CHEAP ERROR ESTIMATION FOR RUNGE-KUTTA METHODS

CH. TSITOURAS[†‡] AND S. N. PAPAKOSTAS [†§]

Abstract. Explicit Runge-Kutta methods in the form of pairs of orders $p\,(p-1)$ provide an attractive means for the solution of initial value problems of first order differential equations. Most existing Runge-Kutta formulae (single methods, as well as pairs) use the minimal number of stages required for achieving a prescribed order. In this article we shall study, in terms of efficiency and reliability, Runge-Kutta pairs of orders $p\,(q)$, whenever $q < p-1$. While in practice pairs of orders $p\,(p-1)$ usually require one or two more stages in addition to those already necessary for a pth order single method, we show here that if $p = 6$, 7, 8, or 10, efficient pairs of orders $p\,(p-2)$, $p\,(p-3)$, or $p\,(p-4)$ may be easily constructed with a reduced cost in function evaluations with respect to pairs of orders $p\,(p-1)$. In general comparing $p\,(q)$ pairs used in local extrapolation mode (the one most frequent in practice), we see that while the propagated solution of a problem is in either case of the same order p, pairs characterized by $q < p-1$ use less function evaluations. Consequently they might be more efficient, provided that they are accompanied by a reliable estimator and an efficient implementation could be found for their application in practical situations. A new step-size selection algorithm proposed here takes full advantage of the potential for increased efficiency inherited on pairs that are accompanied by no-additional-cost estimators. This algorithm, which may be also applied to Nystrom pairs, makes code implementation of these pairs attractive, as in all cases the proportionality of the global error with respect to the requested tolerance is in practice always achieved. Here we shall cover the cases of pairs of orders 4(2), 6(4), 7(5), 8(6), 8(5), 8(4), and 10(6) with nearly minimized truncation error coefficients which use a minimal number of stages (i.e., in almost all cases equal to that currently known to be the minimal required for constructing a single method of order equal to that of the higher order method of the pair). By studying the numerical performance of these pairs we may see that not only these pairs are as reliable as the respective pairs of the type $p\,(p-1)$, but in all cases they seem to be more efficient. Another important consequence of the numerical tests performed here, is that they suggest that for a given number of stages, the best Runge-Kutta pairs that may be attained are those for which the higher order method is of the maximal possible order.

Key words. Initial Value Problems, ODE solvers, one-step methods, Runge-Kutta pairs, step-size change control algorithm, tolerance proportionality

AMS subject classification. 65L05

1. Introduction. Explicit Runge-Kutta (RK) methods in the form of pairs of embedded methods are nowadays considered as one of the most efficient means for solving the non-stiff initial value problem

$$y' = f\,(x, y)\,, \quad y\,(x_0) = y_0$$
$$x \in [x_0, x_e]\,, \quad f : R \times R^m \to R^m.$$

A RK method is characterized by the triple A, b, c (where $A \in R^{s\times s}$, b^T, $c \in R^s$) and is said to be of algebraic order (or simply order) p, whenever the coefficients in A, b, c satisfy a system of order conditions, which are in one-to-one correspondence with the set of (rooted) trees of orders not exceeding p (see Butcher [3], Hairer, Norsett and Wanner [14]). RK pairs are characterized by two RK methods of orders $p\,(q)$, $(p > q)$ with distinct vectors of weights b, $\hat{b}$, which, however, share the same function evaluations (A, c are for both methods the same). In the following by $T^{(p+1)}$ we symbolize a vector consisting of the principal truncation error coefficients of a method. (Here we adopt the usual determination of these coefficients as in [3], [14].)

[†] National Technical University of Athens, Department of Mathematics, Zografou Campus, Athens 157 80, GREECE.

[‡] (tsitoura@@math.ntua.gr)

[§] (spapakos@@softlab.ece.ntua.gr)

Let $s^*(p)$ be the minimal number of stages required for the construction of a pth order RK method and $s^*(p,q)$ for a $p(q)$ pair respectively. The exact value of the function $s^*(p)$ is currently known for orders up to eight and has been established by Butcher [1], [2] (for a brief summary of related known results the reader may also consult [3] and [14]). Most existing RK pairs are of orders $p(p-1)$ and the value of $s^*(p,p-1)$ is known only for orders up to five (for order five see for example [1] and the report by Fehlberg [12]).

The most notable open problem in the algebraic study of explicit RK pairs is, even nowadays, regarded that of the exact determination of the values of the functions $s^*(p)$ and $s^*(p,p-1)$. It has been conjectured by Butcher in [3], p. 194 (in view of the results offered in [1], [3] and [29]) that $s^*(p)-p=O\left(p^2\right)$. It is also highly probable that $s^*(p,p-1)-s^*(p)$ is at most $O(p)$, since for orders up to eight there exist pairs confirming the validity of the relation $s^*(p,p-1)=s^*(p)+1$. However, the only twelve stage family of 8(7) pairs, the authors are aware of, fails on quadrature problems (such pairs are those of type Ib in [17] and for example the pair NEW8(6) in Section 2 if it is equipped with an embedded 7th order method instead of a 6th order one). In general we call *defective* those pairs or families of pairs of order $p(q)$ for which at least one pth order truncation error coefficient of the qth order method vanishes. As this might lead in some cases to unreliable error estimation, in general, defective pairs are not considered good candidates for code implementations (see Shampine [22]).

In practice, the solution of the order conditions for the construction of RK methods or pairs involves the application of a suitable set of simplifying assumptions (see for example [3] and for a more up to date information [17] and the classification and relevant discussion therein). Although the analysis seems to be complicated (especially for higher order methods), in most cases (see [19], [18], and [17]) efficient and easily implementable algorithms have been obtained. These algorithms are characterized by a number of free parameters and define certain families of solution of the respective order conditions. The values of these parameters may be chosen to satisfy specific criteria. A prominent criterion is usually the minimization of the principal truncation error coefficients (in the form of say $\left\|T^{(p+1)}\right\|_2$), as exhibited by Dormand and Prince [9], [20] and more recently by the improved formulas in [18], [19]. This involves the tricky experience of using numerical and/or symbolic routines, which in some cases proves to be rather time consuming. Except of the family of 6(5) pairs studied in [19] (also discovered independently by Verner [32]), another interesting family is that introduced by Dormand, Lockyer, McCorrigan and Prince [8] and studied by Verner [31]. The reader may complete the whole picture of the relevant theory, for the part of the historical background, by consulting also Fehlberg [11], [12], Verner [29], [30], Curtis [5], [6] and Butcher [3]. Another well known formula may be found in Calvo, Montijano, and Randez [4].

