note that the Nordsieck vectors $z(x_n, h_{n+1})$ and $z(x_{n-1}, h_n)$ are connected by the following relation

$$\begin{aligned} z(x_n, h_{n+1}) = D(\delta) \Big(& Ez(x_{n-1}, h_n) + E_p h_n^{p+1} y^{(p+1)}(x_{n-1}) \\ & + E_{p+1} h_n^{p+2} y^{(p+2)}(x_{n-1}) \Big) + O(h_n^{p+3}), \end{aligned}$$

where $\delta = h_{n+1}/h_n$, $D(\delta) = \mathrm{diag}(\delta, \delta^2, \ldots, \delta^p)$, $E_p = [\frac{1}{p!}, \ldots, \frac{1}{2!}, 1]^T$ and $E_{p+1} = [\frac{1}{(p+1)!}, \ldots, \frac{1}{3!}, \frac{1}{2!}]^T$. Expanding using Taylor series leads to the expressions

$$\begin{aligned} Y &= y(x_{n-1}) + Cz(x_{n-1}, h_n) + \left(\frac{c^p}{p!} - \xi \right) h_n^{p+1} y^{(p+1)}(x_{n-1}) + O(h_n^{p+2}), \\ h_n F &= Cz(x_{n-1}, h_n) + \frac{c^p}{p!} h_n^{p+1} y^{(p+1)}(x_{n-1}) + \frac{c^{p+1}}{(p+1)!} h_n^{p+2} y^{(p+2)}(x_{n-1}) \\ &\quad - \xi h_n^{p+2} \frac{\partial f}{\partial y} y^{(p+1)}(x_{n-1}) + O(h_n^{p+3}), \\ y_n &= y(x_{n-1}) + Pz(x_{n-1}, h_n) + \left(\frac{1}{(p+1)!} - \epsilon \right) h_n^{p+1} y^{(p+1)}(x_{n-1}) \\ &\quad + \left(\frac{1}{(p+2)!} - \rho \right) h_n^{p+2} y^{(p+2)}(x_{n-1}) - \sigma\, h_n^{p+2} \frac{\partial f}{\partial y} y^{(p+1)}(x_n) + O(h_n^{p+3}), \\ y^{[n]} &= D(\delta) E z(x_{n-1}, h_n) + \Big(D(\delta) E_p - \delta^{p+1} \alpha \Big) h_n^{p+1} y^{(p+1)}(x_{n-1}) \\ &\quad + \Big(D(\delta) E_{p+1} - \delta^{p+1} \alpha - \delta^{p+2} \beta \Big) h_n^{p+2} y^{(p+2)}(x_{n-1}) \\ &\quad - \delta^{p+2} (\gamma + \epsilon\, e_1) h_n^{p+2} \frac{\partial f}{\partial y} y^{(p+1)}(x_{n-1}) + O(h_n^{p+3}). \end{aligned} \tag{3.5}$$

To determine the quantities ξ, ϵ, ρ, σ, α, β and γ, we substitute the Taylor series expanded quantities (3.5) into the method (1.3) with h replaced by h_n in the first two equations. The result is given in the following theorem.

Theorem 2 *In the variable stepsize case the stage error vector ξ and the error*

constants ϵ, ρ and σ are given by

$$\begin{aligned}
\xi &= \frac{c^{p+1}}{(p+1)!} - A\frac{c^p}{p!} + U\alpha, \\
\epsilon &= \frac{1}{\delta^{p+1}}\left(\frac{1}{(p+1)!} - b^T\frac{c^p}{p!} + v^T\alpha\right), \\
\rho &= \frac{1}{\delta^{p+2}}\left(\frac{1}{(p+2)!} - b^T\frac{c^{p+1}}{(p+1)!} + v^T\beta - \delta^{p+1}\epsilon\right), \\
\sigma &= \frac{1}{\delta^{p+2}}\left(b^T\xi + v^T\gamma\right),
\end{aligned} \tag{3.6}$$

where $\delta = h_{n+1}/h_n$. When fixed stepsizes are used the contributions from the higher order terms α, β and γ in the Nordsieck vector are

$$\begin{aligned}
\alpha &= (I-V)^{-1}\Big(E_p - B\frac{c^p}{p!}\Big), \\
\beta &= (I-V)^{-1}\Big(E_{p+1} - \alpha - B\frac{c^{p+1}}{(p+1)!}\Big), \\
\gamma &= (I-V)^{-1}\Big(B\xi - \epsilon\, e_1\Big).
\end{aligned} \tag{3.7}$$

Proof. To prove the expressions in (3.6), substitute the Taylor series expanded expressions for Y, h_nF and y_n given in (3.5) into the first two equations of (1.3), with h replaced by h_n. Collecting terms leads to expressions (3.6) and the order conditions (1.6). To prove the expressions in (3.7) let the stepsize h be constant in the last equation of the method, substitute the expanded expressions for hF and $y^{[n]}$ and collect terms.

The above theorem determines the value of the constant vectors α, β and γ, when constant stepsizes are used, and we want to maintain these values in a variable stepsize setting. As we stated at the beginning of this section, in the current form of the method, it is not possible to maintain these quantities. To do so we expand upon the scale and modify process, first used in [5]. In [5] the scale and modify process was used to ensure that α remained constant in a variable stepsize setting, but allowed β and γ to vary. The scale part of this process involves scaling the Nordsieck vector by the matrix $D(\delta)$, this takes into account the new stepsize. The modify part of the process involves adding the appropriate contributions from the higher order elementary differentials, both parts of the process need only be updated when the stepsize is changed. The scale and modify process of computing y_n and $y^{[n]}$ from the

approximations y_{n-1} and $y^{[n-1]}$, takes the form

$$\begin{aligned}
Y &= ey_{n-1} + Ah_nF + Uy^{[n-1]}, \\
F_i &= f(Y_i), \qquad\qquad i = 1, 2, \ldots, s, \\
y_n &= y_{n-1} + b^T h_n F + v^T y^{[n-1]}, \\
y^{[n]} &= \Big(D(\delta)B + \theta_1(\delta)\varphi_1^T + \theta_2(\delta)\varphi_2^T + \theta_3(\delta)\varphi_3^T\Big)h_nF \\
&\quad + \Big(D(\delta)V + \theta_1(\delta)\psi_1^T + \theta_2(\delta)\psi_2^T + \theta_3(\delta)\psi_3^T\Big)y^{[n-1]},
\end{aligned} \tag{3.8}$$

where the vectors $\theta_1(\delta), \theta_2(\delta), \theta_3(\delta) \in \mathbb{R}^p$ are chosen so that the vectors α, β and γ remain the constant vectors given in (3.7). The difference with the formulation presented in [5] is that in that paper $\theta_2(\delta)$ and $\theta_3(\delta)$ were each zero and only α was able to be kept constant. The vectors $\varphi_i, \psi_i \in \mathbb{R}^p$, for $i = 1, 2, 3$, are used to estimate the order $p+1$ and $p+2$ elementary differentials and the conditions that they need to satisfy to be correct to within $O(h_n^{p+3})$ are given in the following theorem.

