New Runge-Kutta based schemes for ODEs with cheap global error estimation

J. Makazaga* and A. Murua†
Konputazio Zientziak eta A.A. saila,
informatika fakultatea, EHU/UPV Donostia

October 22, 2001

Abstract

We present a particular 5th order one-step integrator for ODEs that provides an estimation of the global error. It's based on the class of one-step integrator for ODEs of Murua and Makazaga [5] considered as a generalization of the globally embedded RK methods of Dormand, Gilmore and Prince [2].The scheme we present cheaply gives useful information on the behavior of the global error. Some numerical experiments show that the estimation of the global error reflects the propagation of the true global error. Moreover we present a new step-size adjustment strategy that takes advantage of the available information about the global error. The new strategy is specially suitable for problems with exponential error growth.

1 Introduction

Very efficient methods exist for the numerical integration of non-stiff ODEs. They are intended to maintain the global error controlled keeping the local error below a prescribed tolerance. However, they do not provide any information about the global error, unless additional computational effort is done.

Interesting work in obtaining schemes that give information about the propagation of the global error has been done by Dormand, Gilmore and Prince [2]. In this paper we present a methodology to construct non-stiff methods based on the general class of schemes presented in [5]. These schemes are a generalization of the methods due to Dormand, Gilmore and Prince.

We consider (non-stiff) ODE systems in autonomous form

$$\frac{dy}{dt} = f(y),\ y \in \mathbb{R}^D,\ f : \mathbb{R}^D \to \mathbb{R}^D. \tag{1}$$

*e-mail: joxe@si.ehu.es
†e-mail: ander@si.ehu.es

We wish to integrate initial value problems of the form (1) over an interval $[t_0, t_{end}]$ and we want to get, together with the numerical solution, an estimation of the global error.

Recall that the h-flow of the system is a parametric transformation of phase space $\phi_h : \mathbb{R}^D \to \mathbb{R}^D$ such that $\phi_h(y(t)) = y(t+h)$, and when we solve numerically the ODE the flow ϕ_h is replaced by a transformation of phase space $\psi_h : \mathbb{R}^D \to \mathbb{R}^D$ that approximates it. The local error is defined as $\delta(y,h) = \psi_h(y) - \phi_h(y)$, and the method ψ_h is of order p if $\delta(y,h) = O(h^{p+1})$.

For a given time discretization $t_0 < t_1 < \ldots < t_N = t_{end}$ with $h_n = t_n - t_{n-1}$ the integrator computes the numerical solution

$$y_n = \psi_{h_n}(y_{n-1}), \quad n = 1, \ldots, N \tag{2}$$

as an approximations to

$$y(t_n) = \phi_{h_n}(y(t_{n-1})), \quad n = 1, \ldots, N$$

The global error $e_n = y_n - y(t_n)$ satisfies that

$$\begin{aligned} e_n &= \psi_{h_n}(y_{n-1}) - \phi_{h_n}(y(t_{n-1})) \\ &= (\phi_{h_n}(y(t_{n-1}) + e_{n-1}) - \phi_{h_n}(y(t_{n-1}))) + \delta(y_{n-1}, h_n) \end{aligned} \tag{3}$$

This shows that the global error e_n can be viewed as the sum of two different errors,

- The local error $\delta(y_{n-1}, h_n) = \phi_{h_n}(y_{n-1}) - \psi_{h_n}(y_{n-1})$ committed at each step.
- Propagation of the global error at previous step, or equivalently, propagation and accumulation of local errors made at previous steps.

For the family of explicit Runge-Kutta (RK) methods, ψ_h is defined as

$$\psi_h(y) = y + h \sum_{i=1}^{s} b_i f(Y_i), \tag{4}$$

where for $i = 1, \ldots, s$

$$Y_i = y + h \sum_{j=1}^{i-1} a_{i,j} f(Y_j). \tag{5}$$

In this paper we will consider a subfamily of the class of one-step integrators for ODEs introduced in [5]. The general class is briefly introduced in the next section. The conditions on the parameters of the method are shown in Section 3. In Section 4, a particular method is presented. The method has been tested numerically and in Section 5 the global error is compared with the estimation of the global error for different problems. In Section 6 a new strategy to adjust the step-size is proposed. This strategy takes advantage of the information about the global error. Finally, some numerical experiments are presented to test the new strategy.