From a practical point of view, and according to our own tests, we have seen that at least for orders up to eight and for medium to high requested accuracies higher order RK pairs seem to be, in most cases, more efficient than lower order ones. This phenomenon to some extent depends on the underlying families and the number of free parameters defining them, because in general, among different families of the same orders better formulas are usually obtained from those families that are characterized by a larger number of free parameters (especially if these are among the nodes of a method). All existing efficient pairs (with a notable exception to be discussed later) use the minimal currently known number of stages required for achieving a non-

defective pair of a prescribed order; it is very probable that this number is actually the theoretically minimal one as well.

A question that emerges when realizing the relation between p and $s^*(p, p-1)$ is what happens when the stages s of a pth order RK pair are chosen to satisfy the relations

$$\begin{aligned} s^*(p-1, p-2) &< s < s^*(p, p-1), \\ s^*(p) &\leq s. \end{aligned}$$

Here we shall study the possibility of using a pth order RK pair when the number of its stages equals $s^*(p)$ or $s^*(p)+1$. The lower order formulas of these pairs, depending on the occasion, will be equal to $p-2$, $p-3$, or $p-4$. When comparing these pairs with pairs of the type $p(p-1)$, we see that in most cases they use one or two less function evaluations, while retaining the same order of approximation on the propagated solution. Consequently, for higher order pairs, where except of a measure of the size of $T^{(p+1)}$, the numerical performance of the pairs depends also heavily on the cost in function evaluations s, these pairs might offer some good prospects for finding a practical implementation. In any case a proper adjustment of the step-size change algorithm, which takes into account the idiosyncrasy of these pairs, should be made. This will be studied extensively in Section 3. The numerical results of Section 4 will show that in practice a $(p-2)$-order estimator proves to be adequately reliable. Estimators of orders $(p-3)$ and $(p-4)$ are also reliable, provided they are applied at suitably stringent accuracies.

Let $\bar{s}^*(p, q)$ be the currently best known upper bound for $s(p, q)$. All the higher order methods of the pairs proposed in this article are of the maximal order allowed by the number of their stages, i.e., their stages s are chosen so that $s = \bar{s}^*(p, q)$. We shall call these pairs *maximal* and we note here that maximal pairs of the type $p(p-2)$ are quite common when solving the initial value problem

$$\begin{aligned} &y'' = f(x, y), \quad y(0) = y_0, \quad y'(0) = y'_0, \\ &x \in [x_0, x_e], \quad f : R \times R^m \to R^m, \end{aligned}$$

by Nystrom methods, which have always been considered as reliable and efficient. However, as we shall see later there exists a more reliable and effective way for implementing them than that currently being in use.

An alternative possibility is to consider pairs that do not abide by this rule (*non-maximal* pairs). Thus far, the only formulae of this type that have appeared in the literature are the single methods of Shanks [26] and the pairs proposed by Sharp and Smart in [28] (as well as a 3(4) pair by Fehlberg [12] and a nine-stage FSAL 5(6) pair by Butcher [3] when either one of them is used in local extrapolation mode—LEM). Shanks realized that more and more additional stages were required as the order of a RK method was increasing. So, as an alternative he proposed, among others, in [26] methods of one or two stages less than the minimal required for a specified order. These methods were thus essentially of one order less than the claimed one, but with relatively small principal truncation error coefficients. Another approach was recently adopted for RK pairs, this time by Sharp and Smart in [26]. This approach, in one respect similar to that of Shanks, results to the construction of RK pairs which are non-maximal with respect to the order achieved for the number of stages used. But this time their third, fifth, sixth, and seventh-order pairs use more stages than those known to suffice for achieving their order (non-maximal pairs).

In the next section we shall discuss the derivation of pairs of orders $p\,(p-2)$, $p\,(p-3)$, or $p\,(p-4)$, where we shall present some new pairs with minimized truncation error coefficients. In the third section we shall consider some practical aspects of the implementation of the new pairs. Specifically, a new step-size change control algorithm will be presented and its significance, when compared to a classical algorithm of this kind, will be highlighted by applying it, among others, on $p\,(p-2)$ order Nystrom pairs from the literature. In the final section we intend to present some numerical tests and comparisons between the new maximal pairs presented here and some other maximal and non-maximal pairs that have appeared previously. As implemented here the new pairs performed better than the best $p(p-1)$ pairs from the literature on the Toronto nonstiff test set[15].

2. Some Families Leading to Pairs of Orders $p(p-2)$, $p(p-3)$, and $p(p-4)$. First we should answer the question of whether there exist pairs of orders 4(3) with four stages. Consider the vectors in the set

$$\begin{aligned} S &= \{v_1, v_2, v_3\} \\ &= \left\{ (c_2, c_3, c_4)\,, \left(c_2^2, c_3^2, c_4^2\right), \left(-\frac{c_2^2}{2}, a_{32}c_2 - \frac{c_3^2}{2}, a_{42}c_2 + a_{43}c_3 - \frac{c_4^2}{2}\right) \right\}. \end{aligned}$$

S is formed from the order conditions bc, bc^2 and $bAc - c^2/2$. We shall first show that these are linear independent. Assume that this is not true, then for suitable reals κ, λ, μ, not all of them being zero we should have

$$\kappa v_1 + \lambda v_2 + \mu v_3 = 0.$$

Let $C = \mathrm{diag}\,(c)$. Multiplying the above relation from the left by (b_2, b_3, b_4), (b_2c_2, b_3c_3, b_4c_4) and using the order conditions

$$\begin{array}{ll} bc = \frac{1}{2}, & \\ bc^2 = \frac{1}{3}, & bAc = \frac{1}{2\cdot 3}, \\ bc^3 = \frac{1}{4}, & bCAc = \frac{1}{2\cdot 4}, \end{array}$$

we find that

$$\left.\begin{array}{l} \frac{1}{2}\kappa + \frac{1}{3}\lambda = 0 \\ \frac{1}{3}\kappa + \frac{1}{4}\lambda = 0 \end{array}\right\} \Rightarrow \kappa = \lambda = 0,$$

which is a contradiction. Next, multiplying the vectors in S from the left by u^T, where

$$u = \left(b_2 - \hat{b}_2, b_3 - \hat{b}_3, b_4 - \hat{b}_4\right)^T,$$

and using also the order conditions for the lower order method

$$\begin{array}{ll} \hat{b}c = \frac{1}{2}, & \\ \hat{b}c^2 = \frac{1}{3}, & \hat{b}Ac = \frac{1}{2\cdot 3}, \end{array}$$

we find that these vectors are orthogonal to the members of S. So we are led to the fact that both methods of the pair are identical and consequently there do not exist pairs of orders 4(3) with 4 stages.