Theorem 3 *If the function $y(x)$ is sufficiently differentiable in a neighbourhood of x, then*

$$\begin{aligned}
h_n^{p+1} y^{(p+1)}(x_n) &= \varphi_1^T h_n F + \psi_1^T y^{[n-1]} + O(h_n^{p+3}), \\
h_n^{p+2} y^{(p+2)}(x_n) &= \varphi_2^T h_n F + \psi_2^T y^{[n-1]} + O(h_n^{p+3}), \\
h_n^{p+2} \frac{\partial f}{\partial y} y^{(p+1)}(x_n) &= \varphi_3^T h_n F + \psi_3^T y^{[n-1]} + O(h_n^{p+3}),
\end{aligned} \tag{3.9}$$

where the vectors φ_i^T and ψ_i^T, for $i = 1, 2, 3$, satisfy the linear systems

$$\begin{array}{ccc}
\varphi_1^T C + \psi_1^T = 0, & \varphi_2^T C + \psi_2^T = 0, & \varphi_3^T C + \psi_3^T = 0, \\
\varphi_1^T \dfrac{c^p}{p!} - \psi_1^T \alpha = 1, & \varphi_2^T \dfrac{c^p}{p!} - \psi_2^T \alpha = 0, & \varphi_3^T \dfrac{c^p}{p!} - \psi_3^T \alpha = 0, \\
\varphi_1^T \dfrac{c^{p+1}}{(p+1)!} - \psi_1^T \beta = 1, & \varphi_2^T \dfrac{c^{p+1}}{(p+1)!} - \psi_2^T \beta = 1, & \varphi_3^T \dfrac{c^{p+1}}{(p+1)!} - \psi_3^T \beta = 0, \\
\varphi_1^T \xi + \psi_1^T \gamma = 0, & \varphi_2^T \xi + \psi_2^T \gamma = 0, & \varphi_3^T \xi + \psi_3^T \gamma = -1.
\end{array} \tag{3.10}$$

Proof. Substituting the expansions for h_nF and $y^{[n-1]}$ given respectively in equations (3.5) and (3.1) into (3.9) and collecting terms gives the required system of equations (3.10).

This system is uniquely determined for order two, but there is freedom available for orders greater than two. We are now faced with determining the exact form of the vectors $\theta_1(\delta)$, $\theta_2(\delta)$ and $\theta_3(\delta)$ so that the quantities α, β and γ remain unchanged even if the stepsize is changed from step to step. This leads to the following theorem.

Theorem 4 *To ensure that the constant vectors α, β and γ remain unchanged when variable stepsizes are used the vectors $\theta_1(\delta)$, $\theta_2(\delta)$ and $\theta_3(\delta)$ must be chosen so that*

$$\begin{aligned}\theta_1(\delta) &= \left(D(\delta) - \delta^{p+1} I\right) \alpha,\\ \theta_2(\delta) &= \left(D(\delta) - \delta^{p+2} I\right) \beta,\\ \theta_3(\delta) &= \left(D(\delta) - \delta^{p+2} I\right) (\gamma + \epsilon\, e_1).\end{aligned} \tag{3.11}$$

Proof. Substitute the expanded expressions for $y^{[n]}$, $h_n F$ and $y^{[n-1]}$, given respectively in equations (3.5) and (3.1), into the last equation of the scale and modify process (3.8) and collect terms. The coefficient of the $z(x_{n-1}, h_n)$ term is a combination of the order condition (1.6) and the first line of the system of equations (3.10) which determines the error estimators. Using the last three lines of the system of equations (3.10) the coefficients of $h_n^{p+1} y^{(p+1)}(x_{n-1})$, $h_n^{p+2} y^{(p+2)}(x_{n-1})$ and $h_n^{p+2} \frac{\partial f}{\partial y} y^{(p+1)}(x_{n-1})$ simplify respectively to

$$\begin{aligned} D(\delta)E_p - \delta^{p+1}\alpha &= D(\delta)B\frac{c^p}{p!} - D(\delta)V\alpha + \theta_1(\delta),\\ D(\delta)E_{p+1} - \delta^{p+1}\alpha - \delta^{p+2}\beta &= D(\delta)B\frac{c^{p+1}}{(p+1)!} - D(\delta)V\beta + \theta_1(\delta) + \theta_2(\delta),\\ -\delta^{p+2}\gamma &= -D(\delta)B\xi - D(\delta)V\gamma + \theta_3(\delta).\end{aligned}$$

Using (3.7) to simplify these equations leads to the desired result.

4 Zero-stability analysis

In this section we will analyze zero-stability properties of the scale and modify method given in equation (3.8). Applying this method to the test equation

$$y' = 0, \quad y(0) = 1,$$

on the nonuniform grid $\{x_n\}$ we obtain from (3.8) that

$$\begin{aligned} y_n &= y_{n-1} + v^T y^{[n-1]},\\ y^{[n]} &= \left(D(\delta_n)V + \theta_1(\delta_n)\psi_1^T + \theta_2(\delta_n)\psi_2^T + \theta_3(\delta_n)\psi_3^T\right) y^{[n-1]}.\end{aligned}$$

To simplify notation we define the amplification matrix $M(\delta)$ as

$$M(\delta) = D(\delta_n)V + \theta_1(\delta_n)\psi_1^T + \theta_2(\delta_n)\psi_2^T + \theta_3(\delta_n)\psi_3^T. \tag{4.1}$$