2 A general class of schemes

Substitution of $y(t_{n-1}) = y_{n-1} - e_{n-1}$ in (3) gives a recurrence for the global error of the form

$$e_n = E_{h_n}(y_{n-1}, e_{n-1}), \tag{6}$$

where the mapping $E_h : \mathbb{R}^{2D} \to \mathbb{R}^D$ is defined by

$$E_h(y,e) = \psi_h(y) - \phi_h(y-e) = \phi_h(y) - \phi_h(y-e) + \delta(y,h).$$

Obviously, if the exact flow is not available, that mapping E_h will not be available either. If we could compute a mapping $\widetilde{E}_h$ that somehow approximates the true global mapping E_h we would be able to estimate $\widetilde{e}_h$ as

$$\widetilde{e}_h = \widetilde{E}_h(y_{n-1}, \widetilde{e}_{n-1}), \quad n = 1, \ldots, N. \tag{7}$$

Extrapolation would then provide a second approximation $\overline{y}_n = y_n - \widetilde{e}_n$. Thus, the process of obtaining the approximations y_n together with the global error estimates $\widetilde{e}_n$ can be alternatively interpreted as obtaining two approximations y_n and $\overline{y}_n$, and then computing the estimated global error as their difference. More specifically, define $\overline{\psi}_h(y,\overline{y}) = \psi_h(y)$ -$\widetilde{E}_h(y, y-\overline{y})$, and compute

$$y_n = \psi_{h_n}(y_{n-1}),\ \overline{y}_n = \overline{\psi}_{h_n}(y_{n-1}, \overline{y}_{n-1}),\ \widetilde{e}_n = y_n - \overline{y}_n. \tag{8}$$

In particular, *Richardson extrapolation* and the *globally embedded RK methods* due to Dormand, Gilmore, and Prince fit into the format (8).

A possible generalization of processes of the form (8) is as follows: Take $y_0 = \overline{y}_0 = y(t_0)$ and compute for $n = 1, 2, \ldots, N$

$$y_n = \psi_{h_n}(y_{n-1}, \overline{y}_{n-1}),\ \overline{y}_n = \overline{\psi}_{h_n}(y_{n-1}, \overline{y}_{n-1}),\ \widetilde{e}_n = y_n - \overline{y}_n. \tag{9}$$

2.1 Generalized Globally Embedded RK Schemes

The schemes proposed in [5] fits in the general format (9) where the mappings $\psi_h, \overline{\psi}_h : \mathbb{R}^{2D} \to \mathbb{R}^D$ are defined by

$$\psi_h(y,\overline{y}) = y + h\sum_{i=1}^{\overline{s}} b_i f(Y_i), \tag{10}$$

$$\overline{\psi}_h(y,\overline{y}) = \overline{y} + h\sum_{i=1}^{\overline{s}} \overline{b}_i f(Y_i), \tag{11}$$

where for $i = 1, \ldots, \overline{s}$,

$$Y_i = \mu_i y + (1-\mu_i)\overline{y} + h\sum_{j=1}^{i-1} a_{i,j} f(Y_j). \tag{12}$$

The parameters $b_i, \overline{b}_i, a_{ij}, \mu_i$ have to be chosen in such a way that, when applying (9), y_n and $\overline{y}_n$ approximate $y(t_n)$ and $\tilde{e}_n = y_n - \overline{y}_n$ is an useful estimate to the global error. A scheme of the form (10)–(12) has been presented in [5]. This scheme works particularly well for integration in constant step-size mode. Nevertheless, efficient software requires variable step-sizes, and the usual strategies for step-size adjustment require a cheap estimate of the local error computed from the available intermediate values $f(Y_i)$. However, these intermediate values in general depend on both approximations y_{n-1} and $\overline{y}_{n-1}$ and therefore are affected by the size of global error estimate $\tilde{e}_{n-1}$. This causes some difficulties, which can easily be avoided (at the expense of losing some parameters of the method) as follows. If the following conditions are required

$$\mu_i = 1, \qquad i = 1, \dots, s, \tag{13}$$

$$b_i = 0, \qquad i = s+1, \dots, \overline{s}, \tag{14}$$

then the mapping ψ_h is made independent of $\overline{y}$ (actually, ψ_h becomes the RK map (4)–(5)) and the required local error estimates can be obtained in the standard way. In fact, by requiring (13)–(14) we are forcing our scheme to be of the form (8), so that the global error estimation process can be interpreted as replacing the recurrence (6) by (7), with $\tilde{E}_h(y,h) := \psi_h(y) - \overline{\psi}_h(y, y-e)$. In what follows, we will focus in the subclass of the Generalized Globally Embedded RK schemes satisfying (13)–(14). If in addition to (13)-(14) the following conditions hold,

$$\overline{b}_i = 0, \qquad i = 1, \dots, s,$$

$$\mu_i = 0, \qquad i = s+1, \dots, \overline{s},$$

then the family of globally embedded RK schemes of Dormand, Gilmore and Prince [2] is obtained.

3 Conditions on the parameters of the method

In order to construct a method of the form (9) one has to determine appropriate values for the parameters b_i, $\overline{b}_i$, a_{ij} and μ_i. These parameters must be chosen (given s and $\overline{s}$) according to the following criteria

1. The local error $\delta(y,h) = \psi_h(y) - \phi_h(y)$ of the RK scheme is as small as possible. Typically, one requires that $\delta(y,h) = O(h^{p+1})$ for a certain $p \geq 1$ (i.e., that the method be of order p) with reasonably small error constants in $O(h^{p+1})$.

