Martin–Luther–Universität
Halle–Wittenberg
Institut für Mathematik

Variable-stepsize doubly quasi-consistent parallel explicit peer methods with global error control

R. Weiner, G. Kulikov and H. Podhaisky

Report No. 05 (2010)

Editors:
Professors of the Institute for Mathematics, Martin-Luther-University Halle-Wittenberg.

Electronic version: see `http://www2.mathematik.uni-halle.de/institut/reports/`

Variable-stepsize doubly quasi-consistent parallel explicit peer methods with global error control

R. Weiner, G. Kulikov and H. Podhaisky

Report No. 05 (2010)

Prof. Dr. Rüdiger Weiner
Dr. Helmut Podhaisky
Martin-Luther-Universität Halle-Wittenberg
Naturwissenschaftliche Fakultät III
Institut für Mathematik
Theodor-Lieser-Str. 5
D-06120 Halle/Saale, Germany
Email: ruediger.weiner@mathematik.uni-halle.de
Email: helmut.podhaisky@mathematik.uni-halle.de

Prof. Dr. Gennady Yu. Kulikov
CEMAT, Instituto Superior Técnico
TU Lisbon
Av. Rovisco Pais
1049-001 Lisboa, Portugal
Email: gkulikov@math.ist.utl.pt

Variable-stepsize doubly quasi-consistent parallel explicit peer methods with global error control

Rüdiger Weiner,* Gennady Yu. Kulikov, † Helmut Podhaisky‡

Abstract

Variable-stepsize explicit peer methods were designed and studied by Weiner et al. in 2008, 2009. Those schemes have proved their high efficiency in practical computations and are considered to be competitive to the best explicit Runge-Kutta embedded pairs. This paper adds more functionality to the mentioned numerical technique in terms of global error estimation and control. Theoretically, it is based on the new concept of double quasi-consistency introduced by Kulikov in 2009. This property means that the principal terms of the local and global errors coincide. In other words, the global error estimation and control can be done effectively via the conventional local error control facility, that is a standard feature of ODE solvers.

Recently, Kulikov and Weiner implemented the idea of double quasi-consistency in fixed-stepsize doubly quasi-consistent parallel explicit peer methods. They also extended that result to non-equidistant meshes by an accurate polynomial interpolation technique. Here, we prove at first that the class of variable-stepsize doubly quasi-consistent methods is not empty and provide first samples of such numerical schemes. Then, we utilize the notion of embedded formulas to evaluate and control efficiently the local error of the constructed doubly quasi-consistent peer methods and, hence, their global error at the same time. Numerical examples of this paper confirm that the usual local error control implemented in doubly quasi-consistent numerical integration techniques is capable of producing numerical solutions for user-supplied accuracy conditions in automatic mode. A comparison with the third order Matlab solver `ode23` is also presented.

Keywords. variable-stepsize doubly quasi-consistent numerical schemes, embedded formulas, local error estimation, automatic global error control.

*Institut für Mathematik, Universität Halle, D-06099 Halle, Germany (`weiner@mathematik.uni-halle.de`).

†CEMAT, Instituto Superior Técnico, TU Lisbon, Av. Rovisco Pais, 1049-001 Lisboa, Portugal (`gkulikov@math.ist.utl.pt`).

‡Institut für Mathematik, Universität Halle, D-06099 Halle, Germany (`podhaisky@mathematik.uni-halle.de`).

1 Introduction

The property of double quasi-consistency was formulated in the class of Nordsieck methods for the first time (see [6]). Unfortunately, Kulikov proved in the cited paper that there exists no doubly quasi-consistent scheme of such kind. Moreover, he stressed that this property cannot be found in any conventional class of numerical methods such as Runge-Kutta schemes, multistep and one-leg formulas. We point out that the property introduced by Kulikov is stronger than the usual quasi-consistency studied in [16], [19]. Nevertheless, Kulikov [6] mentioned also that doubly quasi-consistent schemes are likely to exist in the class of general linear methods, those are of great interest nowadays (see [2] and [5] for a detailed discussion of these methods).

Later, Kulikov and Weiner [12] proved the existence of doubly quasi-consistent schemes in a special family of general linear methods, namely in fixed-stepsize Explicit Parallel Peer (EPP) methods. They also presented an efficient global error estimation formula there and showed how that technique works in practice. Certainly, fixed-stepsize methods themselves are of limited interest. That is why Kulikov and Weiner [13] accommodated their doubly quasi-consistent schemes to variable meshes by means of an accurate polynomial interpolation technique. Strictly speaking, those interpolating EPP methods cannot be doubly quasi-consistent on non-equidistant grids because of the polynomial interpolation involved in computation (see more explanation in the cited paper), but they use this property in the global error estimation and control technique presented there. Thus, our first task in this paper is to prove that variable-stepsize doubly quasi-consistent methods do exist.

Next, we supply such numerical schemes with an efficient error estimation formula. As usual in numerical methods for differential equations, the error estimation here is to be done by the embedded method approach, which has proved its high efficiency in practical computations. We stress that the embedded method error estimation scheme evaluates the local error of the lower order method in an embedded pair. Nevertheless, the property of double quasi-consistency allows two methods of the same convergence order to be used for error estimation (see [12] for more detail). Most importantly, double quasi-consistency ensures that the principle terms of the local and true errors coincide. This means that the conventional embedded method approach is able to evaluate correctly the global error of the doubly quasi-consistent method in the embedded pair for sufficiently small stepsizes. As an implication, our global error estimation scheme does not involve any Jacobian evaluation when implemented in explicit methods. The latter is unavoidable in many other global error estimation formulas (see, for instance, [1], [7], [8], [9], [10], [11], [15], [17], [18], [20]). Thus, the global error evaluation of this paper is cheap and effective in the numerical schemes under discussion.

However, the most important consequence of double quasi-consistency is not the efficiency of global error estimation, but the efficiency of global error control. It is well known that controlling the global error can be very time consuming because of the need to recompute the numerical solution several times over the integration interval. For example, Skeel [17], [18] mentioned that at least two computations over the integration interval are necessary to calculate a numerical solution for a specific accuracy requirement. He also argued in the cited papers that the cost of global error control was the main obstacle for using it

commonly in ODE solvers. This is no longer the case in doubly quasi-consistent schemes because the standard local error control monitors and regulates the global error of a doubly quasi-consistent method at the same time. In other words, the global error control can be achieved only for one computation over the integration interval. This is the most important feature of doubly quasi-consistent methods (see more explanation in [13]).

The remainder of this paper is organized as follows: In Section 2 we formulate a definition of variable-stepsize doubly quasi-consistent EPP-methods, prove their existence and present a corresponding method of order 2. Section 3 focuses on our error estimation technique based on the embedded method approach. We build two embedded pairs of variable-stepsize 4-stage EPP-methods of orders 3 and 4 that are suitable for practical use. Numerical tests that confirm the power of these new explicit numerical schemes in the sense of efficiency of the global error estimation and control are presented in Section 4. There, a comparison with the built-in Matlab ODE solver `ode23` is also included. The last section summarizes results obtained in this paper and outlines future plans.

2 Variable-stepsize doubly quasi-consistent EPP methods

Here, we explore double quasi-consistency in the class of variable-stepsize explicit two-step peer methods introduced and studied by Weiner et al. [21], [22]. When applied to ordinary differential equations (ODEs) of the form

$$x'(t) = g\big(t, x(t)\big), \quad t \in, \quad x(t_0) = x^0, \tag{1}$$

where $x(t) \in \mathbb{R}^m$ and $g : D \subset \mathbb{R}^{m+1} \to \mathbb{R}^m$ is a sufficiently smooth function, those numerical schemes read

$$x_{ki} = \sum_{j=1}^{s} b_{ij}(k) x_{k-1,j} + \tau_k \sum_{j=1}^{s} a_{ij}(k) g(t_{k-1,j}, x_{k-1,j}) + \tau_k \sum_{j=1}^{i-1} r_{ij}(k) g(t_{kj}, x_{kj}) \tag{2}$$

where $t_{ki} := t_k + c_i\tau_k$, $i = 1, 2, \ldots, s$, and τ_k is the current stepsize. The extra nodes t_{ki} of the integration mesh are fixed by the constants c_i. Notice that the coefficients $a_{ij}(k)$, $b_{ij}(k)$ and $r_{ij}(k)$ of this method depend in general on the stepsize ratio $\theta_k := \tau_k/\tau_{k-1}$.