Expressing $y^{[n]}$ in terms of the initial starting vector $y^{[0]}$ leads to

$$y^{[n]} = M(\delta_n)M(\delta_{n-1})\cdots M(\delta_1)y^{[0]},$$

and the zero-stability of the method (3.8), is equivalent to the uniform boundedness of the product of matrices

$$M(\delta_n)M(\delta_{n-1})\cdots M(\delta_1).$$

We follow the approach proposed in [12,13] to find the conditions under which this is the case. This approach is based on the theory of the joint spectral radius and the notion of a polytope norm for a family of matrices. According to this theory, zero-stability of (3.8) would follow if we can construct a polytope norm $\|\cdot\|_*$ in $\mathbb{R}^p$, such that for the induced matrix norm, denoted by the same symbol, satisfies

$$\|M(\delta)\|_* \leq 1, \tag{4.2}$$

for $\delta \in [0, \delta^*]$. These polytope norms are defined by their unit balls in $\mathbb{R}^p$. Put

$$\delta^* = \max\Big\{\delta :\ \rho(M(\delta)) \leq 1\Big\},$$

where $\rho(M(\delta))$ is the spectral radius of the amplification matrix $M(\delta)$, given in (4.1). As explained in [12], often these polytope norms $\|\cdot\|_*$ can be found by successively applying the matrix $M(\delta^*)$ to the set of vectors

$$\mathcal{S} = \Big\{e_1, e_2, \ldots, e_p\Big\},$$

where e_i are canonical basis vectors in $\mathbb{R}^p$. If

$$M^j(\delta)P, \quad P \in \mathcal{S}, \quad j = 1, 2, \ldots,$$

are contained in a common convex hull, symmetric with respect to the origin, of some points in $\mathbb{R}^p$ for $\delta \in [0, \delta^*]$, then this convex hull defines the unit ball of the polytope norm $\|\cdot\|_*$ satisfying (4.2). This process was illustrated in [13] for variable stepsize 3-step backward differentiation method, in [5] for some GLMs of order $p = 2$, and in [18] for some two-step W-methods of order $p = 2$.

In Section 5, we give the amplification matrix $M(\delta)$ for various methods satisfying (3.8) and the convex hull which defines the unit ball. For methods presented in next section the matrices $M(\delta)$ will have the following block structure

$$M(\delta) = \begin{bmatrix} 0 & 0 \\ M_1(\delta) & M_2(\delta) \end{bmatrix},$$

with square blocks on the diagonal. It can be verified that

$$M(\delta_n)\cdots M(\delta_2)M(\delta_1) = \begin{bmatrix} 0 & 0 \\ M_2(\delta_n)\cdots M_2(\delta_2)M_1(\delta_1) & M_2(\delta_n)\cdots M_2(\delta_2)M_2(\delta_1) \end{bmatrix},$$

and it follows that if

$$\|M_2(\delta_n)\cdots M_2(\delta_2)M_2(\delta_1)\| \leq C,$$

for some constant C, then the nonzero blocks of the matrix

$$M(\delta_n)\cdots M(\delta_2)M(\delta_1),$$

can be bounded by $C\|M_1(\delta_1)\|$ and C, respectively. This means that zero-stability properties of the underlying numerical methods whose stability matrix is $M(\delta)$ are governed by the product of nonzero diagonal blocks

$$M_2(\delta_n)\cdots M_2(\delta_2)M_2(\delta_1).$$

We can take advantage of this fact to investigate stability properties of the methods for which the matrix $M(\delta)$ has the structure described above. This will be illustrated in Section 5.

5 Examples of methods

To achieve stage order and order p requires at least p stages, all the methods reported in this section have $s = p+1$ stages. This choice is prompted mainly because it is known that for this case IRKS methods can be derived using only linear operations [10]. For methods introduced to compete with IRKS methods, $s = p+1$ in every case, this simplifies a direct comparison of accuracy because the computational costs per step are the same. We have chosen an order two and an order three IRKS method and compared each with a PECE scheme, see [19], of the same order.

The first method, of order two, uses the composition of the order two Adams–Bashforth method over two steps of size $\frac{1}{2}h$ then uses a PECE scheme with the composite Adams–Bashforth method as the predictor and the order two Adams–Moulton method as the corrector. We then reinterpret the data passed from step to step to approximate a Nordsieck vector. The overall method with error estimates is

$$\left[\begin{array}{c|cc} A & e & U \\ \hline b^T & 1 & v^T \\ B & 0 & V \\ \hline \varphi_1^T & 0 & \psi_1^T \\ \varphi_2^T & 0 & \psi_2^T \\ \varphi_3^T & 0 & \psi_3^T \end{array}\right] = \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 1 & \frac{1}{2} & \frac{1}{8} \\ \frac{3}{4} & 0 & 0 & 1 & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{4} & 0 & 1 & \frac{1}{2} & \frac{1}{8} \\ \hline \frac{1}{4} & \frac{1}{4} & 0 & 1 & \frac{1}{2} & \frac{1}{8} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -2 & 0 & 2 & 0 & 0 & 0 \\ \hline -20 & 12 & -4 & 0 & 12 & 2 \\ -24 & 16 & -8 & 0 & 16 & 4 \\ 0 & -16 & 16 & 0 & 0 & 0 \end{array}\right]. \tag{5.1}$$

For this method the abscissae vector is $c = [\frac{1}{2}, 1, 1]^T$, the error constant is $\epsilon = \frac{1}{24}$ and the vectors α, β and γ given by formulae (3.7) are

$$\alpha = \begin{bmatrix} 0 \\ \frac{1}{4} \end{bmatrix}, \quad \beta = \begin{bmatrix} 0 \\ -\frac{1}{24} \end{bmatrix}, \quad \gamma = \begin{bmatrix} 0 \\ -\frac{1}{48} \end{bmatrix}.$$

The amplification matrix $M(\delta)$ given by formula (4.1) takes the form

$$M(\delta) = \begin{bmatrix} 0 & 0 \\ \frac{7}{3}\delta^2 - 3\delta^3 + \frac{2}{3}\delta^4 & \frac{1}{3}\delta^2 - \frac{1}{2}\delta^3 + \frac{1}{6}\delta^4 \end{bmatrix}.$$

Applying the procedure described in Section 4 it can be verified that the condition (4.2) is satisfied for $\delta \in [0, \delta^*]$, $\delta^* \approx 2.5747$, for the polytope norm $\|\cdot\|_*$ whose unit ball is a polytope with vertices P_1, P_2, P_3 and P_4 given by

$$P_1 = -P_3 = \begin{bmatrix} 1 & 0 \end{bmatrix}^T, \qquad P_2 = -P_4 = \begin{bmatrix} 0 & 6.4394 \end{bmatrix}^T.$$