2. The difference $\overline{\psi}_h(y+e,y) - \phi_h(y)$ is as small as possible. The smaller this difference, the more similar will be the map $\tilde{E}_h(y,e)$ to the global error map $E_h(y,e)$. Our goal will be obtaining estimates of $\overline{\psi}_h(y+e,y) - \phi_h(y)$ of the form

$$\psi_h(y+e,y) - \phi_h(y) = O(h^{q_0} + h^{q_1}||e|| + h^{q_2}||e||^2 + \cdots) \tag{15}$$

with sufficiently high q_k and small error constants.

The first criterion is standard for explicit RK methods. It is well known how to obtain the conditions on the coefficients of a RK method to achieve a prescribed order using rooted trees [1], [4]. We will refer to these rooted trees as RK-Trees. Each RK-tree is associated to one condition in terms of the parameters b_i, a_{ij} of the method, and the RK method achieve order p if the conditions corresponding to all RK-trees with p or fewer number of vertices are satisfied.

As for the second criterion, (15) holds with given q_k ($k = 0, 1, 2, \dots$) if certain conditions on the parameters $\overline{b}_i, a_{ij}, \mu_i$ hold. There is a nice correspondence between such conditions and the set of rooted trees with vertices of two colors. Table 1 shows the correspondence between the first conditions and rooted trees. Hereafter we use the notation $c_i = \sum a_{ij}$. The exponent q_0 in (15) depends on whether certain conditions written in terms of the parameters $\overline{b}_i, a_{ij}$ (but not in the parameters μ_i) hold. Such conditions are those associated with RK-trees. This is not surprising, since $q_0 - 1$ is the order of the RK tree defined by the map $\overline{\psi}_h(y, y)$, i.e., by the RK scheme determined by the parameters $\overline{b}_i, a_{ij}$. In general, the exponent q_k in (15) depends on the conditions corresponding to rooted trees with k white vertices.

4 Construction of a method of order 5

We have constructed a scheme (10)–(12) satisfying (13)-(14) based on the well known 5th order explicit RK method of Dormand and Prince [3]. We will refer to it as DOPRI5. $s = 6$ is determined by DOPRI5, and we have chosen $\overline{s} = 10$. b_i ($1 \leq i \leq 6$) and a_{ij} ($1 \leq j < i \leq 6$) are determined as the parameters of DOPRI5. Similarly to what is standard in the construction of (locally) embedded RK schemes, we take y_{n-1} and y_n respectively as the first and last stages of the scheme (10)-(12). This is achieved by taking $a_{7i} = b_i$ ($1 \leq i \leq 6$), $\mu_1 = 1$ and $\mu_7 = 1$. DOPRI5 requires 6 function evaluations per step, so the integrator will need three function evaluations per step in order to get the estimation of the global error.

The parameters a_{ij} ($8 \leq i \leq 10$, $j < i$), μ_8, μ_9, μ_{10} and $\overline{b}_i$ ($1 \leq i \leq 10$) have been determined in such a way that the estimate $\overline{\psi}(y, \overline{y}) - \phi(\overline{y}) = O(h^7 + h^3 \|e\| + h^2 \|e\|^2 + h \|e\|^3)$ holds (i.e., $q_0 = 7$, $q_1 = 3$, $q_2 = 2$, $q_3 = 1$) while trying to minimize the error constants of the terms h^7, $h^3 \|e\|$ and $h^2 \|e\|^2$. More precisely, we have tried to minimize in the mean square sense the conditions on the free parameters corresponding to rising each q_k by one. The remaining parameters of the proposed method are given in Table 2.

5 Numerical experiments

The behavior of the estimated global error has been tested with some numerical experiments. The aim of the experiments is to check whether our method gives useful global error estimates when implemented with in variable step-size mode. We have considered three initial value problems :