However 4(3) pairs may be constructed by employing the FSAL device (see Fehlberg [12], Dormand and Prince [9]). Assuming that c_2, c_3 are different from

each other and from 0 and 1, we may easily adapt the analytical solution presented by Butcher [3], p. 179, both to non-FSAL 4(2) pairs and to four-stages FSAL 4(3) pairs. The global minimum of $\left\|T^{(5)}\right\|_2$ for both these families of pairs is obtained when $c_2 = 5/14$ and $c_3 = 13/22$ and the corresponding coefficients may be found in Table 2.2. We should point out here that the so constructed pairs NEW4(2) and NEW4(3) are the optimum pairs among all those possessing these order and stage characteristics, because for their derivation no simplifying assumptions were used (other than the customary $Ae = c$, $e = \underbrace{(1, 1, \ldots, 1)}_{s}^{T}$).

Any fifth-order RK pair requires at least six stages. Moreover, using exactly this number of stages, efficient 5(4) pairs may be obtained (see [18]). Consequently, the consideration of 5(3) pairs does not seem to offer any advantage. It is not known until now if there exist non-defective 6(5) pairs with seven stages, while there exist three categories of families which use eight stages (effectively), [19]. Nevertheless, there do exist a 6(4) family of pairs derived when embedding a fourth order method to the pairs of type Ia, as defined in [17]. Similarly, pairs of orders 7(5) exist with nine stages and of orders 8(6) with twelve stages (all of them being of the type Ia as well). Any one of them may be constructed by first obtaining a $p\,(p-1)$ pair using a general algorithm developed in [17] (when putting $b_{s+1} = 0$) and then embedding a $(p-2)$-order method at no extra cost. The second part of this algorithm involves the solution of only one linear system of equations and it is thus fairly straightforward. The superficial role played in this case by the $(p-1)$-order method simplifies the whole derivation, particularly the part concerning the pth order method of the pair. Moreover 8(5) and 8(4) pairs of type Ib with eleven stages may be constructed in a similar way. However, we may easily show that any 8(5) or 8(4) pair is necessarily quadrature defective and its use will be limited here only for illustration purposes and for reasons to be exhibited in Sections 3, 4. All these pairs are of the maximal order allowed by the number of stages used.

A complete theoretical study of the pairs of type Ia of orders 6, 7, and 8 has been performed (among others) for the first time in [17]. We note here that a subset of the pairs of type Ia of orders 6, however characterized by one parameter less, has been constructed by a different treatment (and not easily generalizable) by Verner [30]. The type of simplifying assumptions used by these pairs has been proposed even earlier (but without any theoretical justification) by Fehlberg [11]. However, the approach of Fehlberg is somewhat limited (and now outdated) for reasons explained in [17]. Hence, as Fehlberg imposed in [11] unnecessary restrictions, his families of pairs are described by fewer parameters and in one case (9(8) pair) even led to the use of one additional stage. A cure for this was later proposed by Verner in [29], but otherwise the number of free parameters in his (implied) enhanced derivation were again unnecessarily small and no theoretical justification or algorithms appeared there as well.

It is interesting to note that in all these cases the principal truncation error coefficients of the higher order method of each pair depends almost exclusively on the values of the free nodes c_i. Consequently as representative optimal pairs, with respect to the value of $\left\|T^{(p+1)}\right\|_2$, we may choose for the 6(4) pairs the selection of the nodes of the optimized pair of category (A) presented in [19]. For the 7(5) and 8(6) cases the choice we make is based on the same selection of the nodes as in the pairs of orders 7(6) and 8(7) presented in [17]. The coefficients of the 6(4) pair are given in Table 2.3. For the other pairs, the last column of A and the coefficients b_i,

$\hat{b}_i$ as needed to be modified, are presented in Tables 2.4, 2.5. The pair 7(5) is based on the same selection of nodes as the pair 7(6) in [17], namely

$$c_2 = \frac{1}{18}, \quad c_4 = \frac{1}{6}, \quad c_5 = \frac{89}{200}, \quad c_7 = \frac{74}{95}, \quad c_8 = \frac{8}{9},$$

with

$$\hat{b}_8 = \frac{1114095023}{9014791121},$$

while the 8(6) pair is based on the nodes

$$\begin{array}{llll} c_2 = \frac{9}{142}, & c_5 = \frac{50}{129}, & c_6 = \frac{34}{73}, & c_7 = \frac{23}{148}, \\ c_8 = \frac{142}{141}, & c_{10} = \frac{83}{91}, & c_{11} = \frac{143}{149}, & \end{array}$$

with

$$a_{87} = \frac{254}{39}, \hat{b}_{11} = -\frac{3}{2},$$

of the respective pair 8(7). The major characteristics of all new pairs, as well as those used for the numerical comparisons of Section 4, are given in Table 2.1. The new 8(5) and 8(4) pairs may be found in Table 2.6.

