Available online at www.sciencedirect.com
ScienceDirect
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
Journal of Computational and Applied Mathematics 218 (2008) 329–341
www.elsevier.com/locate/cam

Global error estimation based on the tolerance proportionality for some adaptive Runge–Kutta codes☆

M. Calvo[b], S. González-Pinto[a,*], J.I. Montijano[b]

[a] *Dpto. de Análisis Matemático, Universidad de La Laguna, 38208, La Laguna-Tenerife, Spain*
[b] *Dpto. Matematica Aplicada, Fac. Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain*

Received 29 September 2006; received in revised form 26 January 2007

Abstract

Modern codes for the numerical solution of Initial Value Problems (IVPs) in ODEs are based in adaptive methods that, for a user supplied tolerance δ, attempt to advance the integration selecting the size of each step so that some measure of the local error is $\simeq \delta$. Although this policy does not ensure that the global errors are under the prescribed tolerance, after the early studies of Stetter [Considerations concerning a theory for ODE-solvers, in: R. Burlisch, R.D. Grigorieff, J. Schröder (Eds.), Numerical Treatment of Differential Equations, Proceedings of Oberwolfach, 1976, Lecture Notes in Mathematics, vol. 631, Springer, Berlin, 1978, pp. 188–200; Tolerance proportionality in ODE codes, in: R. März (Ed.), Proceedings of the Second Conference on Numerical Treatment of Ordinary Differential Equations, Humbold University, Berlin, 1980, pp. 109–123] and the extensions of Higham [Global error versus tolerance for explicit Runge–Kutta methods, IMA J. Numer. Anal. 11 (1991) 457–480; The tolerance proportionality of adaptive ODE solvers, J. Comput. Appl. Math. 45 (1993) 227–236; The reliability of standard local error control algorithms for initial value ordinary differential equations, in: Proceedings: The Quality of Numerical Software: Assessment and Enhancement, IFIP Series, Springer, Berlin, 1997], it has been proved that in many existing explicit Runge–Kutta codes the global errors behave asymptotically as some rational power of δ. This step-size policy, for a given IVP, determines at each grid point t_n a new step-size $h_{n+1} = h(t_n; \delta)$ so that $h(t; \delta)$ is a continuous function of t.

In this paper a study of the tolerance proportionality property under a discontinuous step-size policy that does not allow to change the size of the step if the step-size ratio between two consecutive steps is close to unity is carried out. This theory is applied to obtain global error estimations in a few problems that have been solved with the code Gauss2 [S. Gonzalez-Pinto, R. Rojas-Bello, Gauss2, a Fortran 90 code for second order initial value problems, ⟨http://www.netlib.org/ode/⟩], based on an adaptive two stage Runge–Kutta–Gauss method with this discontinuous step-size policy.

Keywords: Runge–Kutta methods; Global error estimates; Adaptive methods; Tolerance proportionality

1. Introduction

We consider the initial value problem (IVP) for a general differential system

$$y'(t) = f(y(t)), \quad y(0) = y_0 \in \mathbf{R}^m, \quad t \in [0, t_{\text{end}}] \equiv J, \tag{1.1}$$

☆ Work supported by projects MTM2004-06466-C02-01 and MTM2004-06466-C02-02.
* Corresponding author. Tel.: +34 922 318 201.
E-mail addresses: calvo@unizar.es (M. Calvo), spinto@ull.es (S. González-Pinto), monti@unizar.es (J.I. Montijano).

0377-0427/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2007.02.034

which is assumed to possess a unique solution $y(t)$ on J. For simplicity, we will restrict our study to autonomous systems with f sufficiently differentiable in a neighborhood of the trajectory $V_y = \{y = y(t), t \in J\}$. Further, for the derivation of the tolerance proportionality (TP) property, it will be supposed that the product of the length of J by the Lipschitz constant of f in V_y is not too large.

For the numerical solution of (1.1) we consider adaptive Runge–Kutta (RK) methods. That means that for every small tolerance $\delta > 0$, a non-uniform grid $\{t_n\}_{n=0}^{N} \in J$, $t_0 = 0$, $t_N = t_{\text{end}}$, together with approximations $y_n \simeq y(t_n)$, $n = 1, 2, \ldots, N$, both depending on δ, are dynamically generated. In this process the grid is determined so that an estimate of the local errors is maintained under the prescribed tolerance δ. A crucial and natural question is to know how the global errors $\text{ge}_n = y_n - y(t_n)$ depend on the given tolerance. It was established by Stetter [19,20] that for some adaptive RK methods and sufficiently smooth problems the global error has an asymptotically linear dependence on the tolerance in the sense that $\text{ge}_n = v(t_n)\delta + \text{o}(\delta)$, $\delta \to 0^+$, $n = 1, 2, \ldots$, holds for some $\mathscr{C}^1$ function $v(t)$ independent of δ. This property is known in the literature as TP. Later investigations by Higham [10,11] extended Stetter's results to more general adaptive methods that possess a TP dependence of type $\text{ge}_n = v(t_n)\delta^r + \text{o}(\delta^r)$, $\delta \to 0^+$, $n = 1, 2, \ldots$, with a real exponent $r > 0$. A requirement of the Stetter–Higham analysis is the non-vanishing leading term in the local error estimate along the integration interval. When the leading term in the local error estimate vanishes at some isolated points, a modified step-size change technique (regarding the standard one) has been considered in [3] to preserve TP.

In the above-mentioned TP theory developed by Higham [10–12,2] a continuous variation of the step-size along the integration interval was assumed in the sense that the step-size h_{n+1} in advancing from t_n to $t_{n+1} = t_n + h_{n+1}$ satisfies an asymptotic relation $h_{n+1} = \rho(t_n)\delta^{1/q} + \text{o}(\delta^{1/q})$ with a fixed integer $q \geqslant 1$ and a continuous problem-depending function $\rho(t) \geqslant \rho^* > 0$. Such a relation implies that, in general, the step-size will vary from step to step but this does not affect negatively the performance of explicit RK codes in which the computational cost is independent of the size of the step. However, for adaptive codes based on implicit RK formulas, such as Gauss formulas (that may be convenient for special problems which require the preservation of quadratic invariants, simplecticness or some special stability requirements) if the implicit equations of stages are solved by using a simplified Newton iteration, each step-size variation requires a new LU decomposition of the iteration matrix and it increases significantly the computational cost. A way to reduce the number of step-size changes would be to retain the step-size unchanged if h_{n+1}/h_n is close to the unity and this amounts to use a step-size $h_{n+1} = \hat{\rho}(t_n)\delta^{1/q} + \text{o}(\delta^{1/q})$ with a piecewise continuous $\hat{\rho}(t) \geqslant \rho^* > 0$. This technique was already used by Shampine and Gordon in the code DSTEP [17] based on variable coefficient multistep Adams methods to reduce the computational cost. A similar discontinuous strategy have been used in LSODE [13] to improve the stability. Thus, the TP theory of Higham developed under the assumption of a continuous step-size variation cannot be directly applied to those adaptive methods allowing a piecewise constant step-size variation.