The second method, which is of third order is similar to the first method except we use the third order Adams–Bashforth method over three steps of size $\frac{1}{3}h$ as the predictor and the order three Adams–Moulton method as the corrector. The method reinterpreted in Nordsieck form is

$$\left[\begin{array}{c|cc} A & e & U \\ \hline b^T & 1 & v^T \\ B & 0 & V \\ \hline \varphi_1^T & 0 & \psi_1^T \\ \varphi_2^T & 0 & \psi_2^T \\ \varphi_3^T & 0 & \psi_3^T \end{array}\right] = \left[\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 1 & \frac{1}{3} & \frac{1}{18} & \frac{1}{162} \\ \frac{23}{36} & 0 & 0 & 0 & 1 & \frac{1}{36} & \frac{1}{108} & \frac{1}{72} \\ \frac{7}{36} & \frac{23}{36} & 0 & 0 & 1 & \frac{1}{6} & \frac{1}{108} & \frac{1}{72} \\ \frac{11}{18} & \frac{2}{9} & \frac{5}{36} & 0 & 1 & \frac{1}{36} & \frac{1}{108} & \frac{1}{72} \\ \hline \frac{11}{18} & \frac{2}{9} & \frac{5}{36} & 0 & 1 & \frac{1}{36} & \frac{1}{108} & \frac{1}{72} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{3}{2} & -6 & 0 & \frac{9}{2} & 0 & 0 & 0 & 0 \\ 9 & -18 & 0 & 9 & 0 & 0 & 0 & 0 \\ \hline 9 & -\frac{171}{2} & \frac{243}{2} & -\frac{171}{2} & 0 & \frac{81}{2} & 18 & \frac{1}{2} \\ 36 & -99 & 135 & -99 & 0 & 27 & 18 & 2 \\ 0 & 0 & -\frac{972}{5} & \frac{972}{5} & 0 & 0 & 0 & 0 \end{array}\right]. \tag{5.2}$$

The abscissae vector is $c = [\frac{1}{3}, \frac{2}{3}, 1, 1]^T$, the error constant is $\epsilon = \frac{17}{1944}$ and the vectors α, β and γ are

$$\alpha = \begin{bmatrix} 0 \\ \frac{1}{27} \\ \frac{1}{3} \end{bmatrix}, \quad \beta = \begin{bmatrix} 0 \\ -\frac{1}{108} \\ -\frac{7}{108} \end{bmatrix}, \quad \gamma = \begin{bmatrix} 0 \\ -\frac{1}{108} \\ -\frac{5}{108} \end{bmatrix}.$$

The amplification matrix $M(\delta)$ takes the form

$$M(\delta) = \begin{bmatrix} 0 & 0 & 0 \\ \frac{5}{4}\delta^2 - \frac{3}{2}\delta^4 + \frac{1}{4}\delta^5 & \frac{1}{2}\delta^2 - \frac{2}{3}\delta^4 + \frac{1}{6}\delta^5 & -\frac{1}{54}\delta^4 + \frac{1}{54}\delta^5 \\ \frac{47}{4}\delta^3 - \frac{27}{2}\delta^4 + \frac{7}{4}\delta^5 & \frac{29}{6}\delta^3 - 6\delta^4 + \frac{7}{6}\delta^5 & \frac{1}{27}\delta^3 - \frac{1}{6}\delta^4 + \frac{7}{54}\delta^5 \end{bmatrix}.$$

The condition (4.2) is satisfied for $\delta \in [0, \delta^*]$, $\delta^* = 1.621033683$, for the polytope norm $\|\cdot\|_*$ whose unit ball in the three dimensional space (x, y, z) is a diamond-shaped region with vertices $[1, 0, 0]^T$ and $[-1, 0, 0]^T$ connected to the base in (y, z) plane with vertices P_1, P_2, P_3, P_4, P_5 and P_6. This base is plotted in Fig. 2. The (x, y, z) coordinates of the points P_i are

$$\begin{aligned} P_1 = -P_4 &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T, \\ P_2 = -P_5 &= \begin{bmatrix} 0 & 4.2136 & 22.4684 \end{bmatrix}^T, \\ P_3 = -P_6 &= \begin{bmatrix} 0 & 4.2741 & 23.5764 \end{bmatrix}^T. \end{aligned}$$

We can reach the same conclusion about zero-stability taking into account the special form of the matrix $M(\delta)$. For the matrix $M_2(\delta)$ given by the last two rows and columns of $M(\delta)$, it is found that $\rho(M_2(\delta)) \leq 1$ for $\delta \in [0, \delta^\star]$ where $\delta^\star$ is as above. The eigenvalues of $M_2(\delta^\star)$ are $\{-1, 0.0339637790\}$ and furthermore the eigenvector matrix T is given by

$$T = \begin{bmatrix} 0.0882162446 & 0.087063408 \\ 0.4709452399 & 1.598368850 \end{bmatrix}.$$

Form the matrix $T^{-1}M_2(\delta)T$, and evaluate its $\|\cdot\|_\infty$ norm. Then it is found that

$$\|T^{-1}M_2(\delta)T\|_\infty \leq 1, \quad \delta \in [0, \delta^\star].$$

The third example is a second order IRKS method. The free parameters have been chosen in such a way that the method is similar to the order two PECE scheme above. The method coefficients are

$$\left[\begin{array}{c|cc} A & e & U \\ \hline b^T & 1 & v^T \\ B & 0 & V \\ \hline \varphi_1^T & 0 & \psi_1^T \\ \varphi_2^T & 0 & \psi_2^T \\ \varphi_3^T & 0 & \psi_3^T \end{array}\right] = \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 1 & \frac{1}{2} & \frac{1}{8} \\ \frac{5}{3} & 0 & 0 & 1 & -\frac{2}{3} & -\frac{1}{3} \\ \frac{23}{24} & \frac{1}{8} & 0 & 1 & -\frac{1}{12} & -\frac{5}{48} \\ \hline \frac{23}{24} & \frac{1}{8} & 0 & 1 & -\frac{1}{12} & -\frac{5}{48} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -2 & -2 & 4 & 0 & 0 & 0 \\ \hline -20 & -17 & 25 & 0 & 12 & 2 \\ -24 & -20 & 28 & 0 & 16 & 4 \\ 0 & 12 & -12 & 0 & 0 & 0 \end{array}\right]. \tag{5.3}$$

This method has abscissae vector $c = [\frac{1}{2}, 1, 1]^T$, error constant $\epsilon = -\frac{1}{24}$. The vectors α, β and γ, the amplification matrix $M(\delta)$ and the polytope are the same as for the order two PECE method (5.1).