It seems appropriate to note that the 8(5) and 8(4) pairs are also based on the derivation, proofs and algorithms of [17]. Specifically, the free parameters in this case are c_2, c_5, b_9, and b_{10}. Additionally, for a 5th and 4th order embedded method $\hat{b}_9$, and $\hat{b}_7$ respectively are free as well. According to Proposition 4.2 of [17] three of c_6, c_7, c_8, c_9, and c_{10} must be non-zero, distinct and different from unity. Elaborating on this proposition we note that the nodes c_6, c_7, and c_8 must be distinct, while each one of c_9 and c_{10} must be distinct and equal to anyone of the former three nodes, i.e., the following cases

$$c_6 = c_7, \quad c_6 = c_8, \quad c_7 = c_8, \quad c_9 = c_{10},$$

must be excluded. Consequently, the quadrature order conditions restrict the parameters c_6, c_7, and c_8 to assume any permutation of the values $\frac{1}{2}$, $\frac{1}{2}\left(1 \pm \sqrt{\frac{3}{7}}\right)$. In total there are 36 possibilities which all lead to the construction of respective methods. The pairs of Table 2.6 are based on the selection $c_9 = c_6$ and $c_{10} = c_8$. Curtis in [5] studied heuristically just one case, namely that corresponding to $c_9 = c_6$ and $c_{10} = c_7$. In particular his choice $c_6 = \frac{1}{2}\left(1 + \sqrt{\frac{3}{7}}\right), c_7 = \frac{1}{2}\left(1 - \sqrt{\frac{3}{7}}\right)$ and $c_8 = \frac{1}{2}$ leads to $\left\| T^{(9)} \right\|_2 = 1.04 \cdot 10^{-4}$.

A significant characteristic of all these pairs is that, having selected the free $\hat{b}$ parameters suitably, they waste less function evaluations on each rejected step. In particular, $p\,(p-2)$ order pairs save 2 function evaluation per rejected step and $p\,(p-3)$, $p\,(p-4)$ order pairs save 3 and 5 function evaluations respectively, while in general $p\,(p-1)$ pairs save only 1 function evaluation per rejected step. This is due to the selection for example $b_{10} = \hat{b}_{10}$ for the parameters of the low order method of the pair 8(5), which leads to $b_{11} = \hat{b}_{11}$, and accordingly for the others. Under this selection, while, as it is usual for any RK pair, after a rejected step the first function evaluation is not reevaluated, the above conditions allow as well an early termination in the estimation of the local error before all stages are computed.

TABLE 2.1
The characteristics of the RK pairs referenced in this article.

Method of orders $p\,(q)$	Effective Number of Stages	$\left\|T^{(p+1)}\right\|_2$	$\left\|T^{(p+2)}\right\|_2$	I_R	I_{IM}	D_∞
NEW4(3)	4 (FSAL)	$1.19 \cdot 10^{-2}$	$1.36 \cdot 10^{-2}$	$(-2.78, 0)$	$(0, 2.82)$	1.15
SS3(2)	4	$1.28 \cdot 10^{-2}$	$1.39 \cdot 10^{-2}$	$(-3.02, 0)$	$(0, 2.75)$	1.22
NEW4(2)	4	$1.19 \cdot 10^{-2}$	$1.36 \cdot 10^{-2}$	$(-2.78, 0)$	$(0, 2.82)$	1.15
SS5(4)	7	$7.1 \cdot 10^{-5}$	$1.77 \cdot 10^{-4}$	$(-3.91, 0)$	—	0.86
NEW6(4)	7	$2.12 \cdot 10^{-4}$	$3.47 \cdot 10^{-4}$	$(-3.95, 0)$	$(0, 1.76)$	0.83
NEW7(6)	10	$2.83 \cdot 10^{-5}$	$6.24 \cdot 10^{-5}$	$(-4.5, 0)$	$(2.29, 4.61)$	13.9
SS6(5)	9 (FSAL)	$3.24 \cdot 10^{-6}$	$2.44 \cdot 10^{-5}$	$(-4.52, 0)$	$(2.28, 4.60)$	13.3
NEW7(5)	9	$2.83 \cdot 10^{-5}$	$6.24 \cdot 10^{-5}$	$(-4.5, 0)$	$(2.29, 4.61)$	13.9
NEW8(7)	13	$7.35 \cdot 10^{-7}$	$3.45 \cdot 10^{-6}$	$(-5.9, 0)$	$(0, 2.91)$	11.7
NEW8(6)	12	$7.35 \cdot 10^{-7}$	$3.45 \cdot 10^{-6}$	$(-5.9, 0)$	$(0, 2.91)$	11.7
NEW8(5)	11	$8.87 \cdot 10^{-6}$	$2.02 \cdot 10^{-5}$	$(-6.78, 0)$	$(0, 2.13)$	42.8
NEW8(4)	11	$8.87 \cdot 10^{-6}$	$2.02 \cdot 10^{-5}$	$(-6.78, 0)$	$(0, 2.13)$	42.8
PHNW8(5)(3)	12	$6.26 \cdot 10^{-6}$	$1.35 \cdot 10^{-5}$	$(-6.32, 0)$	$(0, 5.96)$	43.5
PHNW8(6)	12	$6.26 \cdot 10^{-6}$	$1.35 \cdot 10^{-5}$	$(-6.32, 0)$	$(0, 5.96)$	43.5
HA10(6)	18	$5.27 \cdot 10^{-6}$	$1.72 \cdot 10^{-5}$	$(-2.7, 0)$	$(0, 1.16)$	1.05
HA11(10)	50	$7.10 \cdot 10^{-9}$	$2.13 \cdot 10^{-8}$	$(-4.66, 0)$	—	1.05

I_{IM}: Imaginary Stability Interval,
I_R: Real Stability Interval,
$D_\infty = \max\left(\max_{i,j}\ a_{ij}, \|b\|_\infty, \left\|\hat{b}\right\|_\infty, \|c\|_\infty\right)$,
A, b, $\hat{b}$, and c are the defining parameters of a RK pair.

Many authors in the past have claimed RK pairs of orders higher than eight as not being particularly suited for practical purposes. However, there have not appeared any comparisons concerning pairs of this type for requested tolerances more stringent than 10^{-14}. Such accuracies are some times requested for problems in astronomy, high energy physics, molecular dynamics, etc. For this reason, as well because we want to test a new step-size selection algorithm to be discussed in the next section, we chose a 10th order, 17-stage RK method by Hairer [13] and we embedded a non-defective 6th order method, to be used for error estimation, by appending one extra stage. Table 2.7 contains the last column of A and the weights for this method. The remaining parameters may be found in [14] or [13]; higher precision data may be obtained from the current authors. As it is customary in the literature the numbers in these tables, except of three cases, are rational approximations accurate in 20 significant digits (this type of presentation allows the easy presentation of a method in a single Butcher tableau).