On the other hand, in convergent adaptive ODE solvers an integration with tolerance δ does not imply that the global errors are smaller than δ. A desirable practical feature of a numerical integration code is the possibility to provide to the interested user, information on the magnitude of the global errors. As remarked in [1] one may wish to improve the numerical solution when the global error estimation is not good enough. In such a case the technique must be reliable, in the sense that the estimation has some significant digits or it is in the size of the true global errors. An overview about most of global error estimations considered in the literature can be seen in [1,18] and the references therein. In general, these estimations are costly, fail in problems with stability difficulties or when the numerical method is near the limit of accuracy and they are not included in standard integrators. An alternative less demanding is to make use of the TP property exhibited for many adaptive codes, in which case the existence of an asymptotic expansion on the tolerance (δ), allows to derive an estimate of the global error [10,11,16,19,20]. This estimation is based on global extrapolation when considering two independent integrations for different tolerances. Observe that since this technique is based on asymptotic expansions on δ, it may fail at crude tolerances because the asymptotic expansion may not be present and also for very stringent tolerances because of the effects of finite precision. In any case it is important to remark that this technique can be easily applied to any standard integrator.

The aim of this paper is two-fold: firstly, to study the TP property for adaptive codes with the abovementioned discontinuous step-size changing policy and secondly, to apply it to estimate the global error in the code [6,7], which is based on the fully implicit two stage RK Gauss method, intended to integrate second order problems. The paper is organized as follows: in Section 2 we collect a number of definitions and results following the ideas of Higham [10,11] that will be used in the rest of the paper. In Section 3 we study the effect of a discontinuous step-size policy on the numerical integration and a suitable modification of the TP theory of Higham is given. In Section 4 we present a global error estimation based on TP. Finally, in Section 5, the results of some numerical experiments with the code [6] for several test problems are presented to show that this global error estimation can be easily incorporated into the code and provide reliable estimates.

2. Definitions and basic results

By an adaptive RK method we mean an algorithm that, for all IVP (1.1) and for all sufficiently small tolerance $\delta > 0$, defines a discrete approximation $\{(t_n, y_n)\}_{n=0}^{N}$, with $N = N(\delta)$, $t_n = t_n(\delta)$ to the solution of (1.1). The adaptive method is specified by the following elements:

- An s-stage RK formula $\phi_h = \phi_{h,f}$ with order p. Thus, for every (t_n, y_n) and step-size h the approximation $\phi_h(y_n)$ must satisfy

$$\phi_h(y_n) - z_{n+1}(t_n + h) = \psi(y_n)h^{p+1} + \mathcal{O}(h^{p+2}) \quad (h \to 0), \tag{2.1}$$

  where $z_{n+1}(t)$ represents the local solution given by

$$z'_{n+1}(t) = f(z_{n+1}(t)), \quad z_{n+1}(t_n) = y_n, \quad t \in (t_n, t_n + h]. \tag{2.2}$$

  The left-hand side of (2.1), denoted by $\mathrm{le}(y_n; h)$, is the local error of ϕ_h at (t_n, y_n) with step-size h.
- A (scalar) continuous local error estimate $e(y_n; h)$ having the form

$$e(y_n; h) = \widehat{\psi}(y_n)h^q + \mathcal{O}(h^{q+1}). \tag{2.3}$$

  It will be assumed to have a leading part, $\widehat{\psi}(y)$ Lipschitz continuous on y and positive

$$\widehat{\psi}(y(t)) \geqslant \kappa > 0, \quad t \in J, \tag{2.4}$$

  for some constant κ.
- A step-size changing policy that includes a predictor for the first step-size h_1, a criterion of acceptance of each step and a prediction for the next step-size h_{n+1} to be used at (t_n, y_n) to advance the numerical solution from $(t_n, y_n) \to (t_{n+1} = t_n + h_{n+1}, y_{n+1})$.

Concerning the local error control function many adaptive RK methods employed in practical codes (see e.g., [4,10, 11,15]) take a norm of the difference between two embedded RK formulas ϕ_h and $\widehat{\phi}_h$ with orders p and $(q-1) < p$, respectively, so that $e(y_n; h) = \|\phi_h(y_n) - \widehat{\phi}_h(y_n)\| = \mathcal{O}(h^q)$. In this case the computation of the local error control has a low computational cost. Most of local error estimates in the literature satisfy (2.3), see e.g., those in [5,9,11].

Next, we recall the main features of the standard step-size policy that are relevant to our study. Firstly, for a given tolerance δ, the initial step-size h_1 is chosen so that $e(y_0, h_1) \simeq \theta^q \delta$ where θ is a fixed constant $0 < \theta < 1$. Secondly, we take $e(y_{n-1}; h_n) \leqslant \delta$ as acceptance criterion of the step $(t_{n-1}, y_{n-1}) \to (t_n = t_{n-1} + h_n, y_n = \phi_{h_n}(y_{n-1}))$. After an accepted step from $t_{n-1} \to t_n = t_{n-1} + h_n$ the step-size h_{n+1} for the next step is predicted by

$$h_{n+1} = r_n h_n \equiv \theta\left(\frac{\delta}{e_n}\right)^{1/q} h_n, \quad e_n \equiv e(y_{n-1}; h_n). \tag{2.5}$$

It is important to notice that with this step-size policy it has been proved in [11] that $e_n = \theta^q \delta + \mathcal{O}(\delta^{(q+1)/q})$, as $\delta \to 0^+$ and, due to the safety factor θ, no step failures will arise for δ sufficiently small. From (2.3)–(2.5) it follows that all step-sizes behave as $\mathcal{O}(\delta^{1/q})$ as $\delta \to 0^+$.

Next, we recall a restricted version of the TP property introduced by Stetter [19] and later refined by Higham [11] to study the relationship between the global error and the tolerance in the numerical solutions generated by adaptive RK methods. To compare discrete solutions obtained with different tolerances and defined on different grids we need an interpolation process that extends a discrete solution to J. Then, for a discrete solution $\{(t_n, y_n)\}_{n=0}^{N}$ computed with an adaptive method with tolerance δ (here $N = N(\delta)$ and $t_n = t_n(\delta)$, but the dependence on δ has been omitted to simplify the notations), we define the "ideal interpolant" $\eta_\delta(t)$, $t \in J$ of this discrete solution as

$$\eta_\delta(t) = \begin{cases} y_0 & \text{for } t = t_0 = 0, \\ z_n(t) + \dfrac{(t - t_{n-1})}{h_n}\mathrm{le}_n & \text{for } t \in (t_{n-1}, t_n], \ n = 1, \dots, N, \end{cases} \tag{2.6}$$

where $\mathrm{le}_n = \mathrm{le}(y_{n-1}; h_n) = \phi_{h_n}(y_{n-1}) - z_n(t_n)$ is the local error of the advancing RK formula ϕ_h in the step from $t_{n-1} \to t_n$. This interpolant is a piecewise $\mathscr{C}^1$ function in J because the left and right derivatives at grid points may be different. Moreover, $\eta_\delta(t)$ is not explicitly computed in practice, but it is merely a tool to define the TP property and to establish some basic results. It must be also observed that according to the step-size change policy (2.5), the end-point $t_{\mathrm{end}} = t_N$, will not be in general a natural grid-point.

Definition 1. An adaptive method is said to possess the TP property if for all sufficiently smooth IVP (1.1) and small δ, the ideal interpolant (2.6) of the discrete solution generated with tolerance δ satisfies

$$\eta_\delta(t) - y(t) = v(t)\delta^{p/q} + g(t;\delta)\delta^{(p+1)/q}, \quad \delta \to 0^+, \ \ t \in J, \tag{2.7}$$

with some $v \in \mathscr{C}^1(J; \mathbf{R}^m)$ independent of δ and some g piecewise $\mathscr{C}^1$ with respect to t and uniformly bounded as $\delta \to 0^+$.

