Journal of Computational and Applied Mathematics 45 (1993) 227–236
North-Holland

CAM 1285

The tolerance proportionality of adaptive ODE solvers

Desmond J. Higham *

Department of Mathematics and Computer Science, University of Dundee, United Kingdom

Received 4 September 1991
Revised 8 December 1991

Abstract

Higham, D.J., The tolerance proportionality of adaptive ODE solvers, Journal of Computational and Applied Mathematics 45 (1993) 227–236.

Modern software for solving ordinary differential equation (ODE) initial-value problems requires the user to specify the ODE and choose a value for the error tolerance. The software can be thought of as a black box with a dial — turning the dial changes the accuracy and expense of the integration process. It is therefore of interest to know how the global error varies with the error tolerance. In this work, we look at explicit Runge–Kutta methods and show that with any standard error control method, and ignoring higher-order terms, the global error in the numerical solution behaves like a known rational power of the error tolerance. This generalises earlier work of Stetter, who found sufficient conditions for the global error to be linear in the tolerance. We also display the order of the next-highest term. We then analyse continuous Runge–Kutta schemes, and show what order of interpolation is necessary and sufficient for the continuous approximation to inherit the tolerance proportionality of the discrete formula. Finally we extend the results to the case of ODE systems with constant delays, thereby generalising some previous results of the author.

Keywords: Global error; interpolation; tolerance proportionality; delay ordinary differential equations.

1. Introduction

When using standard software to solve the ODE

$$y'(t)=f(t, y(t)), \quad y(t_0)=y_0 \in \mathbb{R}^N, \ t_0 \leqslant t, \tag{1.1}$$

a user will be asked to specify an error tolerance $\delta > 0$, which gives an indication of the level of accuracy required. A typical integrator will proceed from t_0, computing discrete approximations $y_n \approx y(t_n)$, for $n = 1, 2, \ldots$. The meshpoints t_n are chosen dynamically, and they depend upon the error tolerance. Usually, decreasing the error tolerance δ will cause the code to refine the mesh and hence to produce a more accurate solution. However, the user may want to know

Correspondence to: Dr. D.J. Higham, Department of Mathematics and Computer Science, University of Dundee, Dundee DD1 4HN, United Kingdom.
* Supported by the Research Initiatives Fund of the University of Dundee.

how much more accuracy is produced — if the tolerance is decreased by a factor of ten, how much does this change the global error?

This question was examined by Stetter [11,12]. Stetter looked for a linear relationship between global error and error tolerance (this is what a user might assume if the documentation did not say otherwise) and said that a method exhibits *tolerance proportionality* if, for any δ, there exists a piecewise C^1 interpolant $\eta(t)$ to the mesh data $\{t_n, y_n\}$ such that

$$\eta(t) - y(t) = v(t)\delta + g(t), \tag{1.2}$$

where $v(t)$ is C^1 and independent of δ, and $g(t)$ is piecewise C^1 with zeroth and first derivatives of $o(\delta)$. The condition (1.2) involves asymptotics and is relevant for sufficiently small tolerances. Note also that (1.2) guarantees asymptotically linear behaviour of $\eta'(t) - y'(t)$. Stetter showed that for a pth-order method, tolerance proportionality can be achieved by controlling any smoothly varying $O(h_i^p)$ quantity on each step. (Here h_i denotes the stepsize $t_i - t_{i-1}$.) We mention that the interpolant $\eta(t)$ for which (1.2) was shown to hold is not computable in general.

In [6] we examined the case where $\eta(t)$ is a *computable* Runge–Kutta interpolant (or continuous extension). Two classes of computable interpolants have been proposed in the literature — *higher*- and *lower*-order interpolants, whose local errors are $O(h_n^{p+1})$ and $O(h_n^p)$, respectively. It was shown that neither type of interpolant can satisfy (1.2), although higher-order interpolants satisfy the weaker condition

$$\eta(t) - y(t) = v(t)\delta + o(\delta). \tag{1.3}$$

We considered systems of ODEs with constant delays in [7], taking the usual approach of applying a Runge–Kutta method to a nearby standard ODE by using an interpolant to approximate the delayed values. We showed that the above results hold if discontinuities in the solution are crossed with sufficiently small steps and delay terms are computed with higher-order interpolants.