Table 1: Correspondence within trees and the conditions on the coefficients

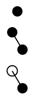
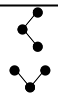
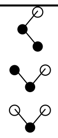
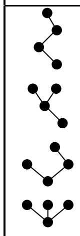
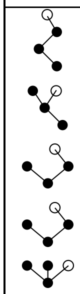
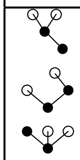
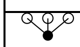
tree	condition	estimate of $\overline{\psi}_h(y, y+e) - \phi_h(y)$
	$\sum_i \overline{b}_i = 1$ $\sum_i \overline{b}_i c_i = 1$ $\sum_i \overline{b}_i \mu_i = 1$	$O(h^3 + h^2 \lVert e \rVert + h \lVert e \rVert^2)$
	$\sum_{i,j} \overline{b}_i a_{ij} c_j = \frac{1}{6}$ $\sum_i \overline{b}_i c_i^2 = \frac{1}{3}$	$O(h^4 + h^2 \lVert e \rVert + h \lVert e \rVert^2)$
	$\sum_{i,j} \overline{b}_i a_{ij} \mu_j = 0$ $\sum_{i,j} \overline{b}_i c_i \mu_i = 0$ $\sum_i \overline{b}_i \mu_i^2 = 0$	$O(h^4 + h^3 \lVert e \rVert + h^2 \lVert e \rVert^2 + h \lVert e \rVert^3)$
	$\sum_{i,jk} \overline{b}_i a_{ij} a_{jk} c_k = \frac{1}{24}$ $\sum_{i,j} \overline{b}_i a_{ij} c_j^2 = \frac{1}{12}$ $\sum_{i,j} \overline{b}_i c_i a_{ij} c_j = \frac{1}{8}$ $\sum_i \overline{b}_i c_i^3 = \frac{1}{4}$	$O(h^5 + h^3 \lVert e \rVert + h^2 \lVert e \rVert^2 + h \lVert e \rVert^3)$
	$\sum_{i,jk} \overline{b}_i a_{ij} a_{jk} \mu_k = 0$ $\sum_{i,j} \overline{b}_i a_{ij} c_j \mu_j = 0$ $\sum_{i,j} \overline{b}_i c_i a_{ij} \mu_j = 0$ $\sum_{i,j} \overline{b}_i \mu_i a_{ij} c_j = 0$ $\sum_{i,j} \overline{b}_i \mu_i c_j^2 = 0$	$O(h^5 + h^4 \lVert e \rVert + h^2 \lVert e \rVert^2 + h \lVert e \rVert^3)$
	$\sum_{i,j} \overline{b}_i a_{ij} \mu_j^2 = 0$ $\sum_{i,j} \overline{b}_i \mu_i a_{ij} \mu_j = 0$ $\sum_{i,j} \overline{b}_i \mu_i^2 c_j = 0$	$O(h^5 + h^4 \lVert e \rVert + h^3 \lVert e \rVert^2 + h \lVert e \rVert^3)$
	$\sum_{i,j} \overline{b}_i \mu_i^3 = 0$	$O(h^5 + h^4 \lVert e \rVert + h^3 \lVert e \rVert^2 + h^2 \lVert e \rVert^3 + h \lVert e \rVert^4)$

Table 2: Remaining parameters of the method based on DOPRI5

i	$1-\mu_i$	c_i	$a_{8,i}$	$a_{9,i}$	$a_{10,i}$	$\overline{b}_i$
1	1	0	$\frac{26251126}{75292183}$	$-\frac{126276029}{115017392}$	$\frac{89178409}{82486612}$	$\frac{56696811}{789712427}$
2	1	$\frac{1}{5}$	$-\frac{30511879}{68834945}$	$\frac{153409379}{49308629}$	$-\frac{275044175}{99029299}$	0
3	1	$\frac{3}{10}$	$\frac{11490887}{155205387}$	$-\frac{107711621}{48274693}$	$\frac{115406143}{68971088}$	$-\frac{47431484}{279691831}$
4	1	$\frac{4}{5}$	$\frac{700737845}{174891007}$	$-\frac{675136779}{64711289}$	$\frac{140298385}{24130572}$	$\frac{72791025}{357831874}$
5	1	$\frac{8}{9}$	$-\frac{5336}{941}$	$\frac{559269939}{36928210}$	$-\frac{344040692}{42025591}$	$\frac{17490085}{349505178}$
6	1	1	$\frac{5735}{1214}$	$-\frac{669687859}{52442748}$	$\frac{121333564}{17575013}$	$-\frac{66245097}{563676842}$
7	1	1	$-\frac{2507}{898}$	$\frac{193952703}{25738526}$	$-\frac{190380249}{47005513}$	$-\frac{24}{611}$
8	$\frac{140719960}{143529893}$	$\frac{204}{823}$	0	$\frac{169021117}{130072535}$	$-\frac{12078143}{165601005}$	$\frac{40757463}{82884629}$
9	$\frac{941}{896}$	$\frac{579}{1036}$	0	0	$\frac{56747365}{92317949}$	$\frac{33159666}{111811519}$
10	$\frac{92493035}{95359057}$	1	0	0	0	$\frac{42422453}{199331202}$

1. The first one has been taken from [4]. It is a 4th dimensional initial value problem, referred as *Arenstorf*. Its initial values correspond to a periodic solution of the restricted 3-body problem.

2. The second one, referred as *Pleiades*, is part of the test set for Initial Value Problems IVPTestSet [7]. It is a 28th dimensional problem.

3. We have taken the third problem from [6]. The system is of dimension $D = 1$, and the initial value problem is defined as

$$y' = cos(t)y,\ y(0) = 1$$

   Its solution is $y(t) = e^{sin(t)}$, a periodic function with period 2π. We refer to this problem as *expsin*.