A main result to assess the TP property of some adaptive RK methods proved by Higham in [11] is the following:

Theorem 1. (1) *If there exist $\gamma \in \mathscr{C}(J, \mathbf{R}^m)$ independent of δ and $s(t,\delta)$ piecewise continuous such that*

$$\eta'_\delta(t) - f(\eta_\delta(t)) = \gamma(t)\delta^{p/q} + \delta^{(p+1)/q}s(t,\delta) \tag{2.8}$$

holds, then the adaptive RK method possess the TP property with $v(t)$ defined as the unique solution of the variational equation of (1.1)

$$v'(t) - f_y(y(t))v(t) = \gamma(t), \quad v(0) = 0. \tag{2.9}$$

(2) *For the standard step-size policy given above, condition* (2.7) *holds with the $\mathscr{C}^1(J, \mathbf{R}^m)$ function $v(t)$ given by* (2.9) *with*

$$\gamma(t) = \frac{\psi(y(t))\theta^p}{\widehat{\psi}(y(t))^{p/q}}. \tag{2.10}$$

Remark 1. The function $v(t)$ of the TP property (2.7) only depends on the continuous function (2.10). This means that, apart of $\psi(y(t))$, only the leading term of $e(y;h)$ along the exact solution of (1.1) contributes to $v(t)$. In other words, two adaptive methods with the same ϕ_h and different local error estimates e and $\widetilde{e}$ of order q, such that $e(y_{n-1};h) - \widetilde{e}(y_{n-1};h) = \mathcal{O}(h^{q+1})$ have the same asymptotic TP behavior.

In particular, if an adaptive RK method with local error estimate $e(y_{n-1};h) = \widehat{\psi}(y_{n-1})h^q + \mathcal{O}(h^{q+1})$ has the TP property, then it also has the TP property with respect to $\widetilde{e}(y_{n-1};h) = \widehat{\psi}(y(t_n))h^q$. Now the new step-size prediction is

$$h_{n+1} = \theta\left(\frac{\delta}{\widehat{\psi}(y(t_n))h_n^q}\right)^{1/q} h_n = \theta\left(\frac{\delta}{\widehat{\psi}(y(t_n))}\right)^{1/q}.$$

This clearly represents a continuous step-size selection of the form $h_{n+1} = \nu(t_n)\delta^{1/q}$ with $\nu(t) = \theta/\widehat{\psi}(y(t))^{1/q}$, $t \in J$.

As remarked in [12] the results of Theorem 1 are relevant to explain the global error behavior in non-stiff problems. For some stiff problems a possible order reduction on the basic method and the error estimate might occur. Thus, by assuming orders $\widetilde{p}$ and $\widetilde{q}$ for the method and the estimate, respectively, it can be expected to achieve a TP behavior with a different exponent $r = \widetilde{p}/\widetilde{q}$.

3. The TP property for adaptive RK methods with the new step-size changing technique

Since adaptive RK methods with the standard step-size predictor (2.5) change continuously the size of the step along the integration interval they deteriorate the efficiency of implicit RK solvers that use modified Newton iterations to solve the implicit equations of stages, because the LU-decomposition of $(I_m - h\gamma J_{\mathrm{acob}})$ must be updated after changing the step-size h (I_m is the identity matrix, γ some constant and J_{acob} is the Jacobian matrix $\partial f/\partial y$ evaluated at some previous point y_m). Then let us consider instead of (2.5) the following (discontinuous) step-size changing predictor

$$h_{n+1} = \widehat{r}_n h_n \quad \text{with } \widehat{r}_n = \begin{cases} 1 & \text{if } r_n \in (r', r''), \\ r_n & \text{otherwise,} \end{cases} \tag{3.1}$$

with $r_n = \theta(\delta/e_n)^{1/q}$ as in (2.5), and $0 < r' < 1 < r''$ are fixed numbers.

Note that, in contrast with the continuous step-size predictor (2.5), with the technique (3.1) the step-size is modified only when either $r_n \geqslant r'' > 1$ or $r_n \leqslant r' < 1$.

As a first remark observe that

$$\begin{aligned} e_n = e(y_{n-1}; h_n) &= \widehat{\psi}(y_{n-1})h_n^q + \mathcal{O}(h_n^{q+1}) = \widehat{\psi}(y_{n-2})\widehat{r}_{n-1}^q h_{n-1}^q + \mathcal{O}(h_n^{q+1}) \\ &= (\widehat{r}_{n-1}/r_{n-1})^q \widehat{\psi}(y_{n-2}) r_{n-1}^q h_{n-1}^q + \mathcal{O}(h_n^{q+1}) \\ &= (\widehat{r}_{n-1}/r_{n-1})^q \theta^q \delta + \mathcal{O}(\delta^{(q+1)/q}). \end{aligned}$$

Hence, for $\widehat{r}_{n-1} > r_{n-1}$ since $\widehat{r}_{n-1}/r_{n-1} \leqslant 1/r'$, we will assume $\theta/r' < 1$ to ensure that for δ small there will be no rejected steps.

In order to analyze the TP property of an adaptive RK method with the new step-size control (3.1), we start defining the strictly positive scalar function $w(t) := \widehat{\psi}(y(t))^{-1/q}$, $t \in J$. Then, the local error estimation (2.3) satisfies

$$e_n = \widehat{\psi}(y(t_{n-1}))h_n^q + \mathcal{O}(h_n^{q+1}) = w(t_{n-1})^{-q} h_n^q + \mathcal{O}(h_n^{q+1}), \quad h_n \to 0^+. \tag{3.2}$$

Also assume that $e_1 = \theta^q \delta + \mathcal{O}(\delta^{(q+1)/q})$ is satisfied for the first step-size, i.e.,

$$h_1 = \theta w(t_0)\delta^{1/q} + \mathcal{O}(\delta^{2/q}), \quad \delta \to 0^+. \tag{3.3}$$

Then, from (3.2) and (3.3), we deduce that

$$r_1 = \theta\left(\frac{\delta}{e_1}\right)^{1/q} = \frac{\theta w(t_0)\delta^{1/q}}{h_1(1 + \mathcal{O}(h_1))} = \frac{w(t_0)}{w(t_0)} + \mathcal{O}(\delta^{1/q}) = 1 + \mathcal{O}(\delta^{1/q}).$$

Consequently, $\hat{r}_1 = 1$ and $h_2 = h_1$, for $\delta \to 0^+$. Next, proceeding as before and taking into account that $h_2 = h_1$, we get that

$$r_2 = \theta\left(\frac{\delta}{e_2}\right)^{1/q} = \frac{\theta w(t_1)\delta^{1/q}}{h_2(1 + \mathcal{O}(h_2))} = \frac{w(t_1)}{w(t_0)} + \mathcal{O}(\delta^{1/q}) = 1 + \mathcal{O}(\delta^{1/q}).$$

Again for $\delta \to 0^+$, we have that $\hat{r}_2 = 1$ and $h_3 = h_2 = h_1$. Proceeding in the same way for the next step-size ratios, we can set

$$r_n = \theta\left(\frac{\delta}{e_n}\right)^{1/q} = \frac{w(t_{n-1})}{w(t_0)} + \mathcal{O}(\delta^{1/q}) \quad \text{and} \quad h_n = h_1 \ \text{ for } n = 1, 2, 3, \ldots, n_1, \tag{3.4}$$

where n_1 is the first positive integer such that $r_{n_1} \notin (r', r'')$. Since the step-sizes are $\mathcal{O}(\delta^{1/q})$, we get that

$$\text{(a) } r_{n_1} = r' + \mathcal{O}(\delta^{1/q}) \quad \text{or} \quad \text{(b) } r_{n_1} = r'' + \mathcal{O}(\delta^{1/q}). \tag{3.5}$$