The purpose of this work is to examine a general error control mechanism and to show that a relationship like (1.2) will always hold with a known rational power of δ in the leading term. We also give a more precise quantification of the higher-order terms — if $\delta^{p/q}$ appears in the leading term, then the remaining terms are $O(\delta^{(p+1)/q})$. The extension also applies to the previous results on computable interpolants and constant-delay equations.

2. Discrete formulae

We begin this section with a generalisation of [6, Theorem 2.1]. Condition A below asks for the global error to be asymptotically proportional to δ^r, where r can be any positive real number, and for the higher-order terms to be $O(\delta^{r+\beta})$, where $r \geqslant \beta > 0$. [6, Theorem 2.1] was restricted to $r = 1$, and only asked for higher-order terms of $o(\delta)$. However, the proof of the new theorem is very similar to that of [6, Theorem 2.1], and both proofs are based on the ideas in [11].

Theorem 2.1. *Given the initial value problem $y'(t) - f(t, y(t)) = 0$, $y(t_0) = y_0$, suppose that to every tolerance value δ there corresponds a piecewiese C^1 approximation $\eta(t)$, satisfying $\eta(t_0) = y_0$.*

Let $\epsilon(t) := \eta(t) - y(t)$ denote the global error in $\eta(t)$, and let $r \geqslant \beta > 0$. Then, for sufficiently smooth f, the conditions **A** *and* **B** *below are equivalent.*

Condition **A**: $\epsilon(t) = v(t)\delta^r + g(t)$, *where $v(t)$ is C^1 and independent of δ, and $g(t)$ is piecewise C^1 with zeroth and first derivatives of* $O(\delta^{r+\beta})$.

Condition **B**: $\eta'(t) - f(t, \eta(t)) = \gamma(t)\delta^r + s(t)$, *where $\gamma(t)$ is continuous and independent of δ, and $s(t)$ is piecewise continuous and* $O(\delta^{r+\beta})$.

Proof. We introduce a third condition, **C**, and then prove that **A** $\Rightarrow$ **B**, **B** $\Rightarrow$ **C** and **C** $\Rightarrow$ **A**.

Condition **C**: $\epsilon'(t) - f_y(t, y(t))\epsilon(t) = \gamma(t)\delta^r + u(t)$, where $\gamma(t)$ is the function appearing in condition **B**, and $u(t)$ is piecewise continuous and $O(\delta^{r+\beta}) + O(\epsilon(t)^2)$.

(**A** $\Rightarrow$ **B**) We have

$$\begin{aligned}\eta'(t) - f(t, \eta(t)) &= y'(t) + \epsilon'(t) - f(t, y(t) + \epsilon(t))\\ &= \epsilon'(t) - f_y(t, y(t))\epsilon(t) + w(t),\end{aligned}$$

where $w(t) = O(\epsilon(t)^2)$, and hence, from **A**, $w(t) = O(\delta^{2r})$. Using **A** in this equation, we obtain

$$\begin{aligned}\eta'(t) - f(t, \eta(t)) = \delta^r\big[v'(t) - f_y(t, y(t))v(t)\big]&\\ + g'(t) - f_y(t, y(t))g(t) + w(t),&\end{aligned}$$

which has the required form.

(**B** $\Rightarrow$ **C**) Subtracting the original ODE $y'(t) - f(t, y(t)) = 0$ from **B** gives

$$\eta'(t) - y'(t) - \big(f(t, \eta(t)) - f(t, y(t))\big) = \gamma(t)\delta^r + s(t).$$

Using a Taylor expansion of $f(t, \eta(t)) = f(t, y(t) + \epsilon(t))$ this becomes

$$\epsilon'(t) - f_y(t, y(t))\epsilon(t) + \overline{w}(t) = \gamma(t)\delta^r + s(t),$$

where $\overline{w}(t)$ is piecewise continuous and $O(\epsilon(t)^2)$.