In conclusion, among the pairs discussed in this section, NEW6(4) corresponds to the absolute minimum of $\left\|T^{(7)}\right\|_2$. Concerning the pairs NEW8(6), NEW8(5), and NEW8(4) better values of $\left\|T^{(9)}\right\|_2$ might be obtained, however, at the expense of increasing D_∞ (not recommended).

3. A New Step-Size Selection Algorithm. There are currently two widely used methods that have appeared in the literature for changing the step-size of $p\,(q)$-

TABLE 2.2
Coefficients of the pairs NEW4(2) and NEW4(3).

0					
$\frac{5}{14}$	$\frac{5}{14}$				
$\frac{13}{22}$	$-\frac{52}{605}$	$\frac{819}{1210}$			
1	$\frac{2576}{4745}$	$-\frac{252}{365}$	$\frac{1089}{949}$		
1	$\frac{19}{130}$	$\frac{343}{1215}$	$\frac{1331}{3159}$	$\frac{73}{486}$	
4th-order	$\frac{19}{130}$	$\frac{343}{1215}$	$\frac{1331}{3159}$	$\frac{73}{486}$	
2th-order	$\frac{4}{55}$	$\frac{203}{990}$	$\frac{13}{18}$		
3th-order	$\frac{11}{130}$	$\frac{637}{1215}$	$\frac{605}{3159}$	$-\frac{73}{243}$	$\frac{1}{2}$

TABLE 2.3
Coefficients of NEW6(4) (exact rationals).

0							
$\frac{4}{27}$	$\frac{4}{27}$						
$\frac{2}{9}$	$\frac{1}{18}$	$\frac{1}{6}$					
$\frac{3}{7}$	$\frac{66}{343}$	$-\frac{727}{1372}$	$\frac{1053}{1372}$				
$\frac{11}{16}$	$\frac{13339}{49152}$	$-\frac{4617}{16384}$	$\frac{5427}{53248}$	$\frac{95207}{159744}$			
$\frac{10}{13}$	$-\frac{6935}{57122}$	$\frac{23085}{48334}$	$\frac{333633360}{273642941}$	$\frac{972160}{118442467}$	$\frac{172687360}{610434253}$		
1	$\frac{611}{1891}$	$-\frac{4617}{7564}$	$\frac{6041007}{13176488}$	$\frac{12708836}{22100117}$	$-\frac{3584000}{62461621}$	$\frac{6597591}{7972456}$	
6th-order	$\frac{131}{1800}$	0	$\frac{1121931}{392080}$	$\frac{319333}{1682928}$	$\frac{262144}{2477325}$	$\frac{4084223}{15177600}$	$\frac{1891}{25200}$
4th-order	$\frac{2694253}{26100360}$	0	$\frac{83647323}{535804360}$	$\frac{691202281}{1789061040}$	$-\frac{1275547648}{10565208225}$	$\frac{2}{5}$	$\frac{1891}{25200}$

order RK codes. The first is to apply the formula (see [15])

$$h_{n+1} = h_n \left(\frac{TOL}{EST_n} \right)^{\frac{1}{p}}, \tag{3.1}$$

where the new step-size sought $h_{n+1} = x_{n+1} - x_n$ is predicted in terms of an estimate of the local error EST_n which is based on the approximation

$$EST_n \approx y_n - \hat{y}_n, \tag{3.2}$$

assuming y_n , $\hat{y}_n$ to be the pth, qth order approximate solutions respectively at the previous grind-point x_n and TOL the requested tolerance. If

$$EST_{n+1} \leq TOL$$

then the computed solution y_{n+1} is accepted and the integration is carried on, otherwise (3.1) is re-evaluated by substituting $EST_n \rightarrow EST_{n+1}$. This methodology is termed error per step (EPS) mode (see Shampine [24]).

An alternative is to consider the same algorithm (3.1), but to use instead of (3.2), the approximation

$$EST_n \approx \frac{y_n - \hat{y}_n}{h_n}. \tag{3.3}$$

TABLE 2.4

Coefficients of NEW7(5). Rational approximations accurate in 20 *significant digits. The coefficients of the corresponding* 7(6) *method may be found in* [18].

$\frac{1}{18}$	$\frac{1}{18}$								
$\frac{1}{9}$	0	$\frac{1}{9}$							
$\frac{1}{6}$	$\frac{1}{24}$	0	$\frac{1}{8}$						
$\frac{89}{200}$	$\frac{2183971}{4000000}$	0	$-\frac{8340813}{4000000}$	$\frac{3968421}{2000000}$					
$\frac{56482}{115069}$	$\frac{695768212}{7463744411}$	0	$-\frac{1803549175}{7007942496}$	$\frac{3474507053}{6790877290}$	$\frac{2188198899}{15264927763}$				
$\frac{74}{95}$	$-\frac{11894934857}{8390623634}$	0	$\frac{53094780276}{9800512003}$	$-\frac{8415376229}{2277049503}$	$-\frac{18647567697}{10138317907}$	$\frac{27551494893}{11905950217}$			
$\frac{8}{9}$	$\frac{30828057951}{7654644085}$	0	$-\frac{4511704}{324729}$	$\frac{16217851618}{1651177175}$	$\frac{282768186839}{40694064384}$	$-\frac{104400780537}{15869257619}$	$\frac{5409241639}{9600177208}$		
1	$-\frac{133775720546}{36753383835}$	0	$\frac{49608695511}{4066590848}$	$-\frac{59896475201}{7901259813}$	$-\frac{48035527651}{5727379426}$	$\frac{86266718551}{10188951048}$	$-\frac{7751618114}{23575802495}$	$\frac{2289274942}{8464405725}$	
7th order	$\frac{597988726}{12374436915}$	0	0	$\frac{3138312158}{11968408119}$	$\frac{480882843}{7850665645}$	$\frac{988558885}{3512253271}$	$\frac{5302636961}{26425940286}$	$\frac{1259489433}{12163586030}$	$\frac{1016647712}{23899101975}$
5th order	$\frac{1421940313}{46193547077}$	0	0	$\frac{1943068601}{5911217046}$	$-\frac{3019049881}{6506827856}$	$\frac{7688913279}{9493187186}$	$\frac{586186883}{5187186385}$	$\frac{1114095023}{8014791121}$	$\frac{1016647712}{23899101975}$

TABLE 2.5

Coefficients of NEW8(6). Rational approximations accurate in 20 significant digits. The coefficients of the corresponding 8(7) method may be found in [18].