By denoting $\bar{t}_1(\delta) \equiv t_{n_1}$, it is expectable in view of (3.4) and (3.5) that,

$$|\bar{t}_1(\delta) - \tau_1| = \mathcal{O}(\delta^{1/q}), \quad \delta \to 0^+, \tag{3.6}$$

where $\tau_1 > t_0$ is the first t-point satisfying

$$\text{either (a) } w(\tau_1) = r' w(t_0) \quad \text{or} \quad \text{(b) } w(\tau_1) = r'' w(t_0). \tag{3.7}$$

Then, for the next step-size we would have that

$$\begin{aligned} h_{n_1+1} &= \widehat{r}_{n_1} h_{n_1} = r_{n_1} h_{n_1} = r_{n_1} h_1 = \theta w(t_{n_1-1})\delta^{1/q} + \mathcal{O}(\delta^{2/q}) \\ &= \theta w(t_{n_1})\delta^{1/q} + \mathcal{O}(\delta^{2/q}) = \theta w(\bar{t}_1(\delta))\delta^{1/q} + \mathcal{O}(\delta^{2/q}) = \theta w(\tau_1)\delta^{1/q} + \mathcal{O}(\delta^{2/q}). \end{aligned} \tag{3.8}$$

Since the situation for h_{n_1+1} in (3.8) is the same as for h_1 in (3.3), the process from (3.3) to (3.7) can be practically repeated. Now, instead of (3.4) we have

$$r_n = \theta\left(\frac{\delta}{e_n}\right)^{1/q} = \frac{w(t_{n-1})}{w(\bar{t}_1(\delta))} + \mathcal{O}(\delta^{1/q}) \quad \text{and} \quad h_n = h_{n_1+1} \text{ for } n = n_1+1, \ldots, n_2, \tag{3.9}$$

where n_2 is the first positive integer such that $r_{n_2} \notin (r', r'')$. Proceeding as before and denoting $\bar{t}_2(\delta) \equiv t_{n_2}$, it can be expected that $|\bar{t}_2(\delta) - \tau_2| = \mathcal{O}(\delta^{1/q})$, $\delta \to 0^+$, where $\tau_2 > \tau_1$ is the first t-point satisfying either (a) $w(\tau_2) = r'w(\tau_1)$ or (b) $w(\tau_2) = r''w(\tau_1)$. This process can be only repeated a finite number of times in the interval $[t_0, t_{\text{end}}]$, independently of δ, since as we next prove

$$\tau_1 - t_0 \geqslant L_w^{-1} w(t_0) \min\{r'' - 1, 1 - r'\}, \tag{3.10}$$

where L_w is a Lipschitz constant for the function $w(t)$, which can be calculated from a Lipschitz constant for $\hat{\psi}(y(t))$, denoted by $L_{\hat{\psi}}$. Below, the constant κ is that one given in (2.4).

$$\begin{aligned}|w(t+\tau) - w(t)| &= |(\widehat{\psi}(y(t+\tau))^{-1/q} - \widehat{\psi}(y(t))^{-1/q})| \\ &\leqslant q^{-1}\kappa^{-1-q^{-1}} |\widehat{\psi}(y(t+\tau)) - \widehat{\psi}(y(t))| \leqslant q^{-1}\kappa^{-1-q^{-1}} L_{\widehat{\psi}}|\tau| = L_w|\tau|.\end{aligned}$$

Then, if $w(\tau_1) = r''w(t_0)$, it follows that $w(\tau_1) - w(t_0) = (r''-1)w(t_0) \leqslant L_w(\tau_1 - t_0)$.

This implies that $\tau_1 - t_0 \geqslant L_w^{-1}(r''-1)w(t_0)$. For the case, $w(\tau_1) = r'w(t_0)$, a similar reasoning can be made. From here, we get (3.10).

Remark 2. The τ_j-points can be defined recursively as: τ_j is the first point on the right of τ_{j-1} satisfying

$$w(\tau_j)/w(\tau_{j-1}) \notin (r', r''), \quad j = 1, 2, \ldots, \quad \tau_0 = t_0. \tag{3.11}$$

In a similar way, $\bar{t}_j(\delta)$ is the first point on the right of $\bar{t}_{j-1}(\delta)$ satisfying

$$w(\bar{t}_j(\delta))/w(\bar{t}_{j-1}(\delta)) + \lambda(\delta)\delta^{1/q} \notin (r', r''), \quad j = 1, 2, \ldots, \quad \bar{t}_0(\delta) = t_0, \tag{3.12}$$

for a certain uniform bounded function $\lambda(\delta)$. Moreover as explained before, in the interval $[\bar{t}_{j-1}(\delta), \bar{t}_j(\delta))$, the step-sizes remain constant with length, $h = \theta w(\bar{t}_{j-1}(\delta))\delta^{1/q} + \mathcal{O}(\delta^{2/q})$.

Under the step-size policy (3.1), the code will advance with constant step-sizes except at an (asymptotically finite) number of grid points $\bar{t}_1(\delta), \ldots, \bar{t}_l(\delta)$ such that $\bar{t}_j(\delta) - \tau_j = \mathcal{O}(\delta^{1/q})$. A sufficient condition to guarantee the last equality is given next.

Theorem 2. *Suppose that function $w(t)$ is strictly monotone at the points τ_j and that the inverse function $w^{-1}(t)$ is Lipschitz continuous in some neighborhood of $w(\tau_j)$, for $j = 1, 2, \ldots$. If the initial step satisfies* (3.3) *then $\bar{t}_j(\delta) \to \tau_j$ as $\delta \to 0^+$. Moreover $|\bar{t}_j(\delta) - \tau_j| = \mathcal{O}(\delta^{1/q})$.*

Proof. The proof follows from the following lemma. □

Lemma 1. (i) *Suppose that $\varphi(t)$ is a Lipschitz continuous function on $[a, b]$, and that $s^* > s$ $(a \leqslant s < s^* < b)$ is the first point on the right of s such that $\varphi(s^*) = r\varphi(s)$, for some positive constant $r \neq 1$. Assume that $\varphi(t)$ is strictly monotone at s^* (i.e., in some neighborhood of s^*) and that the inverse function $\varphi^{-1}(t)$ is Lipschitz continuous in some neighborhood of $t^* = \varphi(s^*)$. Then the equation (below, $\kappa_j \equiv \kappa_j(\delta)$ denotes a uniformly bounded function)*

$$\frac{\varphi(x)}{\varphi(s + \kappa_1\delta^{1/q})} + \kappa_2\delta^{1/q} = r, \quad \delta \to 0^+ \ \ (q > 0 \textit{ is a constant}) \tag{3.13}$$

has a solution $x = x(\delta)$, *satisfying* $x(\delta) - s^* = \mathcal{O}(\delta^{1/q})$ *as* $\delta \to 0^+$. *This solution is unique in some subinterval of the form* $[s, s^* + \tau]$, *where* $\tau > 0$ *is a constant*.

(ii) *If there does not exist* $s^* \in [s, b]$ *such that* $\varphi(s^*) = r\varphi(s)$, *then Eq.* (3.13) *has no solution for* $\delta \to 0^+$.