(**C** $\Rightarrow$ **A**) Let $v(t)$ denote the unique solution to the linear initial-value problem

$$v'(t) - f_y(t, y(t))v(t) = \gamma(t), \quad v(t_0) = 0.$$

Then, from **C**, $\epsilon(t) - \delta^r v(t)$ satisfies

$$\big(\epsilon(t) - \delta^r v(t)\big)' - f_y(t, y(t))\big(\epsilon(t) - \delta^r v(t)\big) = u(t), \quad \epsilon(t_0) - \delta^r v(t_0) = 0.$$

Standard theory (see, for example, [1, p.86]) shows that this linear, inhomogeneous, variable-coefficient initial-value problem has a solution of the form

$$\epsilon(t) - \delta^r v(t) = Y(t)\int_{t_0}^{t} Y^{-1}(\mu)u(\mu)\, \mathrm{d}\mu,$$

where the fundamental solution matrix $Y(t)$ is defined by

$$Y'(t) = f_y(t, y(t))Y(t), \quad Y(t_0) = I.$$

It follows that

$$\epsilon(t) - \delta^r v(t) = g(t),$$

where $g(t)$ is $O(\delta^{r+\beta})+O(\epsilon(t)^2)$ and continuous, and $g'(t)$ is $O(\delta^{r+\beta})+O(\epsilon(t)^2)$ and piecewise continuous, leading to the desired result. □

The usefulness of Theorem 2.1 lies in the fact that condition **B** is easier to verify than the equivalent condition **A**. Our aim is now to examine a very general class of error control schemes and show that there always exists an interpolant $\eta(t)$ that satisfies condition **B** for some r and β.

We will suppose that a pth-order Runge–Kutta method is used to advance the numerical solution, and we will let $z_n(t)$ denote the local solution over a step from t_{n-1} to $t_n := t_{n-1}+h_n$, so that $z_n'(t)=f(t, z_n(t))$ and $z_n(t_{n-1})=y_{n-1}$. We assume throughout that (1.1) is sufficiently smooth to allow the local error expansion

$$\mathrm{le}_n := y_n - z_n(t_n) = \psi(y_{n-1}, t_{n-1})h_n^{p+1} + O(h_n^{p+2}), \tag{2.1}$$

where the function ψ is C^1 and independent of h_n.

We regard the error control process as having two parts:

(1) an acceptance criterion;

(2) a stepsize changing formula.

For (1) an attempted step from t_{n-1} to t_n is deemed acceptable if $\mathrm{est}_n \leqslant \delta$. Here the computed quantity est_n is some measure of the error, and we will assume that it has the form

$$\mathrm{est}_n = \| e(y_{n-1}, t_{n-1}, h_n) \|, \tag{2.2}$$

where the expansion

$$e(y_{n-1}, t_{n-1}, h_n) = \hat{\psi}(y_{n-1}, t_{n-1})h_n^q + O(h_n^{q+1}) \tag{2.3}$$

holds, with $\hat{\psi}$ being a C^1 function, independent of h_n. Any vector norm is allowed in (2.2).

For (2) the next stepsize h_{n+1} is computed according to

$$h_{n+1} = \theta\left(\frac{\delta}{\mathrm{est}_n}\right)^{1/q} h_n, \tag{2.4}$$

where θ is a fixed safety factor in (0, 1). In the case of a rejected step, (2.4) could be used to give a stepsize with which to retry the step, or some other strategy such as halving the stepsize could be used. The particular strategy for retaking steps does not affect the analysis below, since we are concerned with asymptotic $\delta \to 0$ results and the presence of the safety factor ensures that for sufficiently small δ no rejections will occur.

We believe that the error control process outlined above encompasses virtually all widely used Runge–Kutta algorithms. The most common scenario is to advance from y_{n-1} with a subsidiary Runge–Kutta formula of order $\bar{p} \neq p$ to produce an additional approximation $\bar{y}_n$. The error estimate is then taken to be either $\| y_n - \bar{y}_n \|$ (error-per-step) or $\| y_n - \bar{y}_n \| / h_n$ (error-per-unit-step). If $\bar{p} < p$, then "local-extrapolation" is said to have taken place. Each of the four combinations of error-per-step or error-per-unit-step with local-extrapolation or without local-extrapolation has been used in practice [10], and it follows from the local error expansion (2.1) that (2.3) holds with $q = \min(p, \bar{p})+1$ for error-per-step control and with $q = \min(p, \bar{p})$ for error-per-unit-step control.