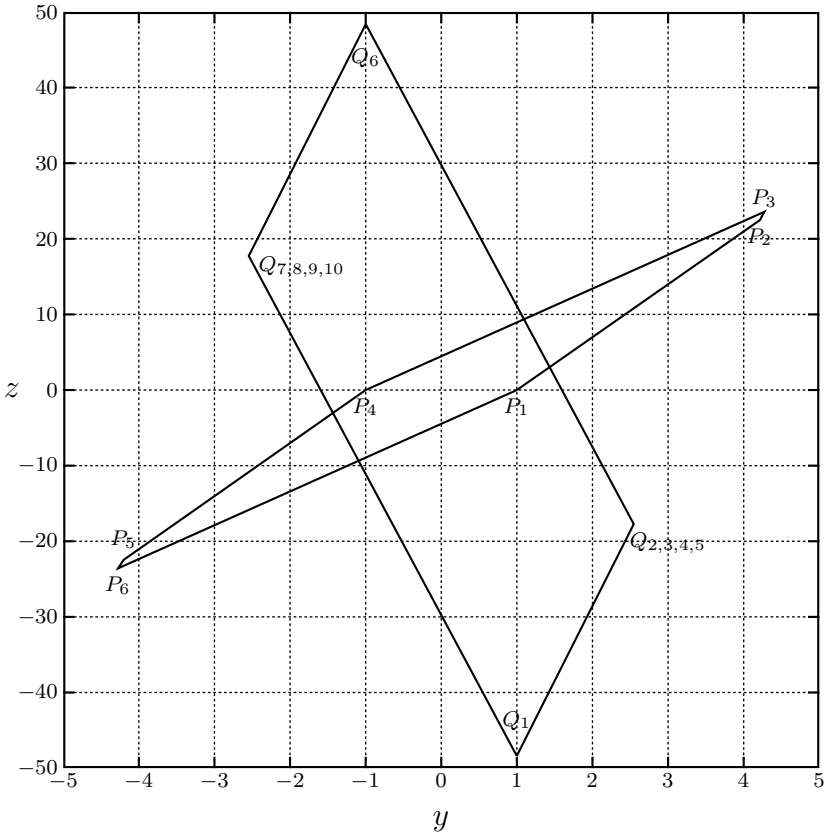

Fig. 2. Bases in (y, z)-plane of unit balls in the polytope norms for the order three PECE scheme (P_i vertices) and the order three IRKS method (Q_i vertices).

The last example is a third order IRKS method. The free parameters have been chosen in such a way that the method is similar to the order three PECE scheme above. The method coefficients are

$$\left[\begin{array}{c|cc} A & e & U \\ \hline b^T & 1 & v^T \\ B & 0 & V \\ \hline \varphi_1^T & 0 & \psi_1^T \\ \varphi_2^T & 0 & \psi_2^T \\ \varphi_3^T & 0 & \psi_3^T \end{array}\right] = \left[\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 1 & \frac{1}{3} & \frac{1}{18} & \frac{1}{162} \\ \frac{3}{5} & 0 & 0 & 0 & 1 & \frac{1}{15} & \frac{1}{45} & \frac{13}{810} \\ \frac{3}{7} & \frac{9}{14} & 0 & 0 & 1 & -\frac{1}{14} & -\frac{1}{14} & 0 \\ \frac{529}{810} & \frac{28}{81} & \frac{7}{81} & 0 & 1 & -\frac{23}{270} & -\frac{14}{405} & \frac{151}{14580} \\ \hline \frac{529}{810} & \frac{28}{81} & \frac{7}{81} & 0 & 1 & -\frac{23}{270} & -\frac{14}{405} & \frac{151}{14580} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{1}{6} & -\frac{31}{6} & -\frac{14}{3} & 9 & 0 & 1 & \frac{1}{6} & -\frac{1}{108} \\ -21 & -3 & -21 & 27 & 0 & 18 & 3 & -\frac{1}{6} \\ \hline 9 & -\frac{171}{2} & -171 & 207 & 0 & \frac{81}{2} & 18 & \frac{1}{2} \\ 36 & -99 & -180 & 216 & 0 & 27 & 18 & 2 \\ 0 & 0 & 270 & -270 & 0 & 0 & 0 & 0 \end{array}\right]. \qquad (5.4)$$

This method has abscissae vector $c = [\frac{1}{3}, \frac{2}{3}, 1, 1]^T$, error constant $\epsilon = \frac{1}{120}$ and the vectors α, β and γ are

$$\alpha = \begin{bmatrix} 0 \\ \frac{1}{27} \\ \frac{1}{3} \end{bmatrix}, \quad \beta = \begin{bmatrix} 0 \\ -\frac{1}{108} \\ -\frac{7}{108} \end{bmatrix}, \quad \gamma = \begin{bmatrix} 0 \\ -\frac{1}{324} \\ -\frac{1}{108} \end{bmatrix}.$$

The amplification matrix $M(\delta)$ takes the form

$$M(\delta) = \begin{bmatrix} 0 & 0 & 0 \\ \frac{9}{4}\delta^2 - \frac{3}{2}\delta^4 + \frac{1}{4}\delta^5 & \frac{2}{3}\delta^2 - \frac{2}{3}\delta^4 + \frac{1}{6}\delta^5 & -\frac{1}{108}\delta^2 - \frac{1}{54}\delta^4 + \frac{1}{54}\delta^5 \\ \frac{119}{4}\delta^3 - \frac{27}{2}\delta^4 + \frac{7}{4}\delta^5 & \frac{47}{6}\delta^3 - 6\delta^4 + \frac{7}{6}\delta^5 & -\frac{1}{6}\delta^4 - \frac{7}{54}\delta^3 + \frac{7}{54}\delta^5 \end{bmatrix}.$$