Proof. (i) By solving for x in (3.13), it follows that $|x - s^*| = |\varphi^{-1}((r + \mathcal{O}(\delta^{1/q}))\varphi(s + \mathcal{O}(\delta^{1/q}))) - \varphi^{-1}(r\varphi(s))| = |\varphi^{-1}(r\varphi(s) + \mathcal{O}(\delta^{1/q})) - \varphi^{-1}(r\varphi(s))| = L_{\varphi^{-1}}\mathcal{O}(\delta^{1/q}) = \mathcal{O}(\delta^{1/q})$. To see the uniqueness of solution, assume for simplicity that $\varphi(s) > 0$ and $r > 1$. Consider some interval $[s^* - \tau, s^* + \tau]$ where φ is strictly increasing, there exists the inverse function φ^{-1} (on $I = \varphi([s^* - \tau, s^* + \tau])$) and $\varphi(t) \leqslant \varphi(s^* - \tau)$, $t \in [s, t^* - \tau]$. It follows that, for $\delta \to 0^+$, Eq. (3.13) cannot have any solution $x \in [s, s^* - \tau]$, since $\max_{t \in [s, s^* - \tau]} \varphi(t) < r\varphi(s) = \varphi(s^*)$. On the other hand, Eq. (3.13) can only admit a solution in the interval $[s^* - \tau, s^* + \tau]$, by virtue of the existence of φ^{-1}. Part (ii) is straightforward to prove. □

Theorem 3. *For an adaptive RK method with step-size control* (3.1) *assume that*:

- *the problem is sufficiently smooth for the expansions* (2.1), (2.3) *and* (2.4) *to hold*;
- *the initial step-size satisfies* (3.3);
- *the function* $w(t)$ *satisfies the assumptions in Theorem* 2.

Then, the adaptive method has the TP property, more precisely the global error satisfies

$$y_\delta(t) - y(t) = v(t)\delta^{p/q} + \mathcal{O}(\delta^{(p+1)/q}), \quad \delta \to 0^+, \ \ t \in J, \tag{3.14}$$

where $v(t)$ *is piecewise* $\mathbf{C}^1(J)$ *and fulfils*

$$v(t) = \Psi(t) \int_0^t \varphi(s)^p \Psi(s)^{-1} \psi(y(s))\,\mathrm{d}s, \quad t \in J \tag{3.15}$$

being $\varphi(t)$ *the piecewise constant function*: $\varphi(t) = \theta w(\tau_j)$, *whenever* $t \in [\tau_j, \tau_{j+1})$, $j = 0, 1, \ldots, l-1$; $\varphi(t) = \theta w(\tau_l)$ *for* $t \in [\tau_l, t_{\text{end}}]$, *and* $\Psi(t)$ *is the solution of* $\Psi'(t) = f_y(t, y(t))\Psi(t)$, $\Psi(0) = I_m$.

Proof. Take $\delta > 0$. From Remark 2, it is clear that the step-sizes are constant on subintervals $[\bar{t}_{j-1}(\delta), \bar{t}_j(\delta))$, $j = 1, 2, \ldots$. Thus, they satisfy

$$h_{n+1} = \tilde{\varphi}(t_n, \delta)\delta^{1/q} \quad \forall t_n \in [t_0, t_{\text{end}}), \tag{3.16}$$

where $\tilde{\varphi}(t, \delta)$ is piecewise continuous in $[t_0, t_{\text{end}}]$. By virtue of Theorem 2 we deduce that

$$|\tilde{\varphi}(t, \delta) - \varphi(t)| = \mathcal{O}(\delta^{1/q}) \quad \forall t \in J \backslash I, \tag{3.17}$$

where

$$I = \bigcup_{j=1}^{l} I_j(\delta), \quad I_j(\delta) = [\min\{\bar{t}_j(\delta), \tau_j\}, \max\{\bar{t}_j(\delta), \tau_j\}].$$

On the other hand, the ideal interpolant (2.6) satisfies for any $t \in (t_{n-1}, t_n)$ that (take into account (3.16) for the latest equality below)

$$\begin{aligned} \eta_I'(t) - f(\eta_I(t)) &= z_{n-1}'(t) + le_n/h_n - f(z_{n-1}(t) + h_n^{-1}(t - t_n)le_n) = le_n/h_n + \mathcal{O}(le_n) \\ &= \psi(y_{n-1})h_n^p + \mathcal{O}(h_n)^{p+1} = \psi(y(t))h_n^p + \mathcal{O}(h_n)^{p+1} = \tilde{\varphi}(t, \delta)^p \psi(y(t))\delta^{p/q} + \mathcal{O}(\delta^{(p+1)/q}). \end{aligned}$$

From here, taking into account that $y'(t) - f(y(t)) = 0$ and that $\eta_I(t_0) = y(t_0) = y_0$, it follows from the standard theory for inhomogeneous linear differential systems that

$$\eta_I(t) - y(t) = \tilde{v}(t, \delta)\delta^{p/q} + \mathcal{O}(\delta^{(p+1)/q}), \quad t \in J,$$

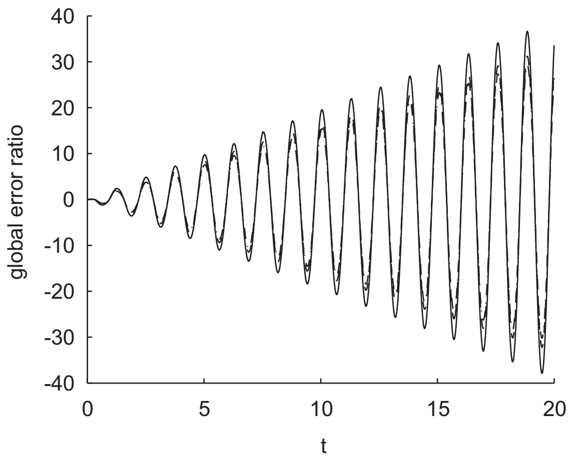

Fig. 1. Global error ratios for Duffing's problem for $\delta = 10^{-k}$, $k = 8, 10, 12, 13$.

with

$$\tilde{v}(t,\delta) = \Psi(t)\int_{t_0}^{t} \tilde{\varphi}(s,\delta)^p\,\Psi(s)^{-1}\psi(y(s))\,\mathrm{d}s, \quad t \in J.$$

Now, the proof is completed after using (3.17). $\square$

It must be observed that the end-point might require an especial adjustment, since it might not be a natural mesh-point. But, this does not represent any problem because the length of latest step-size, namely $t_{\mathrm{end}} - t_{N-1}$ is $\mathcal{O}(\delta^{1/q})$.

To assess the above results on the TP theory, we will present some experiments that illustrate numerically the function $v(t)$. We have considered as test problem Duffing's equation

$$y'' + (\beta^2 + k^2)y = 2k^2y^3, \quad t \in [0, 20],$$

with particular values $\beta = 5$, $k = 0.03$ and initial conditions $y(0) = 0$, $y'(0) = \beta$ which correspond to the periodic solution $y(t) = \mathrm{sn}(\beta t, k/\beta)$. This problem has been integrated with the code *Gauss2* [6], described in [7], that uses the discontinuous step-size policy (3.1) with $p = 4$ and $q = 5$. In Fig. 1 we display the global error ratios $(y_n - y(t_n))/\delta^{p/q}$, that approximate asymptotically $v(t)$ at each mesh point, for the tolerances $\delta = 10^{-j}$, $j = 8, 10, 12, 13$. From this figure it follows that for all tolerances the linear interpolant of the global error ratios display the same shape, and for $\delta \to 0$ appears to converge to a limit as predicted by Theorem 3. In addition, to make clear this convergence we present in Fig. 2 the interpolants in a smaller interval. Note that, even for larger tolerances not included in the figures, the interpolant of the global error ratios is close to the limit function.