0												
$\frac{9}{142}$	$\frac{9}{142}$											
$\frac{24514}{238491}$	$\frac{178422123}{9178574137}$	a_{32}										
$\frac{12257}{79497}$	$\frac{12257}{317988}$	0	a_{43}									
$\frac{50}{129}$	$\frac{2584949729}{6554704252}$	0	a_{53}	$\frac{26222057794}{17776421907}$								
$\frac{34}{73}$	$\frac{4418011}{96055225}$	0	0	$\frac{2947922107}{12687381736}$	$\frac{3229973413}{17234960414}$							
$\frac{23}{148}$	$\frac{2875139539}{47877267651}$	0	0	$\frac{2702377211}{24084535832}$	$-\frac{135707089}{4042230341}$	$\frac{299874140}{17933325691}$						
$\frac{142}{141}$	$-\frac{7872176137}{5003514694}$	0	0	$-\frac{35136108789}{26684798878}$	$-\frac{114433184681}{9760995895}$	$\frac{299204996517}{32851421233}$	$\frac{254}{39}$					
$\frac{14920944853}{17030299364}$	$-\frac{3559950777}{7399971898}$	0	0	$-\frac{29299291531}{4405504148}$	$-\frac{42434013379}{9366905709}$	$\frac{20642871700}{5300635453}$	$\frac{12951197050}{1499985011}$	$\frac{59527523}{6331620793}$				
$\frac{83}{91}$	$-\frac{8196723582}{10570795981}$	0	0	$-\frac{46181454005}{5775132776}$	$-\frac{196277106011}{29179424052}$	$\frac{63575135343}{11491868333}$	$\frac{9348448139}{857846776}$	$\frac{195434294}{9727139945}$	$-\frac{617468037}{15757346105}$			
$\frac{143}{149}$	$-\frac{6373809055}{5357779452}$	0	0	$-\frac{150772749657}{21151088080}$	$-\frac{58076657383}{6089469394}$	$\frac{9252721190}{1221566797}$	$\frac{132381309631}{11748965576}$	$\frac{704633904}{13813696331}$	$\frac{656417033}{8185349658}$	$-\frac{1669806516}{10555289849}$		
1	$-\frac{2726346953}{6954959789}$	0	0	$\frac{24906446731}{6359105161}$	$-\frac{65277767625}{23298960463}$	$\frac{39128152317}{16028215273}$	$-\frac{40638357893}{16804059016}$	$-\frac{7437361171}{21911114743}$	$\frac{1040125706}{5334949109}$	$-\frac{1129865134}{5812907645}$	$\frac{6253441118}{10543852725}$	
8th order	$\frac{438853193}{9881496838}$	0	0	0	0	$\frac{11093525429}{31342013414}$	$\frac{481311443}{1936695762}$	$-\frac{3375294558}{10145424253}$	$\frac{9830993862}{5116981057}$	$-\frac{138630849943}{50747474617}$	$\frac{7152278206}{5104393345}$	b_{12}
6th order	$\frac{289283091}{6008696510}$	0	0	0	0	$\frac{3034152487}{7913336319}$	$\frac{7170564158}{30263027435}$	$\frac{7206303747}{16758195910}$	$-\frac{1059739258}{8472387467}$	$\frac{16534129531}{11550853505}$	$-\frac{3}{2}$	$\widehat{b}_{12}$

$$a_{32}=\frac{685501333}{8224473205}, a_{43}=\frac{12257}{105996}, a_{53}=-\frac{9163901916}{6184003973}, b_{12}=\widehat{b}_{12}=\frac{5118195927}{53798651926}$$

TABLE 2.6
Coefficients of the pairs NEW8(5) and NEW8(4) (rational approximations accurate in 21 significant digits).

0											
$-\frac{1}{25}$	$-\frac{1}{25}$										
$\frac{43}{381}$	$\frac{78991}{290322}$	$-\frac{46225}{290322}$									
$\frac{43}{254}$	$\frac{43}{1016}$	0	$\frac{129}{1016}$								
$\frac{209}{500}$	$\frac{1697713059}{4222509269}$	0	$-\frac{15238032203}{10156496298}$	$\frac{11056598884}{7292015089}$							
$\frac{1}{2}$	$\frac{5543}{107844}$	0	0	$\frac{2048383}{8149188}$	$\frac{1953125}{9902211}$						
$\frac{3512968824}{20344613659}$	$\frac{968421479}{14765520605}$	0	0	$\frac{956894283}{7559277968}$	$-\frac{465115410}{11816446109}$	$\frac{91302285}{4596652571}$					
$\frac{16831644835}{20344613659}$	$-\frac{134489695}{1465284848}$	0	0	$-\frac{95668987870}{6901605883}$	$-\frac{34399893283}{12958171610}$	$\frac{25465019788}{10579016529}$	$\frac{76986202126}{5122674515}$				
$\frac{1}{2}$	$\frac{145536625}{3474014636}$	0	0	$-\frac{13033589681}{17022116763}$	$\frac{55898639}{2339992721}$	$\frac{921475172}{7161215321}$	$\frac{11025931622}{10224678207}$	$-\frac{80727265}{11312405923}$			
$\frac{16831644835}{20344613659}$	$\frac{3439391366}{8230170613}$	0	0	$\frac{1368653752008}{33650418007}$	$\frac{19151417051}{2883993186}$	$-\frac{22521917029}{12057970022}$	$-\frac{953123275013}{22272368203}$	$\frac{3209473745}{8387593463}$	$-\frac{16775244890}{6391208017}$		
1	$\frac{1195929791}{15149569322}$	0	0	$-\frac{35554033801}{20785156544}$	$\frac{8903076353}{16738414228}$	$-\frac{80781378317}{13468382457}$	$\frac{28101089032}{14865674913}$	$\frac{1974790781}{11858655590}$	$\frac{20344613659}{3512968824}$	$\frac{7562197625}{30319520681}$	
8th-order	$\frac{1}{20}$	0	0	0	0	$\frac{7}{45}$	$\frac{49}{180}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{13}{180}$	$\frac{1}{20}$
5th-order	$\frac{1}{20}$	0	0	0	0	$-\frac{29}{45}$	$\frac{49}{180}$	$\frac{1}{5}$	1	$\frac{13}{180}$	$\frac{1}{20}$
4th-order	$\frac{2350230046}{49054484501}$	0	0	$\frac{3649218174}{13461577499}$	$\frac{545839447}{8426176087}$	$\frac{4566413657}{29908515761}$	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{13}{180}$	$\frac{1}{20}$