The condition (4.2) is satisfied for $\delta \in [0, \delta^*]$, $\delta^* = 1.547908766$, for the polytope norm $\|\cdot\|_*$ whose unit ball in the three dimensional space (x, y, z) is a diamond-shaped region with vertices $[1, 0, 0]^T$ and $[-1, 0, 0]^T$ connected to the base in (y, z) plane with vertices Q_1, Q_2, Q_3, Q_4, Q_5, Q_6, Q_7, Q_8, Q_9, and Q_{10}. This base is plotted in Fig. 2. The (x, y, z) coordinates of the points Q_i are

$$\begin{aligned} Q_1 = -Q_6 &= \begin{bmatrix} 0 & 0.9987 & -48.3857 \end{bmatrix}^T, \\ Q_2 = -Q_7 &= \begin{bmatrix} 0 & 2.4927 & -18.7892 \end{bmatrix}^T, \\ Q_3 = -Q_8 &= \begin{bmatrix} 0 & 2.5442 & -17.7671 \end{bmatrix}^T, \\ Q_4 = -Q_9 &= \begin{bmatrix} 0 & 2.5459 & -17.7313 \end{bmatrix}^T, \\ Q_4 = -Q_{10} &= \begin{bmatrix} 0 & 2.5459 & -17.7296 \end{bmatrix}^T. \end{aligned}$$

As before we can reach the same conclusion about zero-stability using the approach described at the end of Section 4. For the matrix $M_2(\delta)$ given by the last two rows and columns of $M(\delta)$, it is found that $\rho(M_2(\delta)) \leq 1$ for $\delta \in [0, \delta^\star]$ where $\delta^\star$ is as above. The eigenvalues of $M_2(\delta^\star)$ are $\{-1, -0.0345118943\}$ and furthermore the eigenvector matrix T is given by

$$T = \begin{bmatrix} 0.0820634390 & 0.0455181560 \\ -0.5714535837 & 0.9016010973 \end{bmatrix}.$$

Again, form the matrix $T^{-1}M_2(\delta)T$, and evaluate its $\|\cdot\|_\infty$ norm. Then it is found that

$$\|T^{-1}M_2(\delta)T\|_\infty \leq 1, \quad \delta \in [0, \delta^\star].$$

We compare these results to those in [7], where it was proved that methods, using a slightly modified form of (3.8), with $\theta_2(\delta) = \theta_3(\delta) = 0$ are zero stable for any choice of stepsize sequence. However, this is at the expense of losing the ability to estimate the higher order terms of the form $h_n^{p+2}y^{(p+2)}(x_n)$ and

$h_n^{p+2}\frac{\partial f}{\partial y}y^{(p+1)}(x_n)$. The construction of highly stable methods (possibly unconditionally stable) which also allow for the estimation of terms of order $p+2$ is the subject of ongoing work.

6 Numerical experiments

In this section, we test experimentally the reliability of the error estimates for $y^{(p+1)}$, $y^{(p+2)}$ and $\frac{\partial f}{\partial y}y^{(p+1)}$ using the approach of Section 3. At the same time we wish to assess the accuracy of low order derivative estimates y', y'', ..., found from outgoing Nordsieck vector approximations. We apply the tests to the methods with $p=2$ and $p=3$ derived in Section 5 in each case using the well-known van der Pol equation (denoted by E2 in the DETEST set [15])

$$\begin{aligned} y_1' &= y_2, & y_1(0) &= 2, \\ y_2' &= (1-y_1^2)y_2 - y_1, & y_2(0) &= 0, \end{aligned} \tag{6.1}$$

with integration interval $[0,8]$. For such a simple system as this, it is possible to find formulae for $y^{(i)}$. Write $y_i = y_1^{(i-1)} = y_2^{(i-2)}$, $i=3,4,\ldots$, so that successive derivatives of the vector valued function $y=[y_1,y_2]^T$ can be found as

$$y' = \begin{bmatrix} y_2 \\ y_3 \end{bmatrix}, \quad y'' = \begin{bmatrix} y_3 \\ y_4 \end{bmatrix}, \quad y^{(3)} = \begin{bmatrix} y_4 \\ y_5 \end{bmatrix}, \quad \ldots.$$

Formulae for y_3, y_4, ... are

$$\begin{aligned} y_3 &= (1-y_1^2)y_2 - y_1, \\ y_4 &= (1-y_1^2)y_3 - 2y_1y_2 - y_2, \\ y_5 &= (1-y_1^2)y_4 - 6y_1y_2y_3 - 2y_2^3 - y_3, \\ y_6 &= (1-y_1^2)y_5 - 8y_1y_2y_4 - 12y_2^3y_3 - 6y_1y_3^2 - y_4, \\ y_7 &= (1-y_1^2)y_6 - 10y_1y_2y_5 - 30y_2y_3^2 - 20y_2^2 - 20y_1y_3y_4 - y_5 \end{aligned}$$

and we also have available the Jacobian matrix

$$\frac{\partial f}{\partial y} = \begin{bmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2y_1y_2 - 1 & 1-y_1^2 \end{bmatrix}.$$

which is needed to evaluate the required values of $\frac{\partial f}{\partial y}y^{(p+1)}$, for $p=2,3$. The numerical experiments were performed in a variable stepsize environment using the standard step changing strategy based on the formula

$$h_{n+1} = h_n \left(\frac{\delta\,\mathrm{Tol}}{\|\mathrm{est}(p,x_n)\|} \right)^{\frac{1}{p+1}},$$

where $\delta = 0.5$, without any limiters or exceptions, compare for example [14,21,22]. Here, Tol is a given accuracy tolerance and $\mathrm{est}(p, x_n)$ is the estimate of the local discretization error, where ϵ is the error constant and the first equation in (3.9) is used to estimate $h_n^{p+1} y^{(p+1)}(x_n)$.

In Fig. 3, we have plotted the norms of the derivative expressions required to compare order 2 behaviour with that of contending methods of orders 1 and 3. This is compared with these quantities computed using the order two PECE scheme and the order two IRKS scheme. Both methods were given in Section 5. Apart from a slight phase shift in the $O(h^{p+2})$ approximations, the numerical estimations are found to be quite accurate.