4. Estimating global errors by TP

Under the assumptions of either Theorem 1 or else Theorem 3, the global errors of the numerical solution $y_\delta(\bar{t})$, obtained from the integration with tolerance δ at some point $\bar{t}$, can be computed just by making another second integration with tolerance $\tau\delta$, where $\tau \neq 1$ is some constant. Thus, denoting by $y_{\tau\delta}(\bar{t})$ the numerical solution obtained for the second integration, from the above theorems we can write

$$y_\delta(\bar{t}) - y(\bar{t}) = v(\bar{t})\delta^{p/q} + \mathcal{O}(\delta^{(p+1)/q}), \quad y_{\tau\delta}(\bar{t}) - y(\bar{t}) = v(\bar{t})(\tau\delta)^{p/q} + \mathcal{O}((\tau\delta)^{(p+1)/q}). \tag{4.1}$$

The standard extrapolation process leads to

$$y_\delta(\bar{t}) - y(\bar{t}) = (1 - \tau^{p/q})^{-1}(y_\delta(\bar{t}) - y_{\tau\delta}(\bar{t})) + \mathcal{O}(\delta^{(p+1)/q}).$$

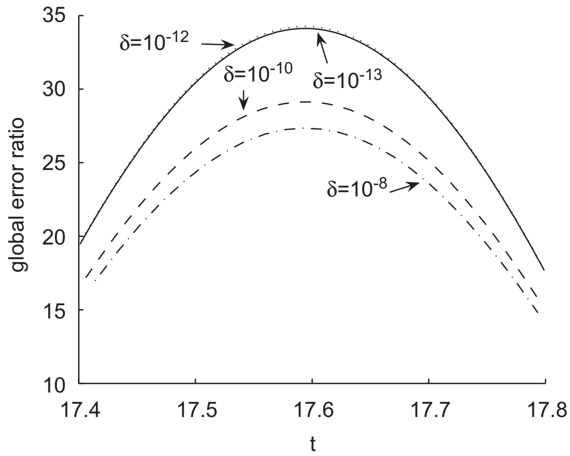

Fig. 2. Zoom of Fig. 1 in a smaller interval.

From here, a computable asymptotically correct global error estimation is obtained by means of

$$y_\delta(\bar{t}) - y(\bar{t}) \simeq (1 - \tau^{p/q})^{-1}(y_\delta(\bar{t}) - y_{\tau\delta}(\bar{t})). \tag{4.2}$$

There is some freedom to choose the value for the constant τ. For practical reasons, it is advisable to take $\tau > 1$, since in this way the second integration carried out with tolerance $\tau\delta$, only for global error estimation purposes, will be in general cheaper than the integration with tolerance δ. It is not advisable to set $\tau \gg 1$, since in that case the estimation in (4.2) becomes meaningless, because the estimates in the previous theorems are only significant for $\delta \to 0^+$. We will discuss more about these topics in next section.

For the numerical experiments presented in the next section, we have applied this technique to estimate the global errors in the code *Gauss2* [6]. This code is described in [7] and it is a second order IVP solver. The advancing formula is the two-stage Gauss method in version RK Nyström, which has order four ($p = 4$). In this case, the local error on the (y, y') components can be represented as

$$\mathrm{le}(t, y, y', h) = h^5(\psi_1(t, y, y'), \psi_2(t, y, y'))^{\mathrm{T}} + \mathcal{O}(h^6), \quad h \to 0. \tag{4.3}$$

Since the local error estimate considered in *Gauss2* is asymptotically correct on the y-component and there is no estimation for the y'-component, then the local error estimate reads

$$e(t, y, y', h) = h^5(\psi_1(t, y, y'), 0)^{\mathrm{T}} + \mathcal{O}(h^6), \quad h \to 0, \tag{4.4}$$

hence, $q = 5$.

The step-size change in the code is made according to formula (3.1), by taking

$$\theta = 0.8, \quad r' = 0.9, \quad r'' = 1.5, \quad r_{\max} = 2,$$

where $r_{\max}$ is the maximum step-size ratio allowed. Thus, assuming that Theorem 3 holds, it is expected that the global errors in *Gauss2* can be estimated by formula (4.2). In practice, this fact will come true especially when the number of rejected steps by the estimator (or by divergence in the iteration) is small compared with the number of successful steps taken by the code to complete the integrations. Observe that from the previous theory, it should not have rejected steps provided that $\delta \to 0^+$.

5. Numerical experiments

We have incorporated the global error estimation (4.2) to the code *Gauss2* and it has been tested on several second order problems, commonly appearing in the related literature. The estimation supplied by (4.2) seems to be reliable, in the sense that it provides a reasonable approximation of the exact global errors, and it is easily implemented (both integrations are totally independent).

Problem 1. The Fermi–Pasta–Ulam problem [8, pp. 17–18], given by the nonlinear second order equations

$$
\begin{aligned}
y_1'' &= (y_2 - y_5 - y_1 - y_4)^3 - (y_1 - y_4)^3, \quad y_1(0) = 1, \quad y_1'(0) = 1,\\
y_2'' &= -(y_2 - y_5 - y_1 - y_4)^3 + (y_3 - y_6 - y_2 - y_5)^3, \quad y_2(0) = 0, \quad y_2'(0) = 0,\\
y_3'' &= -(y_3 - y_6 - y_2 - y_5)^3 - (y_3 + y_6)^3, \quad y_3(0) = 0, \quad y_3'(0) = 0,\\
y_4'' &= (y_2 - y_5 - y_1 - y_4)^3 + (y_1 - y_4)^3 - \omega^2 y_4, \quad y_4(0) = \omega^{-1}, \quad y_4'(0) = 1,\\
y_5'' &= (y_2 - y_5 - y_1 - y_4)^3 + (y_3 - y_6 - y_2 - y_5)^3 - \omega^2 y_5, \quad y_5(0) = 0, \quad y_5'(0) = 0,\\
y_6'' &= (y_3 - y_6 - y_2 - y_5)^3 - (y_3 + y_6)^3 - \omega^2 y_6, \quad y_6(0) = 0, \quad y_6'(0) = 0,
\end{aligned}
$$

where we have taken $\omega = 50$ as recommended in [8, pp. 17–18]. The weighted Euclidean norm of the exact solution at $t_{\text{end}} = 100$ is $\|y(100)\| = 0.344\ldots$.

Problem 2. A slight modification of the second order partial differential equation (PDE) in [14, pp. 426–427], which describes the vibration in a cantilever bar. The modification below consists of introducing a non-homogeneous term $r(t) = (r_i(t))_{i=1}^N$ in the resulting ODE, after discretizing in space the original PDE ($y_{xxxx}(x,t)$ is approximated by using second order centered differences of five points), in such a way that the new ordinary differential system has the same solution as the original PDE [14, pp. 426–427]. The ODE is [14, p. 427]

$$
\begin{pmatrix} y_1'' \\ y_2'' \\ y_3'' \\ \vdots \\ y_{N-2}'' \\ y_{N-1}'' \\ y_N'' \end{pmatrix}
= -\frac{200}{(\Delta x)^4}
\begin{pmatrix}
7 & -4 & 1 & & & & \\
-4 & 6 & -4 & 1 & & 0 & \\
1 & -4 & 6 & -4 & 1 & & \\
 & \ddots & \ddots & \ddots & \ddots & \ddots & \\
 & & 1 & -4 & 6 & -4 & 1 \\
 & 0 & & 1 & -4 & 5 & -2 \\
 & & & & 2 & -4 & 2
\end{pmatrix}
\cdot
\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{N-2} \\ y_{N-1} \\ y_N \end{pmatrix}
+
\begin{pmatrix} r_1(t) \\ r_2(t) \\ r_3(t) \\ \vdots \\ r_{N-2}(t) \\ r_{N-1}(t) \\ r_N(t) \end{pmatrix}, \tag{5.1}
$$

where $\Delta x = 22/N$, N being the number of mesh-points on the x-interval [0,22], and the initial conditions are given by $y_i(0) = f(x_i)$, $y_i'(0) = 0$, $i = 1, \ldots, N$. For our experiments, we have taken $N = 90$ and $t \in [0, 10^4]$. With this choice all eigenvalues of its Jacobian matrix are negative, and it is an oscillatory problem.