Method (2) can be easily represented in the matrix form

$$X_k = (B(k) \otimes I_m) X_{k-1} + \tau_k (A(k) \otimes I_m) g(T_{k-1}, X_{k-1}) + \tau_k (R(k) \otimes I_m) g(T_k, X_k) \tag{3}$$

where I_m is the identity matrix of dimension m and $\otimes$ denotes the Kronecker tensor product (see, for example, [14]). Here, we have utilized the following notation:

$$T_k := (t_{ki})_{i=1}^{s}, \quad X_k := (x_{ki})_{i=1}^{s}, \quad g(T_k, X_k) := g(t_{ki}, x_{ki})_{i=1}^{s},$$
$$A(k) := \big(a_{ij}(k)\big)_{i,j=1}^{s}, \quad B(k) := \big(b_{ij}(k)\big)_{i,j=1}^{s}, \quad R(k) := \big(r_{ij}(k)\big)_{i,j=1}^{s}.$$

We remark that the matrix $R(k)$ is strictly lower triangular because the methods (3) explored in this paper are explicit. Moreover, these peer methods allow for a simple and convenient parallelization by setting $R(k) = 0$. We will further restrict ourselves to this particular case.

Weiner et al. [21], [22] explained benefits of using a constant matrix B in the methods of the form (3). We impose the same restriction on the numerical schemes discussed below.

In what follows, we deal with s-stage *Explicit Parallel Peer (EPP) methods* of the form

$$x_{ki} = \sum_{j=1}^{s} b_{ij} x_{k-1,j} + \tau_k \sum_{j=1}^{s} a_{ij}(k) g(t_{k-1,j}, x_{k-1,j}), \quad i = 1, 2, \ldots, s, \tag{4}$$

or in the matrix form

$$X_k = (B \otimes I_m) X_{k-1} + \tau_k (A(k) \otimes I_m) g(T_{k-1}, X_{k-1}). \tag{5}$$

Notice that method (4) (or (5)) is variable-stepsize, but only the matrix $A(k)$ can depend on stepsize ratio θ_k, here.

The principal feature of peer methods is the fact that all entries of the solution vector X_k possess similar properties in terms of stability and accuracy of numerical integration. This means that any stage value of a peer method can be taken as the output solution. Further, we define this property more formally.

Definition 1 *The vector-function* $L(T_k, x(t), \tau_k) := \big(L_i(t_{ki}, x(t), \tau_k)\big)_{i=1}^{s}$ *where the entries satisfy*

$$L_i(t_{ki}, x(t), \tau_k) := x(t_{ki}) - \sum_{j=1}^{s} \Big(b_{ij} x(t_{k-1,j}) + \tau_k a_{ij}(k) g\big(t_{k-1,j}, x(t_{k-1,j})\big) \Big) \tag{6}$$

is referred to as the defect of the EPP method (4).

Formula (6) is derived by substitution of the exact solution $x(t)$ evaluated at the mesh points into numerical scheme (4).

Definition 2 *The variable-stepsize EPP method* (4) *is consistent of order* p *when the following order conditions hold:*

$$\mathcal{AB}_i(l) := c_i^l - \sum_{j=1}^{s} \left(b_{ij} \frac{(c_j - 1)^l}{\theta_k^l} + l\, a_{ij}(k) \frac{(c_j - 1)^{l-1}}{\theta_k^{l-1}} \right) = 0, \quad l = 0, 1, \ldots, p, \quad i = 1, 2, \ldots, s, \tag{7}$$

for any stepsize ratio $\theta_k := \tau_k / \tau_{k-1}$.

As customary, order conditions (7) are obtained by the Taylor expansion of the defect (6) around the grid point t_k. It is also clear from Definition 2 that fixing entries of the matrix B reduces the consistency order of EPP methods (4) because only entries of the matrix $A(k)$ can be utilized to satisfy a certain number of order conditions (7). To our advantage, this allows to obtain easily numerical schemes that are zero-stable on any variable grid (see [21], [22]). In addition, we require in this paper that no order condition $\mathcal{AB}_i(p+1) = 0$, $i = 1, 2, \ldots, s$, is satisfied for a peer method (4) of the consistency order p.

Definitions 1 and 2 say that $L(T_k, x(t), \tau_k) = \mathcal{O}(\tau_k^{p+1})$. The latter suffices obviously for convergence of order p. However, the property of zero-stability is necessary. This means for

constant matrices B that all powers of B are uniformly bounded. A sufficient condition is that the spectral radius of this matrix is less than or equal to one, i.e. $\varrho(B) \leq 1$, and all eigenvalues of modulus one are simple. Notice that the definition of zero-stability of EPP methods (5) does not depend on a grid. In other words, it will work for any variable mesh. This is the main advantage of fixing the matrix B in the EPP methods under discussion. In what follows, we consider that all methods (5) satisfy this zero-stability condition.

We point out that the consistency of order p is not necessary for convergence of order p. For instance, Weiner et al. [22] found numerical schemes in the class of peer methods (2) that require less consistency for that. In general, that result is based on the idea of quasi-consistency (see [16]) that can be formulated for the EPP methods explored in this paper as follows:

Definition 3 *The variable-stepsize EPP method* (5) *is quasi-consistent of order p when the conditions*

$$\mathcal{AB}(l) = 0, \quad l = 0, 1, \ldots, p-1, \tag{8a}$$

$$B \cdot \mathcal{AB}(p) = 0, \tag{8b}$$

where the vector $\mathcal{AB}(l) := \left(\mathcal{AB}_i(l)\right)_{i=1}^{s}$, hold for any stepsize ratio θ_k.

Notice that the latter definition deals with the form (5) of EPP methods.

Now we extend formally the definition of double quasi-consistency introduced by Kulikov and Weiner [12] for fixed-stepsize EPP-methods to the variable-stepsize methods of the form (5).

Definition 4 *The variable-stepsize EPP method* (5) *is doubly quasi-consistent of order p when the following conditions:*

$$\mathcal{AB}(l) = 0, \quad l = 0, 1, \ldots, p-1, \tag{9a}$$

$$B \cdot \mathcal{AB}(p) = 0, \quad B \cdot \mathcal{AB}(p+1) = 0, \quad A(k) \cdot \mathcal{AB}(p) = 0 \tag{9b}$$

hold for any stepsize ratio θ_k.

Notice that Definition 4 is more general than that introduced earlier in [12] for the fixed-stepsize EPP methods. Both definitions coincide for any equidistant mesh, but only Definition 4 works for variable meshes because of the variable matrix $A(k)$ in formula (9b). Now we prove that conditions (9) are sufficient to retain the main property of double quasi-consistency on variable grids. This implies that the principal term of the true error of method (5) is equal to the principal term of its local error. We stress that, by definition, the local error of the variable-stepsize EPP-method and its defect coincide (see Definition 1). Thus, it suffices to prove this property for the defect of method (5).

Theorem 1 *Let the right-hand side of ODE* (1) *be sufficiently smooth in a neighborhood of the exact solution $x(t)$ on the integration interval $[t_0, t_{end}]$ and the variable-stepsize EPP-method* (5) *of order p be zero-stable and satisfy conditions* (9)*. Suppose additionally that its coefficient matrix $A(k)$ is bounded for any variable mesh and the starting vector X_0 is known with an error of $\mathcal{O}(\tau_0^{p+1})$ where τ_0 is the initial stepsize. Then the true error*

$\Delta X_k := \big((\Delta x_{k1})^T, \ldots, (\Delta x_{ks})^T\big)^T$, $\Delta x_{ki} := x(t_{ki}) - x_{ki}$, $i = 1, 2, \ldots, s$, *of the variable-stepsize EPP-method* (5) *at the mesh points* T_k *and its defect* $L(T_k, x(t), \tau_k)$ *at the same grid nodes satisfy*

$$\Delta X_k = L(T_k, x(t), \tau_k) + \mathcal{O}(\tau^{p+1}) \tag{10}$$

for any variable mesh $w_\tau := \{t_{k+1} = t_k + \tau_k,\ k = 0, 1, \ldots, K-1,\ t_K = t_{end}\}$ *with a sufficiently small diameter* $\tau := \max_{0 \le k \le K-1}\{\tau_k\}$.