More recently a closely related alternative to local error control was proposed in [3]. Here a computable interpolant $w(t)$ is formed on each step and

$$\text{est}_n = \| w'(t^*) - f(t^*, w(t^*)) \|,$$

where $t^* = t_{n-1} + \tau^* h_n$, and τ^* is fixed in (0, 1). In this case est_n satisfies (2.2) and (2.3) with $q = p$ if $w(t)$ is a higher-order interpolant and $q = p - 1$ if $w(t)$ is a lower-order interpolant.

In order to prove our results, we must assume that $\hat{\psi}$ in (2.3) does not vanish, so that the error estimate always behaves like $O(h_n^q)$, and never like some higher power of h_n. It follows from this assumption that $\max\{h_n\} \to 0$ as $\delta \to 0$ and also that $\max\{h_n\} = O(\delta^{1/q})$.

After a successful step to t_n, using (2.4) to give the new stepsize, the error estimate on the next step can be expanded to give

$$\begin{aligned}
\| e(y_n, t_n, h_{n+1}) \| &= \left\| \hat{\psi}(y_n, t_n) \right\| h_{n+1}^q + O\left(h_{n+1}^{q+1}\right) \\
&= \left\| \hat{\psi}(y_{n-1} + O(h_n), t_{n-1} + O(h_n)) \right\| h_{n+1}^q + O\left(h_{n+1}^{q+1}\right) \\
&= \left\| \hat{\psi}(y_{n-1}, t_{n-1}) \right\| h_{n+1}^q + O(h_n h_{n+1}^q) + O\left(h_{n+1}^{q+1}\right) \\
&= \left\| \hat{\psi}(y_{n-1}, t_{n-1}) \right\| \theta^q \frac{\delta}{\text{est}_n} h_n^q + O(\delta^{(q+1)/q}).
\end{aligned}$$

Now, using (2.2) and (2.3), it follows that

$$\text{est}_{n+1} = \theta^q \delta + O(\delta^{(q+1)/q}). \tag{2.5}$$

This shows that on every step the error estimate will be asymptotically equal to θ^q times the tolerance δ. (Note that this also confirms that no step failures will arise for sufficiently small δ.)

We may express the local error on each step in terms of δ. Writing (2.1) as

$$\text{le}_n = \psi(y_{n-1}, t_{n-1}) h_n h_n^p + O\left(h_n^{p+2}\right),$$

and using, from (2.2) and (2.3),

$$h_n^p = \left\{ \frac{\text{est}_n}{\left\| \hat{\psi}(y_{n-1}, t_{n-1}) \right\|} + O\left(h_n^{q+1}\right) \right\}^{p/q},$$

we conclude that

$$\text{le}_n = \psi(y_{n-1}, t_{n-1}) h_n \frac{\text{est}_n^{p/q}}{\left\| \hat{\psi}(y_{n-1}, t_{n-1}) \right\|^{p/q}} + O(\delta^{(p+2)/q}).$$

Now if (2.5) holds on every step, then we find that

$$\text{le}_n = \psi(y_{n-1}, t_{n-1}) h_n \frac{\left\{ \theta^q \delta + O(\delta^{(q+1)/q}) \right\}^{p/q}}{\left\| \hat{\psi}(y_{n-1}, t_{n-1}) \right\|^{p/q}} + O(\delta^{(p+2)/q}),$$

which simplifies to

$$\text{le}_n = \frac{\psi(y_{n-1}, t_{n-1})}{\left\| \hat{\psi}(y_{n-1}, t_{n-1}) \right\|^{p/q}} h_n \theta^p \delta^{p/q} + O(\delta^{(p+2)/q}). \tag{2.6}$$

We are now in a position to show that condition **B** of Theorem 2.1 holds for a particular interpolant, known as the "ideal interpolant", which is defined by

$$\eta_1(t) := z_n(t) + \frac{(t - t_{n-1})}{h_n} \mathrm{le}_n, \quad t \in (t_{n-1}, t_n]. \tag{2.7}$$

Theorem 2.2. *Suppose that a pth-order Runge–Kutta method is used to solve* (1.1), *with an error control process of the form described above. Suppose further that*

- *the problem is sufficiently smooth for the expansions* (2.1) *and* (2.3) *to hold;*
- *for all sufficiently small* δ, $\| \hat{\psi}(y_n, t_n) \|$ *in* (2.3) *never vanishes;*
- *the initial stepsize is chosen so that* (2.5) *holds for* $n = 0$.