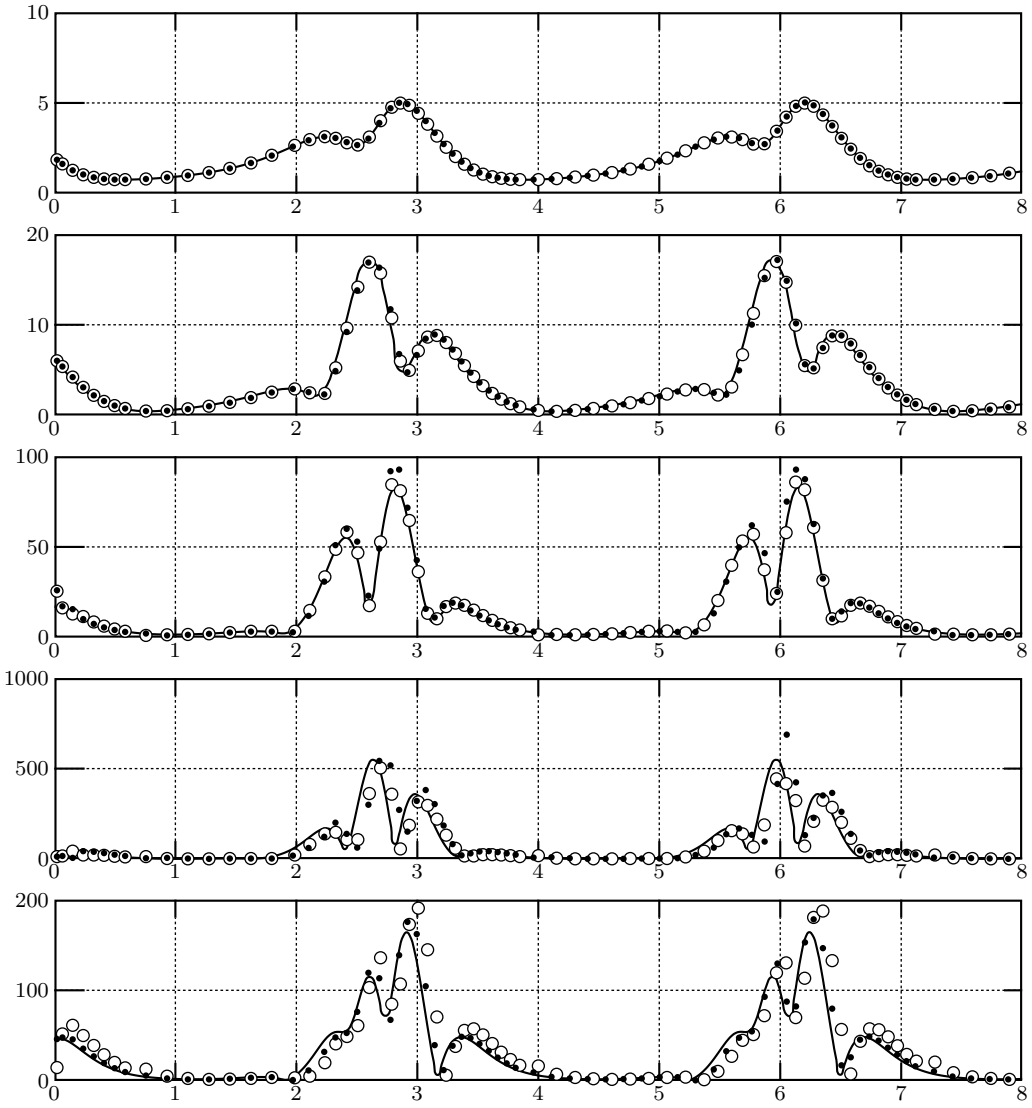

Fig. 3. Each of the derivative expressions (solid line) $y'(x)$, $y''(x)$, $y^{(3)}$, $y^{(4)}$ and $\frac{\partial f}{\partial y} y^{(3)}$, from top to bottom, plotted over the integration interval along with approximations to these quantities computed using order two methods: PECE scheme (•) and IRKS scheme (○). In each case the tolerance was 10^{-3}.

In Fig. 4 we have repeated the experiment reported in Fig. 4 but now using order 3 methods. Although a variable order solver would normally permit switching from order 3 to order 1, we have not included the values of $y'(x)$ in this figure, because the estimates are exact as they were for the order 2 methods in Fig. 3. As for the order 2 experiments, there is a phase shift in the $O(h^{p+2})$ approximations, but otherwise, the results confirm the ability to estimate the quantities we need in a variable order strategy.