Proof 1 *First, having repeated lines of the proof of Theorem 4 in [12] we arrive at the asymptotic formula*

$$\Delta X_k = (B \otimes I_m)\Delta X_{k-1} + \tau_k\big(A(k) \otimes \partial_x g(t_k, x(t_k))\big)\Delta X_{k-1} + L(T_k, x(t), \tau_k) + \mathcal{O}(\tau_k^{p+2}) \tag{11}$$

where $\partial_x g(t_k, x(t_k))$ *denotes the partial derivative (Jacobian) of the mapping* g *with respect to the second argument and evaluated at the point* $(t_k, x(t_k))$. *Notice that all Jacobians have been expanded around the point* $(t_k, x(t_k))$ *with an accuracy of* $\mathcal{O}(\tau_k)$. *Then, we take the order of the peer method into account to yield this error equation.*

Second, we expand all entries of the defect $L(T_k, x(t), \tau_k)$ *of the* s*-stage variable-stepsize EPP-method (5) into the Taylor series*

$$L_i(T_k, x(t), \tau_k) = \mathcal{AB}_i(p)\frac{\tau_k^p}{p!}x^{(p)}(t_k) + \mathcal{AB}_i(p+1)\frac{\tau_k^{p+1}}{(p+1)!}x^{(p+1)}(t_k) + \mathcal{O}(\tau_k^{p+2}), \quad i = 1, 2, \ldots, s, \tag{12}$$

at the node t_k. *Here, we have taking into account condition (9a). With the use of the new notation* $V(k) := B \otimes I_m + \tau_k\big(A(k) \otimes \partial_x g(t_k, x(t_k))\big)$, *(9b), (11) and (12) result in the following chain of asymptotic formulas by induction:*

$$\begin{aligned}
\Delta X_k &= V(k)\Delta X_{k-1} + L(T_k, x(t), \tau_k) + \mathcal{O}(\tau_k^{p+2}) \\
&= V(k)\Big(V(k-1)\Delta X_{k-2} + L(T_{k-1}, x(t), \tau_{k-1}) + \mathcal{O}(\tau_{k-1}^{p+2})\Big) + L(T_k, x(t), \tau_k) + \mathcal{O}(\tau_k^{p+2}) \\
&= \prod_{l=0}^{1} V(k-l)\Delta X_{k-2} + L(T_k, x(t), \tau_k) + \mathcal{O}(\tau_{k-1}^{p+2}) + \mathcal{O}(\tau_k^{p+2}) \\
&= \prod_{l=0}^{2} V(k-l)\Delta X_{k-3} + L(T_k, x(t), \tau_k) + \mathcal{O}(\tau_{k-2}^{p+2}) + \mathcal{O}(\tau_{k-1}^{p+2}) + \mathcal{O}(\tau_k^{p+2}) \\
&\ \ \vdots \\
&= \prod_{l=0}^{k-1} V(k-l)\Delta X_0 + L(T_k, x(t), \tau_k) + \sum_{l=0}^{k-1} \mathcal{O}(\tau_{k-l}^{p+2}).
\end{aligned}$$

Third, the conditions of Theorem 1 transform the last formula in the above chain to the form

$$\Delta X_k = \mathcal{O}(\tau_0^{p+1}) + L(T_k, x(t), \tau_k) + \mathcal{O}(\tau^{p+1}) \sum_{l=0}^{k-1} \tau_{k-l},$$

that proves asymptotic formula (10) and, hence, the theorem. ■

Notice that asymptotic formula (10) does not ensure double quasi-consistency because it does not guarantee that the local error at mesh points T_k will tend to the true error as $\tau_k \to 0$. The second problem with Theorem 1 is that the condition for the coefficient matrix $A(k)$ to be bounded for any variable mesh is too restrictive in practice. It is sufficient to require this on a subset of variable grids, only. Thus, for both reasons, we impose extra conditions on the grids w_τ utilized below.

Definition 5 *For given constants ω_1, ω_2 with the property $\omega_1 \le 1 \le \omega_2$ and Ω, a variable mesh $w_\tau := \{t_{k+1} = t_k + \tau_k,\ k = 0, 1, \ldots, K-1,\ t_K = t_{end}\}$ with a diameter $\tau := \max_{0 \le k \le K-1}\{\tau_k\}$ is called admissible if the following conditions hold:*

$$0 < \omega_1 \le \theta_k \le \omega_2 < \infty, \quad k = 1, 2, \ldots, K-1, \tag{13a}$$

$$\tau/\tau_k \le \Omega < \infty, \quad k = 0, 1, \ldots, K-1. \tag{13b}$$

Theorem 2 *Let the ODE* (1) *and the variable-stepsize EPP-method* (5) *of order p satisfy all conditions mentioned in the formulation of Theorem 1. Then this method is doubly quasi-consistent on any admissible grid w_τ with a sufficiently small diameter τ.*

Proof 2 *The double quasi-consistency condition*

$$\Delta X_k = L(T_k, x(t), \tau_k) + \mathcal{O}(\tau_k^{p+1}) \tag{14}$$

follows from (10) and (13b) at once. Theorem 2 is proved. ■

Remark 1 *Condition (13b) is not restrictive in practice because it means that the ratio of the maximum stepsize to the minimum stepsize is limited. The latter is the case for conventional ODE solvers (see, for instance, [7] or [8] for further explanation).*

To prove the existence of variable-stepsize doubly quasi-consistent numerical schemes, we fix the number of stages $s = 4$, the order of the method $p = 2$ and, with the use of conditions (9), compute in `Maple 11` the following variable-stepsize EPP-method (5):

$$B = \begin{pmatrix} \frac{1}{6} & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}, \qquad c = \begin{pmatrix} 0 \\ \frac{1}{4} \\ \frac{1}{2} \\ 1 \end{pmatrix}. \tag{15a}$$

$$a_{11}(k)=\frac{1-24\theta_k+12\theta_k^2}{96\theta_k},\quad a_{12}(k)=\frac{0.3125}{\theta_k},\quad a_{13}(k)=0.5,$$
$$a_{14}(k)=-\frac{-29+24\theta_k+12\theta_k^2}{96\theta_k},\quad a_{21}(k)=\frac{-39+37\theta_k+62\theta_k^2+50\theta_k^3}{192\theta_k},\quad a_{22}(k)=\frac{0.0625(2\theta_k+5)}{\theta_k},$$
$$a_{23}(k)=-\frac{-41+55\theta_k+92\theta_k^2+50\theta_k^3}{96\theta_k},\quad a_{24}(k)=\frac{17+97\theta_k+122\theta_k^2+50\theta_k^3}{192\theta_k},\quad a_{31}(k)=-\frac{-1+30\theta_k}{96\theta_k},$$
$$a_{32}(k)=\frac{0.0625(4\theta_k+5)}{\theta_k},\quad a_{33}(k)=0.25,\quad a_{34}(k)=\frac{29+30\theta_k}{96\theta_k},$$
$$a_{41}(k)=-\frac{-1+42\theta_k+36\theta_k^2}{96\theta_k},\quad a_{42}(k)=\frac{0.0625(8\theta_k+5)}{\theta_k},\quad a_{43}(k)=0.125,\qquad (15b)$$
$$a_{44}(k)=\frac{29+78\theta_k+36\theta_k^2}{96\theta_k}.$$

The magnitudes of the constants at all entries of the principal term of the local error are the same and equal to $1/4$. More precisely, $\mathcal{AB}(2)=(1/4,-1/4,1/4,1/4)^T$. That is all stage values of method (15) are equally accurate and stable. The latter follows from the optimal stability condition of this numerical scheme on any variable mesh.

As we mentioned earlier, the method's double quasi-consistency follows from conditions (9), but for admissible grids only (see Definition 5) because the coefficient matrix (15b) is not bounded in general. Therefore condition (13a) is necessary for both the convergence and the double quasi-consistency of method (15). This double quasi-consistency will be examined numerically in Section 4.

Thus, we have proved that variable-stepsize doubly quasi-consistent numerical schemes do exist. Further, we intend to utilize this property for an efficient global error estimation and control. We stress that double quasi-consistency allows the global error to be controlled effectively, i.e. for one integration over the interval $[t_0, t_{end}]$.