Then condition **B** *holds for the ideal interpolant defined by* (2.7) *with* $r = p/q$ *and* $\beta = 1/q$.

Proof. We have, from (2.7), for $t \in (t_{n-1}, t_n]$,

$$\eta_1'(t) - f(t, \eta_1(t)) = z_n'(t) + \frac{\mathrm{le}_n}{h_n} - f(t, z_n(t)) + \mathrm{O}(\mathrm{le}_n) = \frac{\mathrm{le}_n}{h_n} + \mathrm{O}(\mathrm{le}_n).$$

Now under the assumptions of the theorem, we have shown that the local error at t_n satisfies (2.6), and hence,

$$\eta_1'(t) - f(t, \eta_1(t)) = \frac{\psi(y_{n-1}, t_{n-1})}{\| \hat{\psi}(y_{n-1}, t_{n-1}) \|^{p/q}} \theta^p \delta^{p/q} + \mathrm{O}(\delta^{(p+1)/q}). \tag{2.8}$$

Now $t_n - t_{n-1} = \mathrm{O}(\delta^{1/q})$ and, from standard convergence theory (see, for example, [4, Theorem 3.4]), $y_{n-1} - y(t_{n-1}) = \mathrm{O}(\max\{h_n\}) = \mathrm{O}(\delta^{1/q})$. Hence we may replace t_{n-1} by t and y_{n-1} by $y(t)$ in (2.8) to give

$$\eta_1'(t) - f(t, \eta_1(t)) = \frac{\psi(y(t), t)}{\| \hat{\psi}(y(t), t) \|^{p/q}} \theta^p \delta^{p/q} + \mathrm{O}(\delta^{(p+1)/q}), \tag{2.9}$$

and hence condition **B** holds with $r = p/q$ and $\beta = 1/q$. □

We mention that the result remains valid when the vector norm in (2.2) involves componentwise absolute and relative weights. In this case the norm $\| \cdot \|$ depends upon δ, but this does not present a serious difficulty; see [6] for details.

Experiments in 386-Matlab [8], based on the built-in `ode23.m` program, have verified Theorem 2.2. Here, due to space limitations, we present a subset of our test results. We give results for the combinations $p = 2$, $q = 5$ and $p = 5$, $q = 2$. Here the fourth- and fifth-order pair HIHA5 from [5, Table 2.1] was used to provide the higher-order approximations. (We mention that we have chosen rather extreme cases of $p > q$ and $p < q$ in order to test the theory fully; we do not advocate the use of such values in practice.) We give results for the logistic equation

$$y' = \tfrac{1}{4} y \left(1 - \tfrac{1}{20} y\right), \quad y(0) = 1, \ 0 \leqslant t \leqslant 20,$$

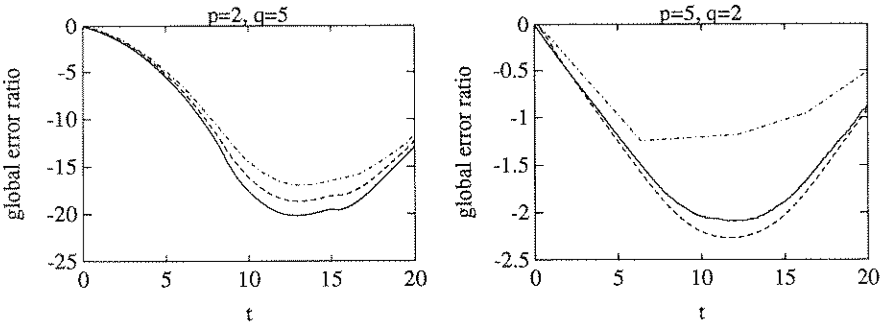

Fig. 1. Global error ratios on logistic equation for three different tolerance values.

which has solution

$$y = \frac{20}{1 + 19\exp\left(-\frac{1}{4}t\right)}.$$

In both cases we recorded the *global error ratio*, $[y_n - y(t_n)]/\delta^{p/q}$, at each meshpoint. If condition A is satisfied with $r = p/q$ and $\beta = 1/q$, then these values should be close to the limiting values $v(t_n)$ for small δ. For the $p < q$ test we used tolerances of $\delta = 10^{-6}$, 10^{-8}, 10^{-10}, while for the $p > q$ test we used $\delta = 10^{-1}$, 10^{-3}, 10^{-5}. (In the latter case, larger tolerances are more realistic, since a high-order formula is coupled with a low-order estimate.) Figure 1 presents the results. Here the discrete values are joined for clarity, and the curves change from dash-dotted to dashed to solid as the tolerance decreases. We see that the global error ratio appears to be settling down to a limit, as predicted by Theorem 2.2.