The exact solution of the problem is given by

$$y_i(t) = f(x_i)\cos(\omega t), \quad \omega = 0.102735464\ldots, \quad l = 22, \quad x_i = il/N, \quad i = 1, 2, \ldots, N,$$

where

$$
\begin{aligned}
f(x) &= 0.1(\cosh(\lambda x) - \cos(\lambda x)) - K(\sinh(\lambda x) - \sin(\lambda x)),\\
K &= (\sinh(\lambda l) + \sin(\lambda l))^{-1}(\cosh(\lambda l) + \cos(\lambda l)) \quad \text{and} \quad \lambda = 0.08523200128726258.
\end{aligned}
$$

Problem 3. A finite difference discretization of the nonlinear partial differential equation [7]

$$
\begin{aligned}
& y_{tt}(x,t) + (1 + y(x,t)^2)y_{xxxx}(x,t) = \eta^4 y(x,t)^3, \quad -\pi/\eta < x \leqslant \pi/\eta, \quad t > 0,\\
& y(x,0) = \cos \eta x, \quad y_t(x,0) = 0, \quad \eta = 1/4,
\end{aligned} \tag{5.2}
$$

Table 1
Global errors (GE_y) and global errors estimated (GEest_y) with `Gauss2` on the y-component for Problem 1 at $t_{\text{end}} = 100$, for different tolerances δ and factors τ

TOL (δ)	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
CPU-time	0.11	0.16	0.25	0.36	0.52
GE_y	$2.3 \cdot 10^{-2}$	$3.4 \cdot 10^{-3}$	$8.3 \cdot 10^{-4}$	$8.3 \cdot 10^{-5}$	$2.1 \cdot 10^{-5}$
$\tau = 1.5$, GEest_y Extra-cost = 92%	$3.5 \cdot 10^{-2}$	$4.3 \cdot 10^{-3}$	$9.1 \cdot 10^{-4}$	$2.4 \cdot 10^{-4}$	$2.0 \cdot 10^{-5}$
$\tau = 2.0$, GEest_y Extra-cost = 82%	$2.7 \cdot 10^{-2}$	$7.5 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$2.3 \cdot 10^{-4}$	$1.9 \cdot 10^{-5}$
$\tau = 5.0$, GEest_y Extra-cost = 73%	$1.7 \cdot 10^{-2}$	$3.5 \cdot 10^{-3}$	$7.5 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$2.8 \cdot 10^{-5}$
$\tau = 10$, GEest_y Extra-cost = 62%	$1.9 \cdot 10^{-2}$	$4.3 \cdot 10^{-3}$	$5.9 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	$1.2 \cdot 10^{-5}$

Also the average of the extra-costs for the integration with tolerance $\tau\delta$ are displayed. The average of failed-steps versus accepted steps was 0.22%.

Table 2
Global errors (GE_y) and global errors estimated (GEest_y) with `Gauss2` on the y-component for Problem 2 at $t_{\text{end}} = 10\,000$, for different tolerances δ and factors τ

TOL (δ)	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}
CPU-time	0.77	1.36	1.89	2.84	4.59
GE_y	$2.3 \cdot 10^{-2}$	$6.0 \cdot 10^{-4}$	$1.6 \cdot 10^{-5}$	$5.5 \cdot 10^{-7}$	$1.5 \cdot 10^{-8}$
$\tau = 1.5$, GEest_y Extra-cost = 92%	$4.6 \cdot 10^{-2}$	$1.2 \cdot 10^{-3}$	$2.1 \cdot 10^{-5}$	$2.0 \cdot 10^{-6}$	$4.4 \cdot 10^{-7}$
$\tau = 2.0$, GEest_y Extra-cost = 86%	$5.6 \cdot 10^{-2}$	$9.0 \cdot 10^{-3}$	$5.2 \cdot 10^{-5}$	$1.2 \cdot 10^{-6}$	$2.4 \cdot 10^{-7}$
$\tau = 5.0$, GEest_y Extra-cost = 74%	$6.2 \cdot 10^{-2}$	$2.9 \cdot 10^{-3}$	$6.3 \cdot 10^{-5}$	$1.8 \cdot 10^{-6}$	$3.1 \cdot 10^{-8}$
$\tau = 10$, GEest_y Extra-cost = 65%	$1.4 \cdot 10^{-2}$	$4.3 \cdot 10^{-3}$	$1.1 \cdot 10^{-4}$	$2.9 \cdot 10^{-6}$	$1.0 \cdot 10^{-7}$

Also the average of the extra-costs for the integration with tolerance $\tau\delta$ are displayed. The average of failed-steps versus accepted steps was 4.3%.

with $2\pi/\eta$-periodic solutions in space. The exact solution is given by $y(x,t) = \cos(x/4)\cos(t/16)$. By using spatial second order centered differences of five points, and adding a residual term $((r_i(t))_{i=1}^N)$ in order to have the same solution for both, the ODE and the PDE, we get the nonlinear ordinary differential system

$$\left.\begin{array}{l} y_i''(t) + (\Delta x)^{-4}(1+y_i^2)(y_{i-2} - 4y_{i-1} + 6y_i - 4y_{i+1} + y_{i+2}) = \eta^4 y_i^3 + r_i(t), \\ \Delta x = 2\pi\eta^{-1}N^{-1}, \quad y_i(0) = \cos(\eta x_i), \quad y_i'(0) = 0, \quad x_i = -\eta^{-1}\pi + i\Delta x, \\ \quad i = 1, 2, \ldots, N, \end{array}\right\} \tag{5.3}$$

where by periodicity of the solution in space, it holds that

$$y_0(t) = y_N(t), \quad y_{-1}(t) = y_{N-1}(t), \quad y_{N+1}(t) = y_1(t), \quad y_{N+2}(t) = y_2(t), \tag{5.4}$$

and the residual terms are given by

$$r_i(t) = \alpha(1+\varepsilon_i(t)^2)\varepsilon_i(t),\ \varepsilon_i(t) = \cos(\eta x_i)\cos(\eta^2 t), \quad i = 1, 2, \ldots, N,$$

$$\alpha := \frac{1}{8}\sum_{k=1}^{\infty}\frac{(1-4^{-k-1})}{(-4)^k}\frac{(\Delta x)^{2k}}{(2k+4)!}. \tag{5.5}$$

Table 3
Global errors (GE_y) and global errors estimated (GEest_y) with `Gauss2` on the y-component for Problem 3 at $t_{\text{end}} = 200$, for different tolerances δ and factors τ