3 Global error control in the variable-stepsize doubly quasi-consistent EPP-method

In this section, we discuss a cheap global error estimation mechanism. It can be done in the doubly quasi-consistent peer method presented above via understanding that the principal terms of its local and true errors coincide. We use an s-stage embedded EPP-method to yield an asymptotic error estimate at any mesh point. Thus, we follow here the idea of global error estimation presented for fixed-stepsize doubly quasi-consistent peer methods in [12].

Having used an embedded peer method (5) with coefficients $A(k)_{emb}$, B_{emb} and $c_{emb}=c$, we arrive at the error evaluation scheme of the form

$$\Delta_1 X_k = X_{k,emb}-X_k = \big((B_{emb}-B)\otimes I_m\big)X_{k-1}+\tau_k\big((A(k)_{emb}-A(k))\otimes I_m\big)g(T_{k-1},X_{k-1}) \quad (16)$$

where $\Delta_1 X_k$ denotes the principal term of the true error of the doubly quasi-consistent peer method and X_{k-1} implies the numerical solution computed by the same peer method. We see that the global error estimation formula (16) is cheap in practice because it is a linear combination of the values known from method (15). However, our error estimation will work correctly for the numerical solution from the embedded EPP-scheme substituted into the

right-hand side of (16) as well if the embedded method is at least one order more accurate (see Remark 2 below for more details).

It is also worthwhile to notice that the global error estimation strategy (16) does not require any Jacobian evaluation, which is needed in many other global error estimation schemes (see, for instance, [1], [8], [9], [10], [13], [15], [17], [18], [20]). Certainly, the evaluation of the Jacobian of ODEs is natural in implicit numerical methods and, hence, can be utilized effectively in global error estimation techniques. However, it is unacceptable in explicit numerical methods, which do not require any Jacobian computation.

As in [12], we have to impose conditions on the embedded peer method utilized in the error estimation scheme (16) to ensure its correctness.

Theorem 3 *Let the ODE* (1) *and the variable-stepsize doubly quasi-consistent EPP-method* (5) *of order p satisfy all conditions mentioned in the formulation of Theorem 1. Suppose that an embedded EPP-method of the form* (5) *and with the coefficients* $A(k)_{emb}$, B_{emb}, c *bounded on any admissible mesh* (13) *is used in the error estimation scheme* (16). *Then formula* (16) *computes the principal term of the true error of the variable-stepsize doubly quasi-consistent EPP-method* (5) *on any admissible grid* w_τ *with a sufficiently small diameter* τ *if and only if the coefficients* $A(k)_{emb}$, B_{emb} *and* c *of the embedded method satisfy the following conditions:*

$$\mathcal{AB}(l)_{emb} = 0, \quad l = 0, 1, \ldots, p, \tag{17a}$$

$$B_{emb} \cdot \mathcal{AB}(p) = 0 \tag{17b}$$

where the vectors $\mathcal{AB}(l)_{emb}$, $l = 0, 1, \ldots, p$, *are calculated for the coefficients of the embedded formula* (5) *and the vector* $\mathcal{AB}(p)$ *is evaluated for the coefficients of the doubly quasi-consistent peer method in the embedded pair.*

The proof of Theorem 3 can be done in the same way as the proof of Theorem 5 in [12]. Here, we have only to take into account the boundedness of the variable matrices $A(k)$ and $A(k)_{emb}$ for any admissible mesh w_τ and the double quasi-consistency condition (14).

Remark 2 *It is clear from Theorem 3 that the error estimation formula (16) is correct even for embedded variable-stepsize EPP-schemes (5) of the same order as that of the doubly quasi-consistent EPP-method in the pair, provided that conditions (17) hold. Certainly, this error evaluation technique will work for any higher order embedded variable-stepsize EPP-method (5) if we continue with the higher order numerical solution and utilize this solution on the right-hand side of the error estimation formula (16).*

Remark 3 *In practice, condition (17b) imposed on the coefficients of the embedded variable-stepsize EPP-scheme (5) is simplified to the form*

$$B_{emb} = B. \tag{18}$$

This increases the accuracy of the error estimation (16). On the other hand, it is evident that the matrix B_{emb} *determined by condition (18) satisfies Theorem 3 as well. This follows from Definition 4.*

Now we construct embedded EPP-schemes for the method (15). Again, the number of stages is $s = 4$. We present below two embedded EPP-schemes of orders $p = 3$ and $p = 4$ and with the matrices B_{emb} are given by formula (18). In other words, these matrices are fixed by (15a). For each nonzero vector β, taking into account $R = 0$ formula

$$A(k)_{emb} = \left(CV_0D^{-1} - \frac{1}{4}\beta e_4^T \right) S(k)V_1^{-1} - \frac{1}{\theta_k} B_{emb}(C - I_4)V_1D^{-1}V_1^{-1} \tag{19}$$

defines an embedded method of order 3. When $\beta = 0$ we have the uniquely defined method of order 4, cf. [21]. Here $e_4 := (0, 0, 0, 1)^T$,

$$C := \mathrm{diag}(c_1, c_2, c_3, c_4), \quad D := \mathrm{diag}(1, 2, 3, 4), \quad S(k) := \mathrm{diag}(1, \theta_k, \theta_k^2, \theta_k^3),$$

and

$$V_0 := \left(c_i^{j-1}\right)_{i,j=1}^4, \quad V_1 := \Big(\big(c_i - 1\big)^{j-1}\Big)_{i,j=1}^4.$$

4 Numerical tests

In this section, we present numerical results of tests in Matlab for our doubly quasi-consistent method (15).

We use the following test problems:

Problem I: (from [6]).

$$\begin{aligned} x_1' &= 2tx_2^{1/5}x_4, \quad & x_2' &= 10t\exp\Big(5\big(x_3 - 1\big)\Big)x_4, \\ x_3' &= 2tx_4, & x_4' &= -2t\ln\big(x_1\big) \end{aligned}$$

where $t \in [0, t_e]$ and $x(0) = (1, 1, 1, 1)^T$. The exact solution is

$$x_1 = \exp\big(\sin t^2\big), \quad x_2 = \exp\big(5\sin t^2\big), \quad x_3 = \sin t^2 + 1, \quad x_4 = \cos t^2.$$

Global error control for this problem is hard for larger values of t_e since the derivatives of the solution become large. We use $t_e = 2, 3, 4$ in our experiments, below.

Problem II: This problem is defined by

$$x_1' = \frac{x_4^4}{x_2} - x_1^2 - x_3^2 - x_3, \quad x_2' = x_4^4 - 3x_2, \quad x_3' = x_1, \quad x_4' = -0.5x_2^{1/4}$$

with $x_0 = (1, 1, 0, 1)^T$ and $t \in [0, 10]$. The exact solution is given by

$$x_1 = \cos t, \quad x_2 = e^{-2t}, \quad x_3 = \sin t, \quad x_4 = e^{-0.5t}.$$

Problem III: (the Kepler problem in [3, p. 86]):

$$x_1' = x_3, \quad x_2' = x_4, \quad x_3' = \frac{-x_1}{(x_1^2 + x_2^2)^{3/2}}, \quad x_4' = \frac{-x_2}{(x_1^2 + x_2^2)^{3/2}}$$

with initial conditions $x_0 = \left(1-\varepsilon, 0, 0, \sqrt{\frac{1+\varepsilon}{1-\varepsilon}}\right)^T$. This problem is to be solved for $t_e = 20$. We consider two cases when its parameter ε is chosen to be 0 or 0.9. The exact solution of the Kepler problem is well known and presented by the formulas

$$x_1 = \cos\varepsilon_1 - \varepsilon, \quad x_2 = \sqrt{1-\varepsilon^2}\sin\varepsilon_1, \quad x_3 = \frac{\sin\varepsilon_1}{\varepsilon\cos\varepsilon_1 - 1}, \quad x_4 = \frac{\sqrt{1-\varepsilon^2}\cos\varepsilon_1}{1-\varepsilon\cos\varepsilon_1}.$$

Here, for each point t, ε_1 is defined as the solution of the nonlinear equation

$$\varepsilon_1 - \varepsilon\sin\varepsilon_1 = t.$$

It is trivial for $\varepsilon = 0$. However, we solve this equation at each time step in order to determine the global error of the methods under examination when $\varepsilon = 0.9$.