3. Interpolants

In general, the interpolant $\eta_I(t)$ in (2.7) is not computable, and hence Theorem 2.2 should be regarded as a result about the discrete numerical solution $\{t_n, y_n\}$. Recently a great deal of work has been done on the derivation of computable interpolants, or continuous Runge–Kutta formulae; see, for example, [9] and the references therein. Here, a continuously differentiable function $w(t)$ is made available, such that for any fixed $\tau \in [0, 1]$,

$$w(t_{n-1} + \tau h_n) - z_n(t_{n-1} + \tau h_n) = O(h_n^l). \tag{3.1}$$

If $l = p$, then the order of the local error in the interpolant is generally one less than the order of the local error in the discrete formula. In this case we have a "lower-order" interpolant. In the case where $l = p + 1$, the local errors in the interpolant and the discrete formula are of the same order, and we have a "higher-order" interpolant. Both higher- and lower-order interpolants have been proposed in the literature.

If we assume that condition A holds for the ideal interpolant, then we may split the global error in $w(t)$ into three parts:

$$w(t) - y(t) = [w(t) - z_n(t)] - [\eta_I(t) - z_n(t)] + [\eta_I(t) - y(t)]. \tag{3.2}$$

The first term, $w(t)-z_n(t)$, is $O(h_n^l)=O(\delta^{l/q})$, the second term, $\eta_I(t)-z_n(t)$, is $O(h_n^{p+1})=O(\delta^{(p+1)/q})$, and the third term, $\eta_1(t)-y(t)$, satisfies condition A with $r=p/q$ and $\beta=1/q$. Hence we find

$$w(t)-y(t)=v(t)\delta^{p/q}+O(\delta^{\min\{l,p+1\}/q}). \tag{3.3}$$

For a higher-order interpolant, we have $l=p+1$, so that

$$w(t)-y(t)=v(t)\delta^{p/q}+O(\delta^{(p+1)/q}). \tag{3.4}$$

This shows that the tolerance proportionality in $\eta_1(t)$ is inherited by $w(t)$. However, for a lower-order interpolant, the $O(\delta^{p/q})$ local error term is of the same asymptotic order as the first term in (3.3) and, as discussed in [6], does not behave smoothly as the tolerance varies. (In fact, $w(t)-z_n(t)$ has a smooth expansion in τ, where $\tau=(t-t_{n-1})/h_n$, but as δ decreases for fixed t, τ behaves in a sawtooth manner.)

Differentiating (3.2) gives a splitting for the global error in the first derivative approximation. Here it can be shown that $w'(t)-z_n'(t)=O(h_n^{l-1})$ for standard computable interpolants. It follows that for a higher-order interpolant this local error term contaminates the $\eta_1'(t)-y'(t)$ term, and we cannot extend (3.4) to first derivatives. Note, however, that if (3.4) holds, then

$$f(t, w(t))-y'(t)=f_y(t, y(t))v(t)\delta^{p/q}+O(\delta^{(p+1)/q}),$$

and hence $f(t, w(t))$ is a computable approximation to $y'(t)$ with a global error that is asymptotically proportional to $\delta^{p/q}$.

For lower-order interpolants, if we differentiate (3.2), then the local error term actually dominates in the expansion, and hence the ratio $[w'(t)-y'(t)]/\delta^{p/q}$ will not even remain bounded as $\delta\to 0$.

These results generalise those given in [6], and numerical results that illustrate the effect of the $w(t)-z_n(t)$ and $w'(t)-z_n'(t)$ terms can be found there.