TOL (δ)	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}
CPU-time	0.05	0.06	0.08	0.09	0.19	0.31
GE_y	$4.7 \cdot 10^{-4}$	$6.2 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$	$2.9 \cdot 10^{-6}$	$4.6 \cdot 10^{-7}$	$7.8 \cdot 10^{-8}$
$\tau = 1.5$, GEest_y Extra-cost = 90%	$1.1 \cdot 10^{-3}$	$1.5 \cdot 10^{-4}$	$1.5 \cdot 10^{-5}$	$4.8 \cdot 10^{-6}$	$5.5 \cdot 10^{-7}$	$8.8 \cdot 10^{-8}$
$\tau = 2.0$, GEest_y Extra-cost = 87%	$8.3 \cdot 10^{-4}$	$9.2 \cdot 10^{-5}$	$2.5 \cdot 10^{-5}$	$2.7 \cdot 10^{-6}$	$4.0 \cdot 10^{-7}$	$8.8 \cdot 10^{-8}$
$\tau = 5.0$,GEest_y Extra-cost = 69%	$8.5 \cdot 10^{-4}$	$9.1 \cdot 10^{-5}$	$1.4 \cdot 10^{-5}$	$2.6 \cdot 10^{-6}$	$4.8 \cdot 10^{-7}$	$8.9 \cdot 10^{-8}$
$\tau = 10$, GEest_y Extra-cost = 62%	$1.1 \cdot 10^{-3}$	$7.7 \cdot 10^{-5}$	$8.9 \cdot 10^{-6}$	$2.3 \cdot 10^{-6}$	$4.5 \cdot 10^{-7}$	$7.2 \cdot 10^{-8}$

Also the average of the extra-costs for the integration with tolerance $\tau\delta$ are displayed. The average of failed-steps versus accepted steps was 2.6%.

Table 4
Global errors and estimated global errors at the y'-component with Gauss2, for Problems 1–3 and $\tau = 5$

	TOL (δ)	10^{-5}	10^{-6}	10^{-7}	10^{-8}
(Prob.1)	GEest_y$'$	$1.4 \cdot 10^{0}$	$1.4 \cdot 10^{-1}$	$4.7 \cdot 10^{-2}$	$5.9 \cdot 10^{-3}$
	GE_y$'$	$7.4 \cdot 10^{-1}$	$1.8 \cdot 10^{-1}$	$3.1 \cdot 10^{-2}$	$4.7 \cdot 10^{-3}$
(Prob.2)	GEest_y$'$	$1.1 \cdot 10^{-3}$	$1.7 \cdot 10^{-4}$	$2.8 \cdot 10^{-5}$	$5.6 \cdot 10^{-6}$
	GE_y$'$	$1.1 \cdot 10^{-3}$	$1.9 \cdot 10^{-4}$	$3.1 \cdot 10^{-5}$	$5.5 \cdot 10^{-6}$
(Prob.3)	GEest_y$'$	$1.2 \cdot 10^{-5}$	$2.3 \cdot 10^{-6}$	$4.2 \cdot 10^{-7}$	$9.0 \cdot 10^{-8}$
	GE_y$'$	$1.4 \cdot 10^{-5}$	$2.5 \cdot 10^{-6}$	$4.4 \cdot 10^{-7}$	$7.8 \cdot 10^{-8}$

For our experiments we have taken $N = 100$. For this problem the eigenvalues of the Jacobian matrix at $t = 0$ are real and negative.

In Tables 1–3 we have displayed a few results obtained with *Gauss2* for the three problems abovementioned. The tolerance for the step $t_n \to t_{n+1}$ for local error control is computed as $\text{tol}_n = (1 + \|y_n\|)\delta$. Hence, the relative and absolute tolerances are the same. GE_y and GEest_y, respectively, denote the global error and the global error estimate on the y-component at the end-point. CPU-time is the time in seconds taken for the main integrations (tolerances δ). τ is the factor considered for the tolerance in subsidiary integrations (tolerances $\tau\delta$). Extra-cost measures the additional CPU-time (in average) for the subsidiary integrations. In Table 4, GE_y$'$ and GEest_y$'$, respectively, denote the global error and the global error estimate on the y'-component at the end-point.

From our numerical experiments we conclude that:

- The global error estimations provided by the tolerance-based global extrapolation technique, provides reliable estimates of the magnitude of the global errors, even for τ-values of moderate size.
- Although the cost of the subsidiary integration decreases when τ increases, the estimates are only reliable for moderate τ-values. A value of $\tau = 5$ can be suggested because there are no significant changes in the accuracy for smaller values of τ.
- The process can be easily parallelized without requiring additional cost.

Acknowledgment

The authors would like to thank the anonymous referees for their kind help and useful comments.

References

[1] R. Aı¨d, L. Levacher, Numerical investigations on global error estimation for ordinary differential equations, J. Comput. Appl. Math. 82 (1–2) (1997) 21–39.

[2] M. Calvo, D.J. Higham, J.I. Montijano, L. Rández, Global error estimation with adaptive explicit Runge–Kutta methods, IMA J. Numer. Anal. 16 (1996) 47–63.

[3] M. Calvo, D.J. Higham, J.I. Montijano, L. Rández, Stepsize selection for tolerance proportionality in explicit Runge–Kutta codes, Adv. Comput. Math. 7 (1997) 361–382.

[4] J.R. Dormand, P.J. Price, A family of embedded Runge–Kutta formulae, J. Comput. Appl. Math. 6 (1980) 19–26.

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

[6] S. Gonzalez-Pinto, R. Rojas-Bello, Gauss2, a Fortran 90 code for second order initial value problems, ⟨http://www.netlib.org/ode/⟩.

[7] S. Gonzalez-Pinto, R. Rojas-Bello, A code based on the two-stage Gauss method for second order initial value problems, preprint in ⟨http://pcmap.unizar.es/numerico/⟩, ACM Trans. Math. Soft., submitted for publication.

[8] E. Hairer, C. Lubich, G. Wanner, Geometric Numerical Integration, Springer, Berlin, 2002.

[9] D.J. Higham, Robust defect control with Runge–Kutta methods schemes, SIAM J. Numer. Anal. 26 (1989) 1175–1183.

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

[11] D.J. Higham, The tolerance proportionality of adaptive ODE solvers, J. Comput. Appl. Math. 45 (1993) 227–236.

[12] D.J. Higham, The reliability of standard local error control algorithms for initial value ordinary differential equations, in: Proceedings: The Quality of Numerical Software: Assessment and Enhancement, IFIP Series, Springer, Berlin, 1997.

[13] A.C. Hindmarsh, ODEPACK, a systematized collection of ODE solvers, Scientific computing, Montreal, Que., 1982, pp. 55–64, IMACS Transactions on Scientific Computation, vol. I, IMACS, New Brunswick, NJ, 1983.

[14] P.J. van der Houwen, B.P. Sommeijer, Diagonally implicit Runge–Kutta Nyström for computing oscillatory problems, SIAM J. Numer. Anal. 26 (2) (1989) 414–429.

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

[16] L.F. Shampine, Tolerance proportionality in ODE codes, in: A. Bellen et al. (Ed.), Numerical Methods for Ordinary Differential Equations, Lecture Notes in Mathematics, vol. 1386, Springer, Berlin, 1989, pp. 118–136.

[17] L.F. Shampine, M.K. Gordon, Computer Solution of Ordinary Differential Equations. The Initial Value Problem, W. H. Freeman and Co., San Francisco, CA, 1975.

[18] R.D. Skeel, Thirteen ways to estimate global error, Numer. Math. 48 (1) (1986) 1–20.

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

[20] H.J. Stetter, Tolerance proportionality in ODE codes, in: R. März (Ed.), Proceedings of the Second Conference on Numerical Treatment of Ordinary Differential Equations, Humbold University, Berlin, 1980, pp. 109–123.