For each of the problems we display two plots, one corresponds to a stringent tolerance and the other one to a lax tolerance. The plots of Figure 1 correspond to the *Arenstorf* problem. It has been integrated over two periods. The two plots show time versus infinity norm of the error in double logarithmic scale. Both plots compare the estimated global error (dashed line) with the exact global error (continuous line). The first one has been integrated with a tolerance of 10^{-6} while in the second one the more stringent tolerance of 10^{-9} is used. The results show that, for the considered initial value problems, the estimated global error behaves in a similar way to the true global error.

The results of the integration of the *Pleiades* problem can be seen in Figure 2. Again the reader can see two plots, each one refers to a different tolerance

Figure 1: The *Arenstorf* problem solved with a tolerance of 10^{-6} (left, 309 steps) and 10^{-9} (right, 1268 steps).

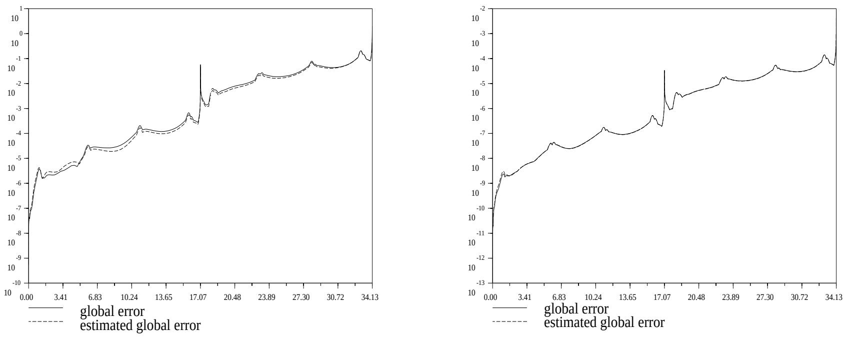
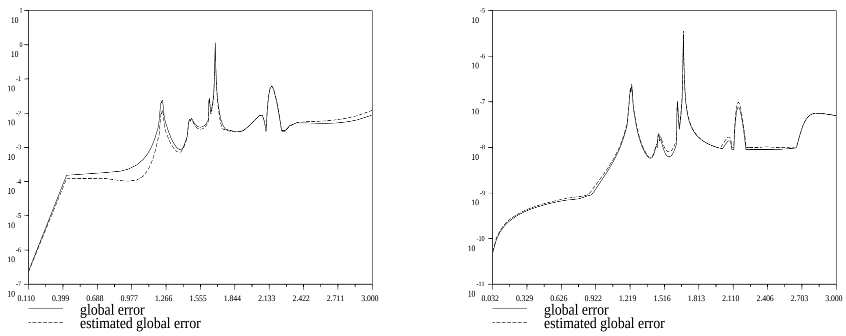


Figure 2: The *Pleiades* problem solved with a tolerance of 10^{-4} (left, 182 steps) and 10^{-9} (right, 1603 steps).



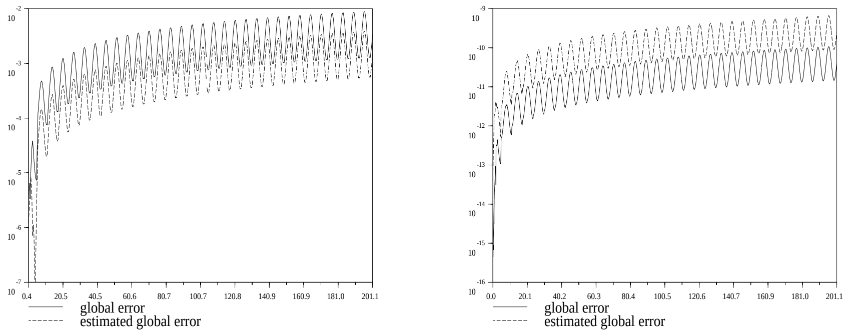
(10^{-4} and 10^{-9}) and in both plots it can be seen that the estimated global error (dashed line) behaves as a small perturbation of the true global error (continuous line). In both problems the more stringent tolerance the more accurate estimation of the global error is obtained.

Figure 3 shows the results for the third problem. Again we have two plots, the first one corresponds to an absolute tolerance of 10^{-4} and the other one to 10^{-9}. This problem has a greater difference within the true global error and the estimated global error, but the behaviors in both plots are similar.

6 Local error tolerance and global error

We want to explore the possibility of using available information about the propagation of the global error provided by our integrator to improve the efficiency of the step-size adjustment strategy. This seems especially promising when the

Figure 3: The *Expsin* problem solved with a tolerance of 10^{-4} (left, 416 steps) and 10^{-9} (right, 7467 steps).



global error grows exponentially. In this case the local errors of previous steps have an exponential effect in the global error, so one can think that it makes no sense maintaining the local error below a fixed prescribed tolerance, because the increase of the global error in the new step is mainly due to the propagation of previous errors. It then seems that a gain in efficiency could be obtained by increasing the tolerance of the local error as the propagation of previous errors grows.