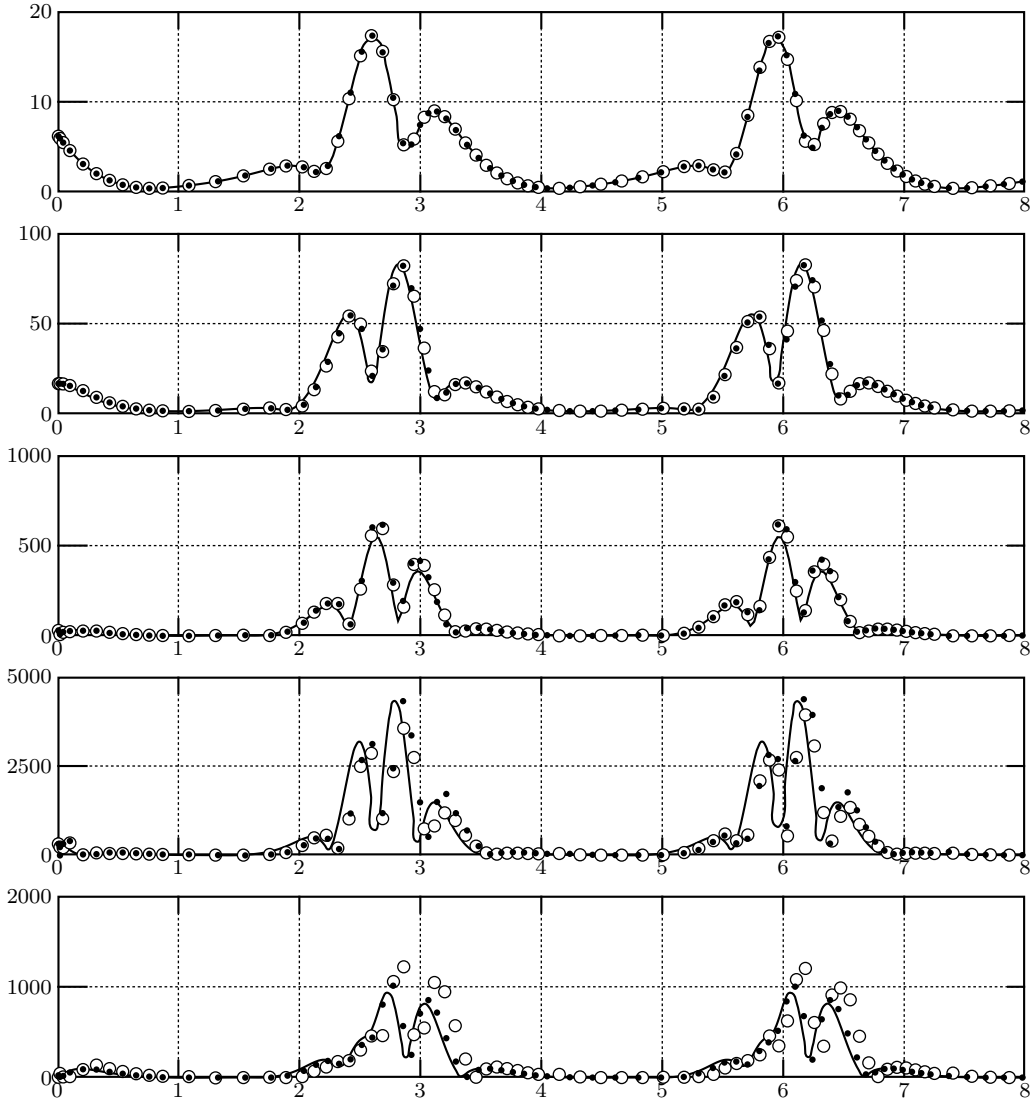

Fig. 4. Each of the derivative expressions (solid line) $y''(x)$, $y^{(3)}$, $y^{(4)}$, $y^{(5)}$ and $\frac{\partial f}{\partial y}y^{(4)}$, from top to bottom, plotted over the integration interval along with approximations to these quantities computed using order three methods: PECE scheme (•) and IRKS scheme (○). In each case the tolerance was 10^{-4}.

Examining the numerical experiments shows that the quality of the estimators for the higher order terms $h_n^{p+1}y^{(p+1)}(x_n)$, $h_n^{p+2}y^{(p+2)}(x_n)$ and $h_n^{p+2}\frac{\partial f}{\partial y}y^{(p+1)}(x_n)$ is reasonably good. The exact quality of these estimates depends very much on the method chosen. Future work will focus on determining how we can

identify which methods are the most suitable and use these methods in a variable stepsize, variable order environment.

7 Concluding remarks

In this paper we have considered the error propagation of a subclass of general linear methods, which have stage order equal to the order and passes a Nordsieck vector from step to step. This choice makes it possible to estimate the elementary differentials $h_n^{p+1}y^{(p+1)}(x_n)$, $h_n^{p+2}y^{(p+2)}(x_n)$ and $h_n^{p+2}\frac{\partial f}{\partial y}y^{(p+1)}(x_n)$ to within $O(h_n^{p+3})$. This makes available an asymptotically correct local error estimator and also asymptotically correct local error estimator of a method of one higher order. This information can be used by an algorithm to effectively choose the most efficient scheme from the methods of order $\{p-1, p, p+1\}$.

Acknowledgments

One of the authors (ZJ) acknowledges support from the Foundation under grant DMS–0509597. Another (WMW) was supported by the Royal Society of New Zealand. Each of the authors, (especially JCB) acknowledge assistance and support from the Marsden Fund.

References

[1] J.C. Butcher, Diagonally-implicit multi-stage integration methods, Appl. Numer. Math. 11(1993), 347–363.

[2] J.C. Butcher, The Numerical Solution of Ordinary Differential Equations, Wiley, 2003.

[3] J.C. Butcher, P. Chartier and Z. Jackiewicz, Nordsieck representation of DIMSIMs, Numerical Algorithms 16(1997), 209–230.

[4] J.C. Butcher, P. Chartier and Z. Jackiewicz, Experiments with a variable-order type 1 DIMSIM code, Numerical Algorithms 22(1999), 237–261.

[5] J.C. Butcher and Z. Jackiewicz, A new approach to error estimation for general linear methods, Numer. Math. 95(2003), 487–502.

[6] J.C. Butcher and Z. Jackiewicz, Construction of general linear methods with Runge–Kutta stability properties, Numerical Algorithms 36(2004), 53–72.

[7] J.C. Butcher and Z. Jackiewicz, Unconditionally stable general linear methods for ordinary differential equations, BIT 44(2004), 557–570.

[8] J.C. Butcher and Z. Jackiewicz, Towards a code for nonstiff differential systems based on general linear methods with inherent Runge–Kutta stability, in preparation.

[9] J.C. Butcher and H. Podhaisky, On error estimation in general linear methods for stiff ODEs, Appl. Numer. Math. 56(2006), 345–357.

[10] J.C. Butcher and W.M. Wright, The construction of practical general linear methods, BIT 43(2003), 695–721.

[11] P. Chartier, The potential of parallel multi-value methods for the simulation of large real-life problems, CWI Quarterly 11(1998), 7–32.

[12] N. Guglielmi and M. Zennaro, On the asymptotic properties of a family of matrices, Linear Algebra Appl. 322(2001), 169–192.

[13] N. Guglielmi and M. Zennaro, On the zero-stability of variable stepsize multistep methods: the spectral radius approach, Numer. Math. 88(2001), 445–458.

[14] E. Hairer, S.P. Nørsett and G. Wanner, Solving Ordinary Differential Equations I. Nonstiff Problems, Springer-Verlag, Berlin, Heidelberg, New York 1993.

[15] T.E. Hull, W.H. Enright, B.M. Fellen and A.E. Sedgwick, Comparing numerical methods for ordinary differential equations, SIAM J. Numer. Anal. 9(1972), 603–637.

[16] Z. Jackiewicz, Implementation of DIMSIMs for stiff differential systems, Appl. Numer. Math. 42(2002), 251–267.

[17] Z. Jackiewicz, Construction and implementation of general linear methods for ordinary differential equations. A review, J. Sci. Comput. 25(2005), 29–49.

[18] Z. Jackiewicz, H. Podhaisky and R. Weiner, Construction of highly stable two-step W-methods for ordinary differential equations, J. Comput. Appl. Math. 167(2004), 389–403.

[19] J.D. Lambert, Computational Methods in Ordinary Differential Equations, John Wiley & Sons, Chichester, New York 1973.

[20] A. Nordsieck, On numerical integration of ordinary differential equations, Math. Comp. 16(1962), 22–49.

[21] L.F. Shampine, Numerical Solution of Ordinary Differential Equations, Chapman & Hall, New York, London 1994.

[22] L.F. Shampine, I. Gladwell and S. Thompson, Solving ODEs with MATLAB, Cambridge University Press, Cambridge 2003.

[23] W.M. Wright, General linear methods with inherent Runge–Kutta stability, Ph.D. thesis, The University of Auckland, New Zealand, 2002.