4. Constant delays

We now consider a system of ODEs with k constant delays, which we write as

$$\begin{aligned} &y'(t)=F(t, y(t), y(t-\tau_1), y(t-\tau_2),\ldots, y(t-\tau_k))\in\mathbb{R}^N, \quad t\geqslant 0,\\ &y(t)=\Phi(t), \quad t\in[-\tau_k, 0], \ 0<\tau_1<\tau_2<\cdots<\tau_k. \end{aligned} \tag{4.1}$$

We will assume that F and Φ are smooth functions, but we will not assume that $\Phi'(0)=F(0, \Phi(0), \Phi(-\tau_1), \Phi(-\tau_2),\ldots,\Phi(-\tau_k))$. Hence, in general, the solution has a first derivative jump at $t=0$, and this is propagated into higher-order derivative discontinuities at later points. The points where $y^{(i)}(t)$ is discontinuous for $2\leqslant i\leqslant p+1$ can be determined a priori, and we will label them $\{\hat{t}_i\}_{i=1}^m$.

A standard approach for solving (4.1) is to use an interpolant to approximate the delayed values, and to apply a Runge–Kutta method to the resulting ODE. Formally we assume that a method as described in Section 3 is applied to the ODE

$$\begin{aligned} &y^{w\prime}(t)=F(t, y^w(t), w(t-\tau_1), w(t-\tau_2),\ldots,w(t-\tau_k)), \quad t\geqslant 0,\\ &y^w(0)=\Phi(0), \end{aligned} \tag{4.2}$$

where $w(t) := \Phi(t)$ for $t \in [-\tau_k, 0]$, and for $t > 0$, $w(t)$ denotes either a higher- or lower-order interpolant to the discrete approximation. We must also be careful to "tiptoe" over the points of discontinuity — we assume that the stepsize selection is altered so that each point $\{\hat{t}_i\}$ is crossed with a step of length $O(\delta^{2/q})$. We will also effectively restart the integration after crossing a discontinuity by choosing the stepsize in the same manner as the initial stepsize in Theorem 2.2.

In [7] we derived sufficient conditions for the error control to cause the global error to be asymptotically linear as a function of δ. Here, we extend those results to allow for the more general error control scheme discussed in the last section. The extension of the analysis in [7] is of a similar spirit to the extension of the analysis in [6] given in Sections 2 and 3, and so we state the final result without proof.

Theorem 4.1. *Suppose that we solve* (4.1) *in the manner described above. Suppose further that if we were given exact back-values $y(t - \tau_i)$ and applied the Runge–Kutta method to the corresponding standard ODE, then $\| \hat{\psi}(y_n, t_n) \|$ in* (2.3) *would not vanish for sufficiently small δ.*

If $w(t)$ is a higher-order interpolant, then there exists an interpolant $\eta(t)$ through the mesh data such that

$$\eta(t) - y(t) = V(t)\delta^{p/q} + G(t), \tag{4.3}$$

where $V(t)$ is continuous and is C^1 over $[0, \hat{t}_1)$, over each $(\hat{t}_i, \hat{t}_{i+1})$ and over $(\hat{t}_m, \infty)$, and is independent of δ, and $G(t)$ is piecewise C^1 with zeroth and first derivatives of $O(\delta^{(p+1)/q})$. Further,

$$w(t) - y(t) = V(t)\delta^{p/q} + O(\delta^{(p+1)/q}). \tag{4.4}$$

However, if $w(t)$ is a lower-order interpolant, then (4.3) *and* (4.4) *do not hold.*

Essentially, Theorem 4.1 says that the previous results for standard ODEs also hold for constant-delay ODEs if and only if a higher-order interpolant is used to approximate the delayed values. We mention that in [2] it was also found that higher-order interpolation offered an important advantage over lower-order interpolation in the context of local error estimation for uniformly corrected implicit Runge–Kutta methods.

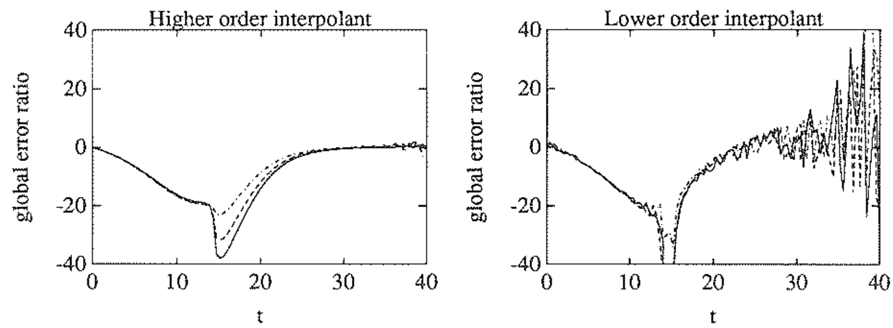

Fig. 2. Global error ratios on delayed logistic equation for three different tolerance values.