6.1 A new variable step-size strategy

From (3), and assuming e_{n-1} is sufficiently small, the error in the nth step satisfies that

$$e_n \simeq R_{n,n-1} e_{n-1} + \delta_n \tag{16}$$

where $\delta_n = \delta(y_{n-1}, h_n) = \psi_{h_n}(y_{n-1}) - \phi_{h_n}(y_{n-1})$, and

$$R_{n,j} = \frac{\partial \phi_{t_n - t_j}}{\partial y} \left(y(t_j)\right).$$

In order to simplify the exposition, we hereafter will assume that the equality $e_n = R_{n,n-1} e_{n-1} + \delta_n$ holds exactly. Due to the properties of the flow, it holds that $R_{n,j} = R_{n,k} R_{k,j}$ for $n \geq k \geq j$, which implies that

$$e_n = R_{n,j} e_j + \sum_{k=j+1}^{n} R_{n,k} \delta_k, \quad 1 \leq j < n. \tag{17}$$

Our aim is to find a condition for the local error δ_n that guarantees that the effect in the global error is comparable to the effect of the propagation of previous local errors.

Lemma 1 *Under the assumption that (17) holds, if*

$$\|\delta_k\| \leq \frac{h_k \|R_{k,j}\| \varepsilon_j}{C}, \quad \varepsilon_j = \frac{\|e_j\|}{t_j - t_0}, \quad C = \max_{j<k<n} \left(\frac{\|R_{n,k}\| \|R_{k,j}\|}{\|R_{n,j}\|} \right), \tag{18}$$

then the following holds for $1 \leq j < n$

$$\varepsilon_n \leq \|R_{n,j}\| \varepsilon_j.$$

Proof: From the hipothesis of the lemma and (17) we arrive at the inequality

$$\|e_n\| \leq \|R_{n,j}\| (t_j - t_0)\varepsilon_j + \sum_{k=j+1}^{n} \|R_{n,k}\| \frac{\|R_{k,j}\| h_k \varepsilon_j}{C},$$

which can be rewritten as

$$\|e_n\| \leq \|R_{n,j}\| \varepsilon_j \left((t_j - t_0) + \sum_{k=j+1}^{n} h_k \frac{1}{C} \frac{\|R_{n,k}\| \|R_{k,j}\|}{\|R_{n,j}\|} \right)$$

and by definition of C we arrive at

$$\|e_n\| \leq \|R_{n,j}\| \varepsilon_j \left((t_j - t_0) + \sum_{k=j+1}^{n} h_k \right) = \|R_{n,j}\| \varepsilon_j (t_n - t_0),$$

which leads to the required result. □

Lemma 2 *Under the assumptions of Lemma 1, if* $\varepsilon_{n-1} \leq \|R_{n-1,j}\| \varepsilon_j$ *and* $\|\delta_n\| \leq \frac{1}{C^2} \|R_{n,n-1}\| h_n \varepsilon_{n-1}$, *then it holds that*

$$\varepsilon_n \leq \|R_{n,j}\| \varepsilon_j. \tag{19}$$

Proof: Using both inequalities in the hypothesis of the lemma, we arrive at

$$\|\delta_n\| \leq \frac{1}{C^2} \|R_{n,n-1}\| h_n \|R_{n-1,j}\| \varepsilon_j = \frac{1}{C} \frac{\|R_{n,n-1}\| \|R_{n-1,j}\|}{C} h_n \varepsilon_j,$$

and taking into account that, by definition of C,

$$\frac{\|R_{n,n-1}\| \|R_{n-1,j}\|}{C} \leq \|R_{n,j}\|$$

holds, the first inequality in (18) is obtained, which leads to the required inequality. □

Remarks

1. If the local errors satisfy

$$\|\delta_n\| \le \frac{h_n \varepsilon_{n-1}}{C^2}, \tag{20}$$

for each n, then $\|\delta_n\| \le \frac{1}{C^2} \|R_{n,n-1}\| h_n \varepsilon_{n-1}$, and Lemma 2 can be applied. In particular, we have $\varepsilon_n \le \|R_{n,1}\| \varepsilon_1$, that is, $\|e_n\| \le \frac{t_n - t_1}{t_1 - t_0} \|R_{n,1}\| \|\delta_1\|$. Compared to the bound $\|R_{n,1}\| \|\delta_1\|$ of the effect of the local error of the first step (that is, the global error of an integration where the rest of steps are extremely precise), the global error e_n of a numerical integration with local errors satisfying (20) admits a reasonable bound.