Problem IV: (the Arenstorf orbit, problem AREN in [4]):

$$\begin{aligned} x_1' &= x_2, \quad x_2' = x_1 + 2x_4 - m_1(x_1+m)/D_1 - m(x_1-m_1)/D_2, \\ x_3' &= x_4, \quad x_4' = x_3 - 2x_2 - m_1 x_3/D_1 - m x_3/D_2 \end{aligned}$$

where $m = 0.012277471$, $m_1 = 1-m$, $D_1 = ((x_1+m)^2+x_3^2)^(3/2)$, $D_2 = ((x_1-m_1)^2+x_3^2)^(3/2)$ and with initial values $x_0 = (0.994, 0, 0, -2.00158510637908252240537862224)^T$.

Its exact solution has no closed form, but the solution-path of this problem is periodic with the period $T = 17.065216560157962558891$ (see [4, p. 129, 130] for more details). Thus, we merely monitor the error at the point $t_e = T$ to verify the quality of numerical solutions computed.

Our two-step method requires starting values for the first step. They are taken from the exact solution if known, otherwise they are computed by the Matlab solver `ode45`.

In our tests we observed that roundoff errors can have a significant influence on the accuracy of computation when the number of integration steps is large. Their influence is much stronger than in one-step methods due to the term BX_{k-1}, in contrast to simply x_{k-1} in Runge-Kutta methods. To reduce this influence, we have implemented the peer methods with *compensated summation*. An explanation of this technique in the context of Runge–Kutta methods can be found in [2], but we need some adaption for peer methods: We start with the observation that there are three different sources of roundoff errors in recursion (5): 1) the product $(B \otimes I_m)X_{k-1}$, 2) the summation of the right-hand side term $\tau_k g(T_{k-1}, X_{k-1})$ and 3) the evaluation of the latter expression. When $\tau_k \to 0$ the dominant part will come from the first two operations while the third one can be neglected. The first operation, i.e. the product of the matrix B and the vector X_{k-1} can be computed more accurately after a basis change since B has only one nonzero eigenvalue 1 with eigenvector $(1, 1, \ldots, 1)^T$. In our implementation, we use the transformation: $Z_k := (V_1^{-1} \otimes I_m)X_k$ and $\widetilde{B} := V_1^{-1}BV_1$ where the matrix V_1 is defined at the end of Section 3 in this paper. Then, the matrix $\widetilde{B}$ has the dominant eigenvector $e_1 := (1, 0, \ldots, 0)^T$ reflecting the fact that only the first entry Z_{1k} approximates the solution and that all other entries approximate derivatives, i.e. $\mathcal{O}(\tau_k)$-terms. After this preparation, we can apply the usual compensated summation to the

first component Z_{1k}, as follows:

$$\begin{aligned} \phi_k &= (\widetilde{B}_{1,2:s} \otimes I_m) Z_{k,2:s} + \tau_k (e_1^T V_1^{-1} A(k) \otimes I_m) g(T_k, X_k), \\ Z_{1,k+1} &= Z_{1,k} + (\phi_k + err_{k-1}), \\ err_k &= (\phi_k + err_{k-1}) - (Z_{1,k+1} - Z_{1,k}). \end{aligned} \tag{20}$$

err_k in (20) is always zero in the exact arithmetic, but when a floating point arithmetic (with no compiler optimizations) is used it will contain some part of the roundoff error. This error is added to ϕ_{k+1} in the next step. The above compensated summation mechanism reduces roundoff errors of our computations when very stringent tolerances utilized by several magnitudes. So we recommend this technique for practical use.

The local error estimation (and, consequently, the global error estimation) in the doubly quasi-consistent method (15) is done by the embedded EPP-method of order 3 defined by the vector $\beta = (1/40, 1/40, 1/40, 1/40)^T$ substituted in formula (19). We continue integrations with the numerical solution from the above-mentioned doubly quasi-consistent method of order 2. In all figures this numerical scheme is denoted by `dqc2(3)`.

On the other hand, it is a usual practice to continue integration with the higher order numerical solution in majority of embedded Runge-Kutta formulas. This is natural because of two reasons: First, this allows for a correct theoretical substantiation of the error estimation scheme. Second, the resulting global error is expected to be smaller for the higher order method. Therefore, we also test this variant in our code:

- We continue with the embedded method of order 3 and present outcome results marked by `dqc3(2)` in the figures, below.
- We use the numerical solution from the method of order 4 ($\beta = 0$ in (19)) in our error estimation formula (16) and continue with the same numerical solution. The results are denoted by `dqc4(2)` for the latter computation scheme.

We estimate the local error corresponding to (16) by

$$EST = \|X_{k,emb} - X_k\|_\infty.$$

The new stepsize is then determined by the formula

$$\tau_{new} = \min \left\{ 1.5, \max \left\{ 0.5, 0.9 \left(\frac{TOL}{EST} \right)^{0.5} \right\} \right\} \tau_k,$$

where TOL is the prescribed tolerance. When $EST \leq TOL$ we will proceed the integration with $\tau_{k+1} = \tau_{new}$. Otherwise, the current step will be repeated for $\tau_k = \tau_{new}$.

We compare our methods with the Matlab code `ode23`, which is an embedded Runge-Kutta pair of order 3(2). All test problems are solved for $TOL = 10^{-2}, 10^{-3}, \ldots, 10^{-10}$. We use the standard stepsize control with $RelTol = AbsTol = TOL$ in the built-in solver `ode23`. The starting stepsize is computed by the formula $\tau_0 = \min\{10^{-4}, TOL\}$.

For problems with known exact solution x_{exact} we compute the exact error at the point t_k by the formula

$$ERR_k = \|x_{exact} - x_{ks}\|_\infty$$

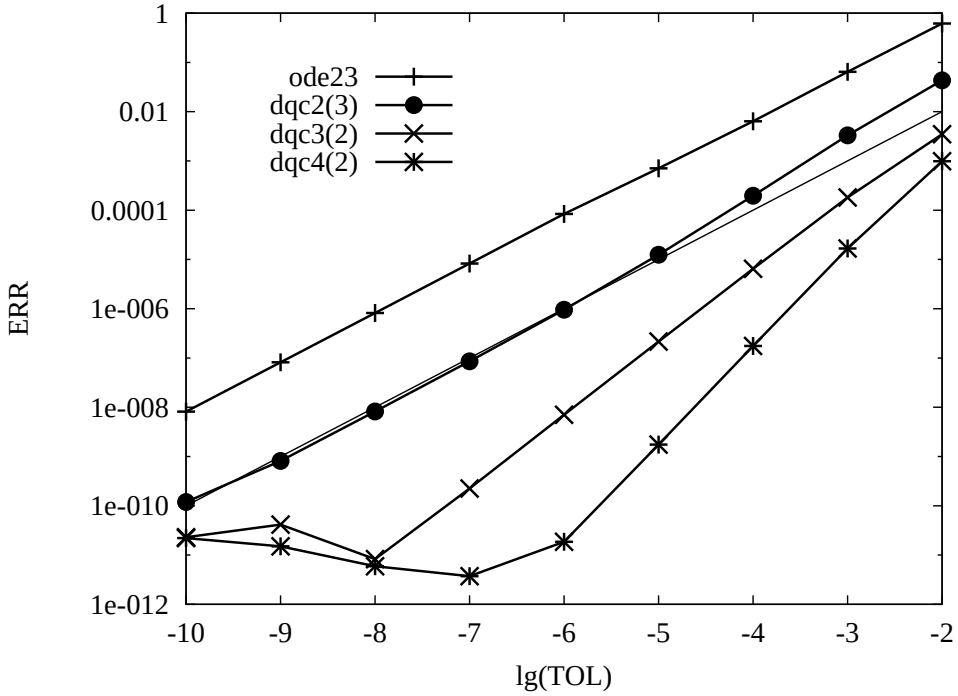

Figure 1: Results for problem I, $t_e = 2$.

for the adaptive scheme `dqc2(3)` or

$$ERR_k = \|x_{exact} - x_{emb,ks}\|_\infty$$

for the embedded pairs denoted by `dqc3(2)` and `dqc4(2)`. The Figures 1–7 show the global errors of the numerical solutions obtained, i.e.

$$ERR = \max_k \{ERR_k\}.$$

The exact error at the end point is displayed for Problem IV. In all figures a solid black line corresponds to $ERR = TOL$.