For the 8(5) pair the weights $\hat{b}_1$, $\hat{b}_6$, $\hat{b}_7$, $\hat{b}_8$, and $\hat{b}_{11}$ are used for satisfying the Vandermonde-type order conditions. The selection $\hat{b}_9 \neq b_9$ leads to a discrete 5th-order method, while the choice $\hat{b}_{10} = b_{10}$ (which leads to $\hat{b}_{11} = b_{11}$) reduces the function evaluation cost by two on each rejected step.

For the 8(4) pair respectively, $\hat{b}_1$, $\hat{b}_4$, $\hat{b}_5$, and $\hat{b}_6$ are also determined by the Vandermonde-type order conditions. We choose $\hat{b}_7 \neq b_7$ for differentiating the two formulas of the pair. By enabling $\hat{b}_8 = b_8$, $\hat{b}_9 = b_9$, $\hat{b}_{10} = b_{10}$, and $\hat{b}_{11} = b_{11}$ we save 4 additional function evaluations per rejected step.

TABLE 2.7

The 18th stage and the weights of a 6th order method embedded on the 10th order method by Hairer [14] *accurate in 35 significant digits (the coefficients not shown here are zero).*

$a_{181} = 0.12445832373808099546998998797622128 95$
$a_{184} = 0.11623459471215480044491475631609552 77$
$a_{185} = 0.55832690162183036227975081779574613 7$
$a_{1817} = 0.36045534484612611942365211362230357 0$
$\hat{b}_1 = 0.03936309050258979550561326646956641 38$
$\hat{b}_6 = -0.19159578689967917147734588646163453 9$
$\hat{b}_7 = -0.30094255173780179047919596278257355 0$
$\hat{b}_8 = 0.05028631001268821613329126296305208 46$
$\hat{b}_9 = 0.31295594458050420712864177697530131 4$
$\hat{b}_{10} = 0.27254498327898779738035125061626510 8$
$\hat{b}_{11} = 0.22745289524138973738724252162809142 7$
$\hat{b}_{12} = 0.11866725870014105544806033617260465 75$
$\hat{b}_{13} = 0.187$
$\hat{b}_{14} = 0.3$
$\hat{b}_{17} = -0.02573214367881984702665856558067291 51$
$\hat{b}_{18} = 0.01$

This is called error per unit step (EPUS) [24].

An heuristic argument in favor of the asymptotic validity of EPS mode for $p\,(p-1)$ pairs used in local extrapolation (or higher order) mode, may be found in Hairer et al. [14].

What is ideally expected by en efficient algorithm for step-size change is to

- allow a RK method to perform the integration with a as few as possible steps and a reasonable number of rejections;
- keep the maximal global error (ge) on the whole of the integration interval $[x_0, x_e]$ in direct proportionality to the requested tolerance.

The second of these requirements stems from the desire of code developers to allow the user an a priori knowledge of the global error behavior of the code with respect to the imposed tolerances. Assume a relation of the form

$$ge = C \cdot TOL^E \tag{3.4}$$

where C, E are constants (see Enright and Pryce [10]). If tolerance proportionality holds, then the user can easily adjust TOL in order to attain a predictable lower value of the global error. For example, assume that two integrations are performed on the same problem and by the same code by imposing TOL equal to 10^{-t}, 10^{-t-1} respectively. If the solutions at both tolerances agree at say d decimal digits, then we may trust the solution corresponding to the stringent tolerance as being accurate at $d+1$ decimal digits. Consequently $d+d'$ correct digits may hopefully be obtained by simply shifting to $TOL = 10^{-t-d'}$. This argument is explained as follows. Assuming $E = 1$, the respective global errors ge_1 and ge_2 satisfy the relations

$$\left.\begin{array}{c} ge_1 \approx C \cdot 10^{-t} \\ ge_2 \approx C \cdot 10^{-t-1} \end{array}\right\} \Rightarrow ge_2 \approx \frac{1}{10} ge_1,$$

and we may trust the solution provided at $TOL = 10^{-t-1}$ for accuracy at $d+1$ decimal

digits. Next, for $TOL = 10^{-t-d'}$ a similar argument shows that $ge_3 \approx 10^{-d'} ge_1$ and we expect that the third integration offers $d + d'$ accurate decimal digits.

In total, if a particular code is expected to offer global errors directly proportional to the requested tolerances, then it is very likely that just three integrations of a particular problem will lead to a solution accurate at any desired degree. Of course, in practice, a posteriori fourth integration is able to verify this expectation. Alternatively, if the global error of a code is not expected to be proportional to the requested tolerance this reasoning is destroyed and a (possibly large) number of trial and errors might be necessary for estimating the solution of a problem at a given accuracy.

In practice, tolerance proportionality holds whenever E is close to 1. For a step-size change algorithm which results to tolerance proportionality we expect asymptotically to hold $E \approx 1$ for any test problem (or at least we expect this to approximately hold for a sufficiently wide set of problems). In practical numerical experiments we have verified the observation of Shampine [25] that for $p\,(p-1)$ pairs, E is close to 1 only when LEM and EPS or a Lower Order Mode (LOM) and EPUS are used. In all other combinations it seems that tolerance proportionality does not hold. Clearly, the second of the above requirements may be interchanged with the exact knowledge of E in (3.4) (C being again problem dependent). An a priori determination of these constants seems to be adequate for the efficient global error prediction in a specific code (using the same reasoning as in the previous case when $E = 1$). We should note that both these requirements characterize a code implementation and not a method itself. Different choices (for example the application of (3.3) instead of (3.2)), in general, lead to different code behaviors. Sometimes this may result to tolerance proportionality being achieved, in some others it might destroy this property.