   This argument motivates us to design variable step-size strategies that ensures that (20) holds for the local error committed at each step.

2. Unfortunately, the parameter C depends on the problem to be solved. In a linear problem of the form $y' = Ay$, it holds that $C = 1$ if A is a symmetric matrix. However, C can be considerably higher, even for linear problems. Hence, the actual value of C used for the variable step-size implementation based on the condition (20) should be adjusted for each problem.

The standard way to adjust the step size (see [6] (pp. 334–340) and [4] (pp. 167)) depends on a fixed tolerance τ provided by the user and on a norm $\|\cdot\|$ in which the local error is measured. The step is accepted depending on the ratio between these two values, and the new step size is computed in terms of this ratio. If we take (20) into account, we can consider $K\varepsilon_{n-1}$ as a variable tolerance, where $K = \frac{1}{C^2}$. In order to increase the efficiency of the integrator, we propose to use a variable tolerance $\tau_n = \max(\tau, K\varepsilon_{n-1})$ for each step.

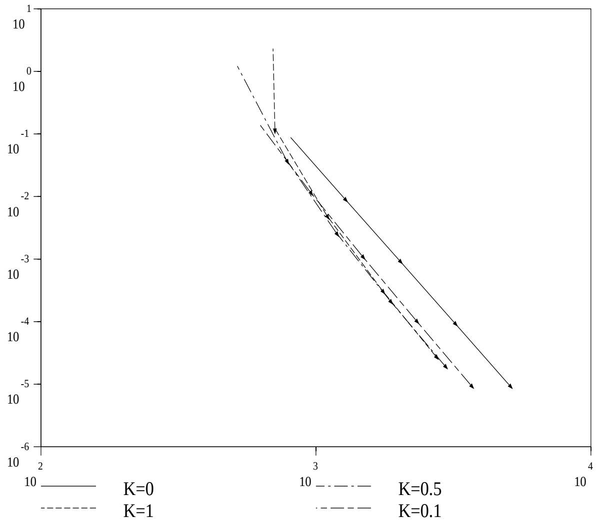
The parameter $K \in (0, 1]$ depends on the problem and it has to be adjusted. If we take a value near 0, the new strategy will become the standard strategy because $\tau > K\varepsilon_{n-1}$. If K is too large, so that $K > \frac{1}{C^2}$ where C is given by (18) then we have the risk of going out of control because in that case (20) will not hold.

7 Numerical experiments

We have implemented the possibility of using the new strategy based on (20) in our experimental code and we have performed several numerical experiments for different initial value problems. We next present the results obtained for four of them. The first two problems in Section 5, the initial value problem referred as *Lorenz*, taken from [4] (pp. 120–121) and the problem referred as *Twobody* taken from [6] (pp. 121) with eccentricity=0.5.

The plots show the efficiency of the integration in the sense of number of steps needed in the integration versus the global error at the end. We have integrated the problems with different local error tolerances and we show this

Figure 4: The comparison of the cost for the *Arenstorf* problem between the new step-size strategy and the usual strategy.



curve in a double logarithmic scale. In each plot there are many curves, the first one corresponds to the usual step-size strategy ($K = 0$), and the rest of the curves correspond to the new strategy but each one has been obtained using a different value of the parameter K.

We can see the results obtained for the *Arenstorf* problem in Figure 4. The continuous line corresponds to the usual step-size strategy, and in all cases the new strategy gets better results. The gain of efficiency for this problem in the case of $K = 0.2$ is around 33% for all tolerances. It can be seen that for $K = 1$ and lax tolerance the integration goes out of control.

Figure 5 shows the difference between standard step-size strategy and the new step-size strategy for the *Pleiades* problem. Again the continuous line corresponds to the usual strategy and the rest have been obtained using the new strategy. The gain of efficiency for $K = 0.8$ is between 20% and 45%, depending on the tolerance. In this problem we can see that for $K = 0.1$ the results are similar to the usual strategy. Nevertheless higher values of K take advantage of the new strategy.

The results for the *Lorenz* problem are shown in Figure 6. It can be seen that applying the new strategy we need fewer steps to get the same accuracy: for $K = 0.3$ only 55% of steps are needed, so the gain of efficiency is 45%. In this figure it can be seen that incrementing the value of the parameter K the integration needs fewer steps, but in the case of $K = 1$ the efficiency diagram shows strange behavior for lax tolerances.

Note that in the three problems presented above the error grows very fast,

Figure 5: Efficiency for the *Pleiades* problem with usual step-size strategy and the new step-size strategy.