We illustrate Theorem 4.1 using the delayed logistic equation

$$y'(t) = \tfrac{1}{4}y(t)\big(1 - \tfrac{1}{20}y(t-1)\big), \quad t \geqslant 0,$$
$$y(t) = 1, \quad t \in [-1, 0].$$

We used a second- and third-order Runge–Kutta formula in extrapolated error-per-unit-step mode, so that $p = 3$ and $q = 2$. Two interpolants were tested: the Hermite piecewise cubic polynomial defined over each step by $w(t_{n-1}) = y_{n-1}$, $w(t_n) = y_n$, $w'(t_{n-1}) = f(t_{n-1}, y_{n-1})$, $w'(t_n) = f(t_n, y_n)$; and the quadratic piecewise polynomial which satisfies only the first three of the above interpolation conditions. Tolerance values of $\delta = 10^{-4}, 10^{-5}, 10^{-6}$ were used. In Fig. 2 we plot the *global error ratios* $[w(t_i) - \hat{y}(t_i)]/\delta^{p/q}$ for 101 equally spaced values t_i. (Since the true solution is not known, we used the numerical solution with $\delta = 10^{-7}$ to generate $\hat{y}(t_i)$.) Note that the cubic interpolant is higher order (locally $O(h_n^4)$), and gives global error ratios that seem to approach a limit as δ decreases. For the lower-order quadratic interpolant, which is locally $O(h_n^3)$, the ratios do not settle down to a fixed limit function, and the characteristic sawtooth oscillations can be seen. These results are in agreement with Theorem 4.1.

Acknowledgement

I thank R. Kid for useful comments on this manuscript.

References

[1] U.M. Ascher, R.M.M. Mattheij and R.D. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations* (Prentice-Hall, Englewood Cliffs, NJ, 1988).

[2] A. Bellen and M. Zennaro, Numerical solution of delay differential equations by uniform corrections to an implicit Runge–Kutta method, *Numer. Math.* **47** (1985) 301–316.

[3] W.H. Enright, A new error-control for initial value solvers, *Appl. Math. Comput.* **31** (1989) 288–301.

[4] E. Hairer, S.P. Nørsett and G. Wanner, *Solving Ordinary Differential Equations I* (Springer, Berlin, 1987).

[5] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II* (Springer, Berlin, 1991).

[6] D.J. Higham, Global error versus tolerance for explicit Runge–Kutta methods, *IMA J. Numer. Anal.* **11** (1991) 457–480.

[7] D.J. Higham, Error control for initial value problems with discontinuities and delays, Numerical Analysis Report NA/129, Dept. Math. Comput. Sci., Univ. Dundee, 1991.

[8] C.B. Moler, J.N. Little and S. Bangert, *PC-Matlab User's Guide* (MathWorks, Inc., South Natick, MA, 1987).

[9] B. Owren and M. Zennaro, Derivation of efficient continuous explicit Runge–Kutta methods, Technical Report 240/90, Dept. Comput. Sci., Univ. Toronto, 1990; also: *SIAM J. Sci. Statist. Comput.*, to appear.

[10] L.F. Shampine, Local error control in codes for ordinary differential equations, *Appl. Math. Comput.* **3** (1977) 189–210.

[11] H.J. Stetter, Considerations concerning a theory for ODE-solvers, in: R. Burlisch, R.D. Grigorieff and J. Schröder, Eds., *Numerical Treatment of Differential Equations*, Proc. Oberwolfach, 1976, Lecture Notes in Math. **631** (Springer, Berlin, 1978) 188–200.

[12] H.J. Stetter, Tolerance proportionality in ODE-codes, in: R. März, Ed., *Proc. Second Conf. on Numerical Treatment of Ordinary Differential Equations*, Seminarberichte **32** (Humboldt University, Berlin, 1980) 109–123; also: in: R.D. Skeel, Ed., *Working Papers for the 1979 SIGNUM Meeting on Numerical Ordinary Differential Equations*, Dept. Comput. Sci., Univ. Illinois, Urbana-Champaign.