In Figure 3, marks are missing for the EPP methods for $TOL = 10^{-10}$ because the number of steps exceeded the maximal number which we fixed by 3000000 steps. For Problem III, we omit the result for the embedded EPP-pairs with $TOL = 10^{-10}$ when $\varepsilon = 0.9$ since the computation of the exact solution at each grid point becomes too expensive.

The results are in good agreement with our theoretical prediction:

- The global error of `dqc2(3)` converges for more stringent tolerances to the prescribed tolerance.

- The accuracy curve of `ode23` is in general in parallel to the line $ERR = TOL$, which is shown as a strait line in all figures of this paper, except for the last one. The factor between the prescribed and achieved accuracies obtained for this Matlab solver is very large for some problems and does not become smaller even for more strict tolerances.

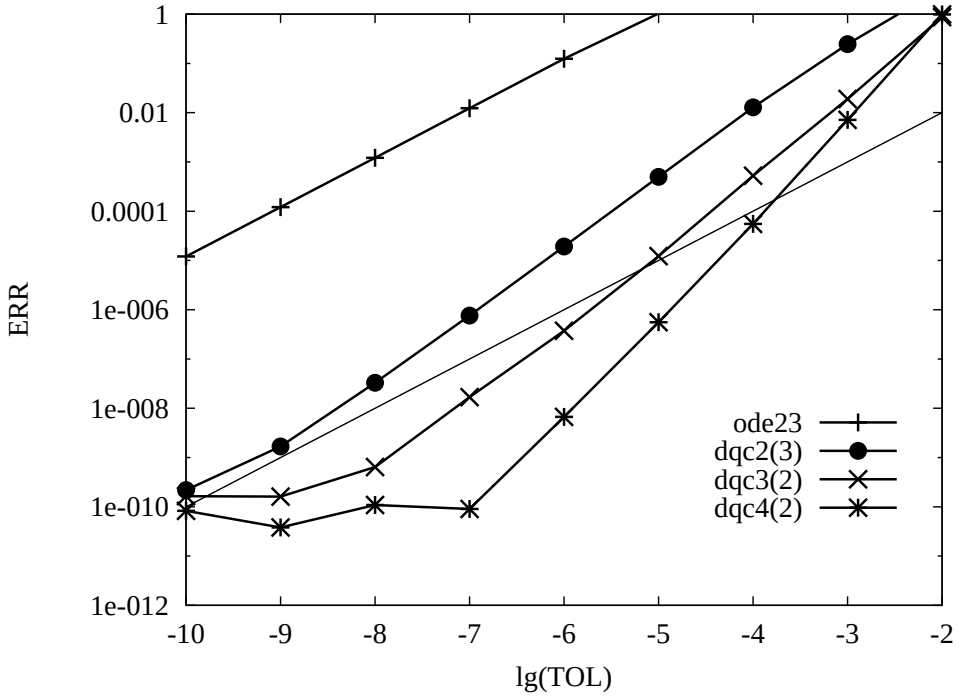

Figure 2: Results for problem I, $t_e = 3$.

- As expected, the results from the higher order EPP-methods `dqc3(2)` and `dqc4(2)` are more accurate than from `dqc2(3)`.

We point out that (14) does not guarantee that the global error will always match to the given accuracy because the $\mathcal{O}$-term depends on higher derivatives of the exact solution and on the length of the integration interval. However, the double quasi-consistency condition (14) guarantees that for strict tolerances, implying $\tau_k \to 0$, the achieved accuracy of numerical integration tends to the prescribed tolerance. This is in contrast to the results obtained for the Matlab solver `ode23`. Notice that both methods (EPP-schemes and Runge-Kutta formulas) use the same stepsize control via estimating the local error.

For very sensitive problems the convergence to the required accuracy of computation can be slow. As example, we present in Figure 8 the result for the Lorenz attractor (see problem LRNZ in [4])

$$x_1' = \sigma(x_2 - x_1), \quad x_2' = -x_1 x_3 + r x_1 - x_2, \quad x_1 x_2 - b x_3$$

with $\sigma = 10$, $r = 28$, $b = 8/3$, $t_e = 16$ and $x_0 = (-8, 8, 27)^T$.

The doubly quasi-consistent computational scheme `dqc2(3)` requires 4 function calls per step and it is of order $p = 2$. Thus, this is more expensive than `ode23`, whose outcome numerical solution is of order 3. However, our code is in much better agreement with the prescribed tolerance than the built-in one.

Figure 9 shows the number of function calls against the achieved accuracy of integration for all methods applied to Problem I with $t_e = 4$. The behavior of the efficiency graphs for

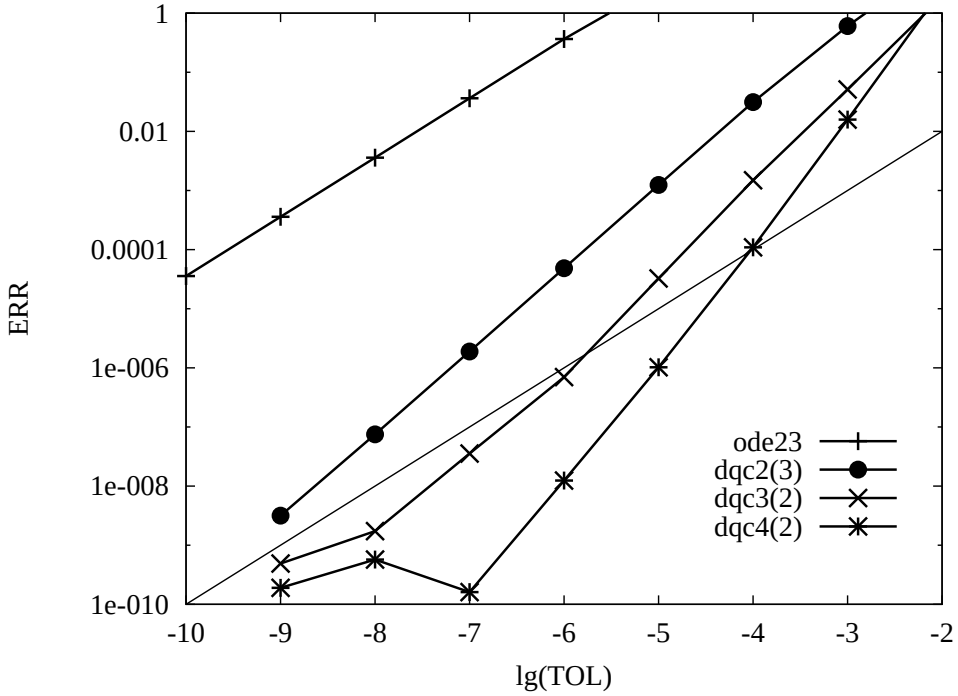

Figure 3: Results for problem I, $t_e = 4$.

the remaining test problems of this paper is similar.

5 Conclusion

We have proved the existence of doubly quasi-consistent methods in the class of parallel peer methods and have constructed one doubly quasi-consistent 4-stage EPP-method of order 2. For such numerical schemes, the principal part of the local and global errors is the same. This theoretical result has been illustrated by numerical tests and showed new possibilities for global error control. Our future work will include the construction of variable-stepsize doubly quasi-consistent methods of higher order. We also plan to use our doubly quasi-consistent methods in practical computations.

Acknowledgements

The second author thanks the support of *Fundação para a Ciência e a Tecnologia* (FCT), cofinanced by the European community fund FEDER.