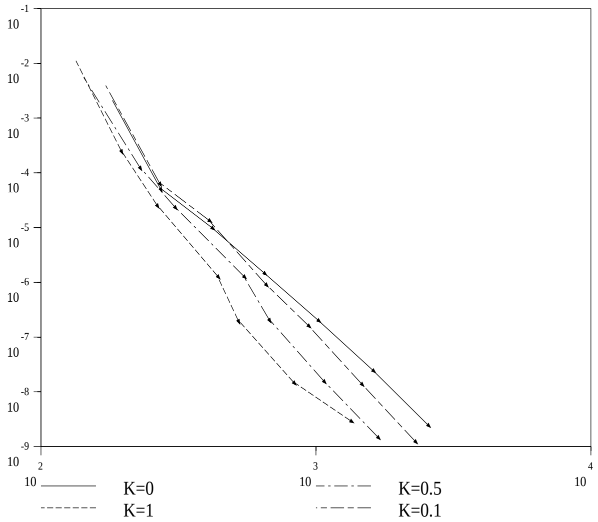


Figure 6: Efficiency comparison for the *Lorenz* problem.

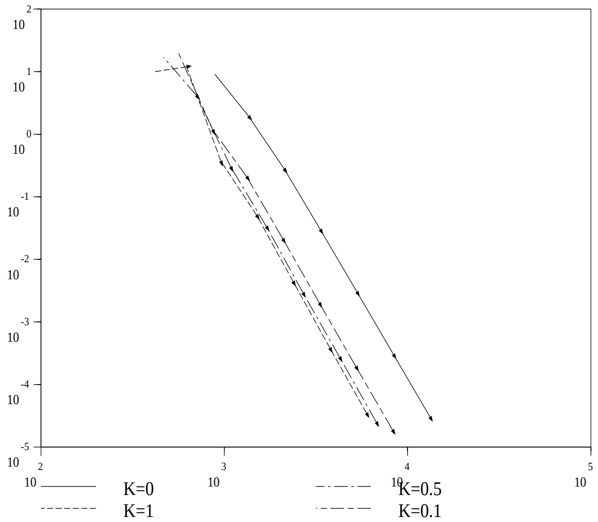
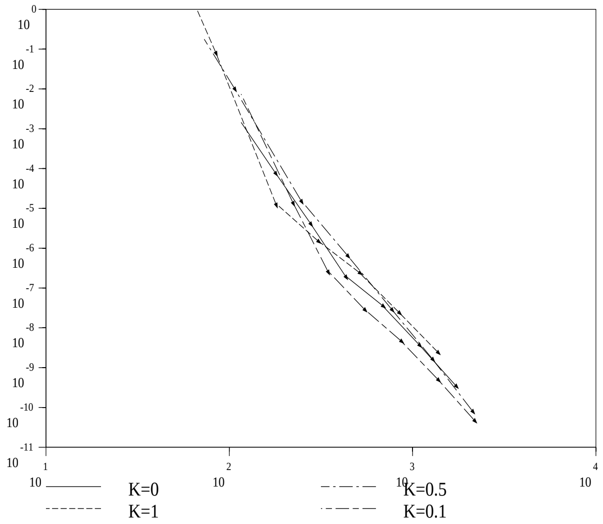


Figure 7: Efficiency comparison for the *Twobody* problem.



and so the new strategy gives very good results. The new strategy does not seem to improve efficiency for problems whose error does not grow fast. Figure 7 shows the efficiency diagram obtained with the *Twobody* problem with eccentricity equal to 0.5. In this case, the new strategy acts like a change of tolerance, there is no gain of efficiency.

We have seen that the new step-size strategy can be used to get a gain in efficiency but one can think that the estimation of the global error may be affected by the new step-size strategy. We have performed extensive numerical experiments to test that issue. We have observed in all cases that the quality of the estimation of the global error is not significantly affected by the new step-size strategy.

Acknowledgments

This work has been supported by grant Orden Foral 328/99 of Gipuzkoako Foru Aldundia.

References

[1] J.C. Butcher, *The Numerical Analysis of Ordinary Differential Equations.* John Willey and Sons 1987.

[2] J.R. Dormand, J.P. Gilmore and P.J. Prince, *Globally embedded Runge-Kutta schemes,* Annals of Numerical Mathematics **1** (1994), 97-106

[3] J.R. Dormand, P.J. Prince, *A family of embedded Runge-Kutta formulae.* J. Comp. Appl. Math. **6** (1980), 19-26.

[4] E. Hairer, S.P. Nørsett and G. Wanner, *Solving Ordinary Differential Equations I. Non-Stiff Problems,* Second Edition, Springer-Verlag, 1993

[5] A. Murua and J. Makazaga. *Cheap one-step global error estimation for ODEs,* New Zealand journal of mathematics **29** (2000), 211-221.

[6] L.F Shampine, *Numerical Solution of Ordinary Differential Equations.* Chapman and Hall, 1994.

[7] Walter M. Lioen and Jacques J.B. de Swart, *Test Set for Initial Value Problem Solvers, Release 2.1 September 1999, http://www.cwi.nl/cwi/projects/IVPtestset/.*