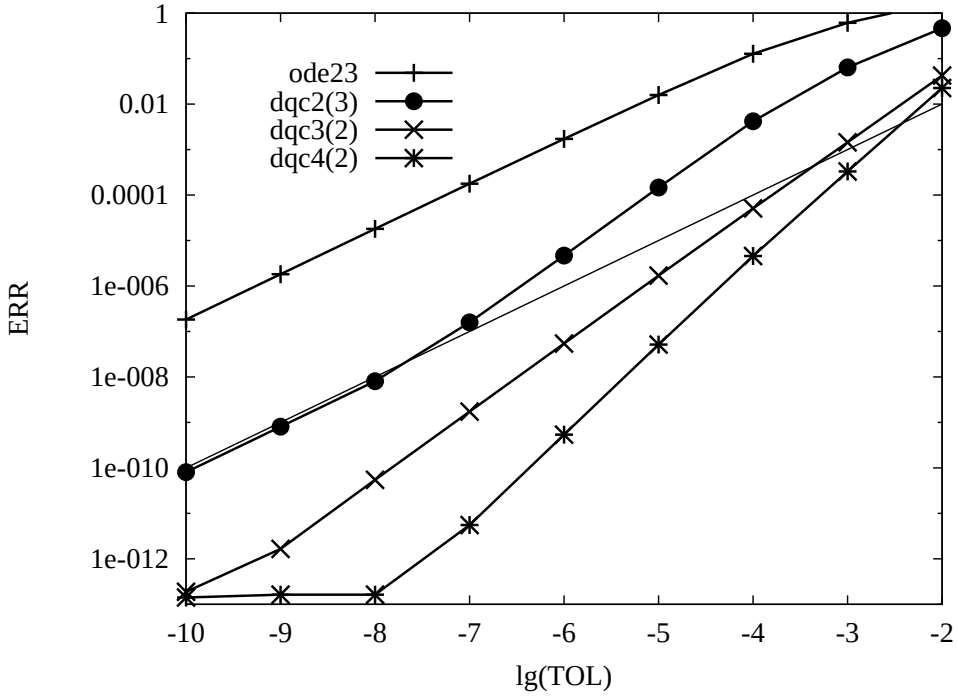

Figure 4: Results for problem II.

References

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

[2] J.C. Butcher, *Numerical methods for ordinary differential equations*, John Wiley & Sons, Chichester, 2003.

[3] J.R. Dormand, *Numerical methods for differential equations*, CRC Press 1996.

[4] E. Hairer, S.P. Nørsett, G. Wanner, *Solving ordinary differential equations I: Nonstiff problems*, Springer-Verlag, Berlin, 1993.

[5] Z. Jackiewicz, *General Linear Methods for Ordinary Differential Equations*, John Wiley & Sons, 2009.

[6] G.Yu. Kulikov, On quasi-consistent integration by Nordsieck methods, *J. Comput. Appl. Math.* **225** (2009) 268–287.

[7] G.Yu. Kulikov, S.K. Shindin, On interpolation-type multistep methods with automatic global error control, (in Russian) *Zh. Vychisl. Mat. Mat. Fiz.* **44** (2004) 1400–1421; translation in *Comput. Math. Math. Phys.* **44** (2004) 1314–1333.

[8] G.Yu. Kulikov, S.K. Shindin, Global error estimation and extrapolated multistep methods for index 1 differential-algebraic systems, *BIT* **45** (2005) 517–542.

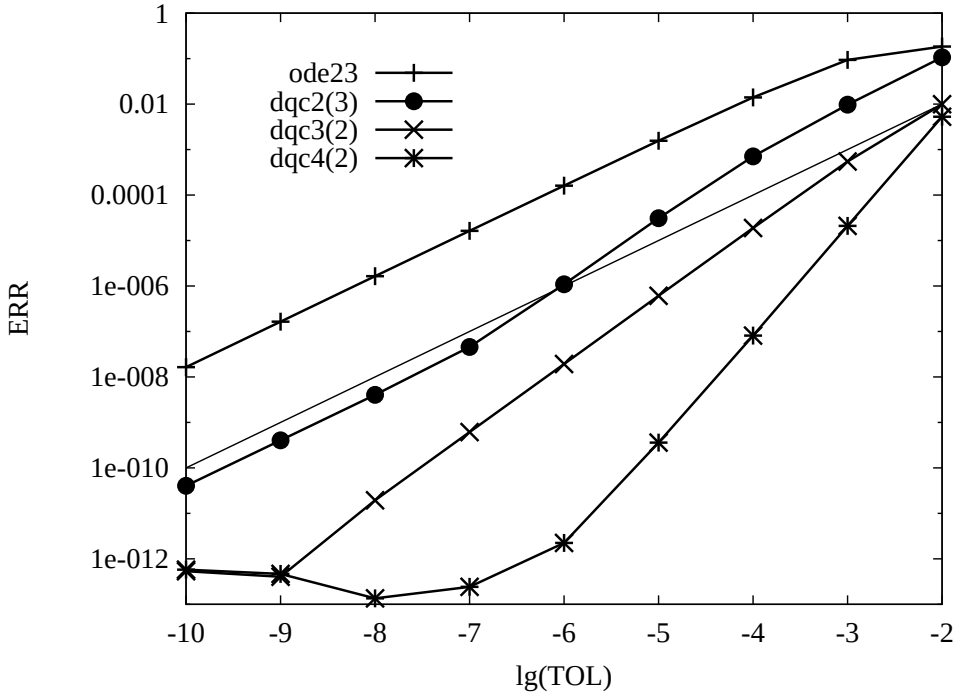

Figure 5: Results for the Kepler problem, $\varepsilon = 0$.

[9] G.Yu. Kulikov, S.K. Shindin, One-leg variable-coefficient formulas for ordinary differential equations and local-global step size control, *Numer. Algorithms* **43** (2006) 99–121.

[10] G.Yu. Kulikov, S.K. Shindin, Local and global error estimation in Nordsieck methods, *Russ. J. Numer. Anal. Math. Model.* **23** (2008) 567–595.

[11] G.Yu. Kulikov, R. Weiner, Global error control in implicit parallel peer methods, *Russ. J. Numer. Anal. Math. Model.* **25** (2010) pp. 131–146.

[12] G.Yu. Kulikov, R. Weiner, Doubly quasi-consistent parallel explicit peer methods with built-in global error estimation, *J. Comput. Appl. Math.* **233** (2010) 2351–2364.

[13] G.Yu. Kulikov, R. Weiner, Variable-stepsize interpolating explicit parallel peer methods with inherent global error control (in preparation).

[14] P. Lancaster, *Theory of matrices*, Academic Press, New York and London, 1969.

[15] J. Lang, J.G. Verwer, On global error estimation and control for initial value problems, *SIAM J. Sci. Comput.* **29** (2007) 1460–1475.

[16] R.D. Skeel, Analysis of fixed-stepsize methods, *SIAM J. Numer. Anal.* **13** (1976) 664–685.

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

[18] R.D. Skeel, Global error estimation and the backward differentiation formulas, *Appl. Math. Comput.* **31** (1989) 197–208.

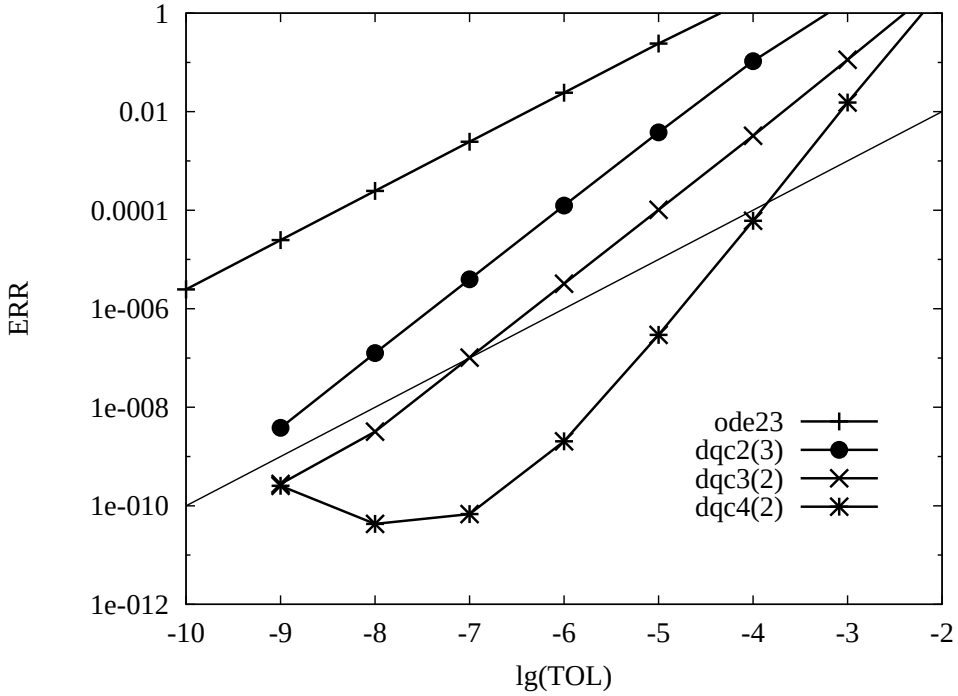

Figure 6: Results for the Kepler problem, $\varepsilon = 0.9$.

[19] R.D. Skeel, L.W. Jackson, Consistency of Nordsieck methods, *SIAM J. Numer. Anal.* **14** (1977) 910–924.

[20] H.J. Stetter, Global error estimation in ODE-solvers, in *Numerical integration of differential equations and large linear systems*, Proceedings, Bielefeld 1980, J. Hinze, ed., *Lecture Notes in Mathematics*, Vol. 968, Springer-Verlag, Berlin, 1982, pp. 269–279.

[21] R. Weiner, K. Biermann, B.A. Schmitt, H. Podhaisky, Explicit two-step peer methods, *Comput. Math. Appl.* **55** (2008) 609–619.

[22] R. Weiner, B.A. Schmitt, H. Podhaisky, S. Jebens, Superconvergent explicit two-step peer methods, *J. Comput. Appl. Math.* **223** (2009) 753–764.

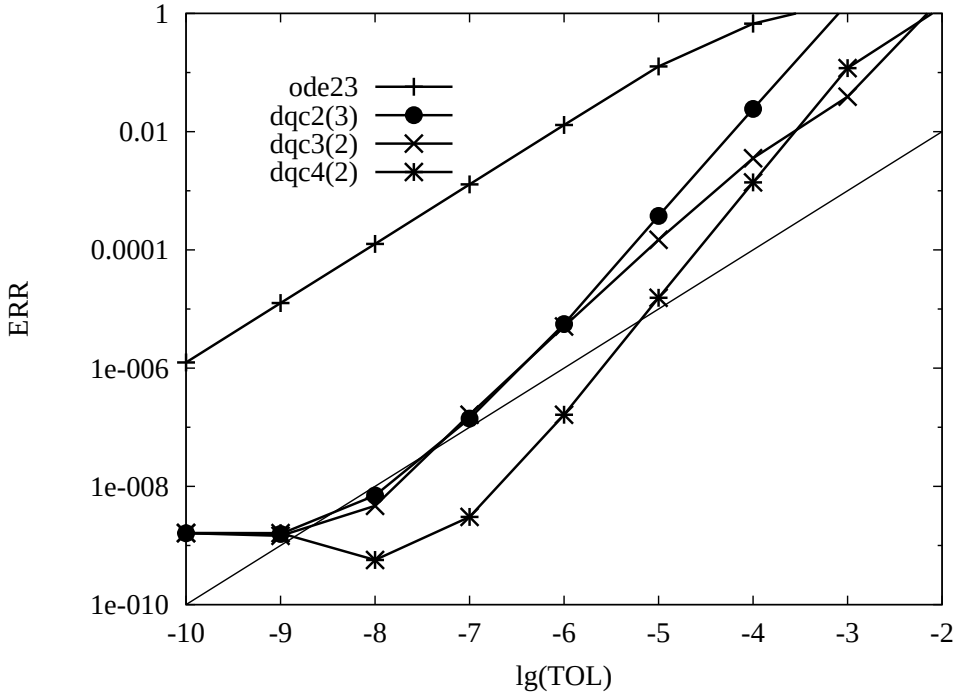

Figure 7: Results for the Arenstorf orbit.

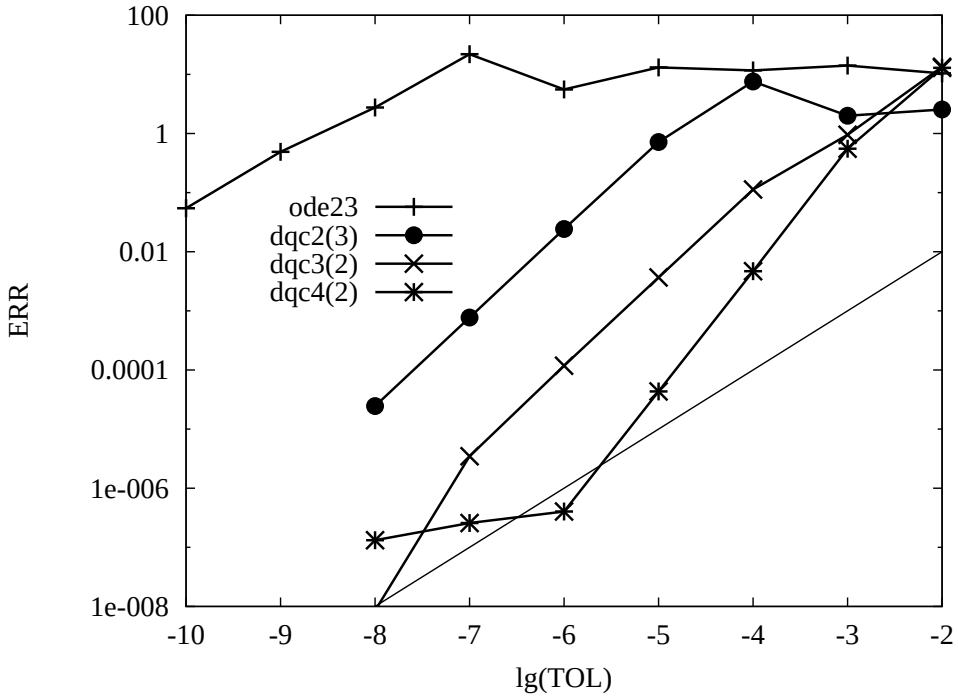

Figure 8: Results for the Lorenz attractor.

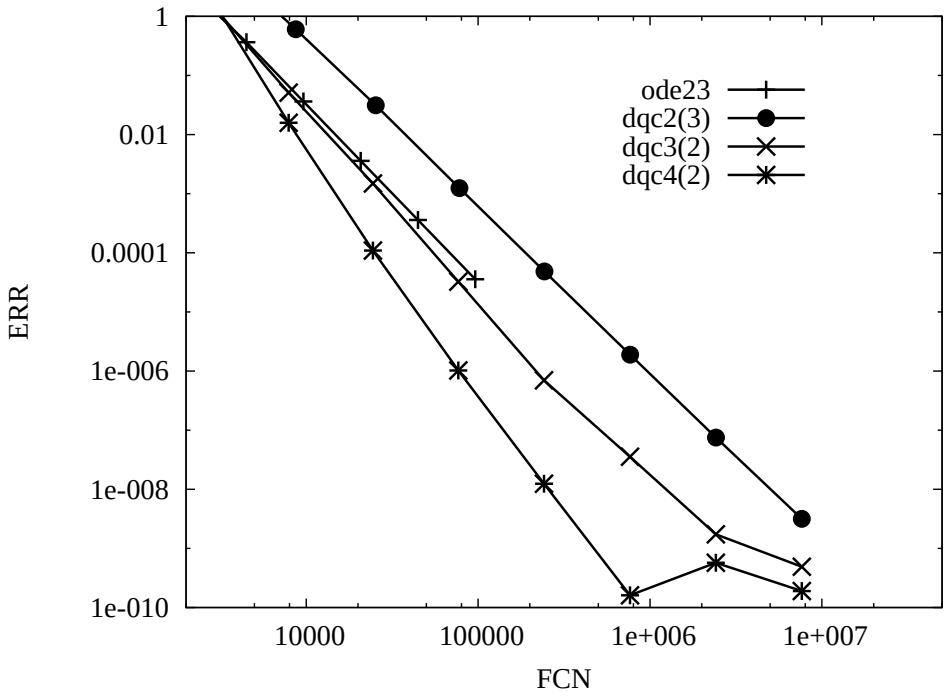

Figure 9: Number of function calls for Problem I, $t_e = 4$.

Reports of the Institutes 2010

01-10. Kristian Debrabant, Anne Kværnø, *B-series analysis of iterated Taylor methods*

02-10. Jan Prüss, Mathias Wilke, *On conserved Penrose-Fife type models*

03-10. Bodo Dittmar, *Maxima for the expectation of the lifetime of a Brownian motion in the ball*

04-10. M. Köhne, J. Prüss, M. Wilke, *Qualitative behaviour of solutions for the two-phase Navier-stokes equations with surface tension*

Reports are available via WWW: http://www2.mathematik.uni-halle.de/institut/reports/