Some authors in the past have claimed lower order mode to be more reliable because this implementation supposedly provides a more accurate error estimation. We should note that both methods of implementation (LEM and LOM), under the estimator of the proper type, use the same $O\left(h^{q+1}\right)$ local error estimation, while the first of them just propagates a more accurate solution (Shampine [21]). In short, and in view of the tests performed in the next section, we will characterize (in accordance to [10]) the efficiency of a pair according to the function evaluation cost for a given problem and tolerance. The respective reliability will be primarily connected here to the ability of a method to perform the integration of a specific set of test problems with an average of the resulting global errors in direct proportionality to the requested tolerances. In the same respect, of a secondary importance will be the ability of the pair to induce relatively small values of the quantity $\max\left(h^{\beta}\frac{LE}{TOL}\right)$ (see equation 3.5 below; see also Table 4.2 and the relative footnote inside there). The reliability of a pair is thus to a larger extend related to the error estimator of a pair and the step-size selection algorithm, than the higher-order method itself.

Next assume we have a pair of orders $p\,(q)$, accepting also here the possibility $q > p$. A close inspection of (3.1) reveals that tolerance proportionality holds whenever the estimation EST (as provided by (3.2), (3.3)), is locally of the same order of h, as the order of the global error induced by the method that propagates the solution. Specifically $O\left(h^{p}\right)$ for $p\,(p-1)$ pairs used in LEM and $O\left(h^{p-1}\right)$ for the same type of pairs used in LOM. Consequently, we expect the same type of behavior for a general $p\,(q)$ code when the step-size change algorithm (3.1) is used with

$$EST_n = h^{p-min(p,q)-1}\left(y_n - \hat{y}_n\right) = h^{\beta}\left(y_n - \hat{y}_n\right). \tag{3.5}$$

The estimation provided by (3.2) is also in frequent use in $p\,(p-2)$ order Nystrom

pairs, where tolerance proportionality in general does not hold under this type of implementation. However, we expect this to happen under the estimation (3.5). A numerical example included in the next section supports this claim.

A problem occurring when comparing results with an order difference greater than 1, is that we loose a desirable scale invariance. Thus for example if we are solving a circuit problem with an independent variable t in units of seconds and after working on the same problem a bit, decide to change units to nanoseconds the error tolerance we should use for obtaining the same accuracy changes. So if we integrate twice a problem with independent variables t and λt, using a $p(p-1-\beta)$ method, then we must use tolerances TOL and $TOL\ /\ \lambda^{\beta}$ in order to achieve the same accuracy.

4. Numerical Tests—Conclusions. We distinguish the pairs that are going to be tested into five groups, primarily according to their effective number of stages, and secondly according to the order of each one of them. In the first group we include the pairs NEW4(3), SS3(2) [28], and NEW4(2) . In the second group we include the pairs SS5(4) and NEW6(4). In the third group we include the pairs NEW7(6), SS6(5) and NEW7(5). The pairs NEW8(7) and NEW8(5) are tested with respect to NEW8(6), all of them are members of group 4. The final group includes the pair HA11(10), a 10th order method with 17 stages (see [13]), applied as a pair using Richardson extrapolation with 50 stages (= 17 + 17 + 16; see Shampine [23]), HA10(6) and as a reference the pairs NEW8(6) from group 4 and NEW8(4). The latter is included here mainly for assessing the asymptotic validity of tolerance proportionality when using the local error estimation provided by (3.5) in the algorithm (3.1). We did not include in these tests 5(4) or 6(5) pairs, as these use 6 or 8 stages respectively and no new methods, among those presented here, utilize such a number of stages.

The efficiency gains comparisons were conducted among the guidelines of the tests performed in [28], [19] (see also Enright and Pryce [10]). We must note that since there is no stepsize limit on the DETEST implementation we used for the numerical results presented here, tolerance proportionality holds even for the problems of class C tested, [10]. So we included in all the tables of efficiency gain the problems of this class as well.

The pairs in [28], especially the third, fifth, and sixth order ones, may alternatively be considered as similar to those of Shanks if we just consider them as resulting under the employment of Shanks device on pairs of one order higher than that of which they actually are. (Thus essentially, the Sharp and Smart pais differ from those of Shanks only on the number of stages being used.) So in these tests we classified them according to their stages and not according to their order. All the methods of each group were tested for the same range of requested tolerances as shown in Table 4.1.

Before we proceed with efficiency comparisons among the various pairs, it is essential to check their reliability. There are two measures of reliability. Tolerance proportionality and local error estimation performance (see Sharp [27]). The outline of the results concerning tolerance proportionality are presented in Table 4.2. We present there the value $\overline{E}$ which is the average of the observed estimations of E for all 25 DETEST problems. This value must be as close to 1 as possible. In the next two columns we indicated the variance of E from 1 and $\overline{E}$, both of which have to be small enough. Then the average of mean square residuals [10] are recorded. Finally, $\max\left(h^{\beta}\frac{LE}{TOL}\right)$ over all tolerances and problems for each method is also given. This quantity exhibits a wide variation among the methods tested here. We see that the lower this quantity is, the greater the achieved accuracy by a specific pair, irrespectively of the requested tolerance (and consequently the function evaluation cost).

Recently, a different step-size selection algorithm has been used for the pair PHNW8(5)(3) in Hairer, Norsett, Wanner [14], based on a technique similar to one that has been proposed for numerical quadratures. We used the type Ia algorithm of [17] to reproduce the coefficients of this 8th order method and, for error estimation, we embedded, at no cost, a 6th order method. We name the resulting pair PHNW8(6). In ([14], page 255), the obligatory selection $b_{12} = \widehat{b}_{12}$, led to the rejection of that pair, since the error estimator uses no stage with $c_i = 1$. This may cause problems if a discontinuity lays just before 1. Then as an alternative, the authors proposed the pair PHNWH8(5)(3). On the contrary, NEW8(6) simply uses c_i's very close to 1, while PHNW8(6) do not. The NEW8(6) results over the problem EULR [14] were satisfactory enough and encourages us to further exploit of $p\,(p-1-\beta)$ pairs. In the left picture of figure 10.7 of the same book [14], there is also a good example of how tolerance proportionality is lost when a traditional step-size algorithm is used in conjunction with an 8(6) pair. Here, we shall also assume the opportunity to test the step-size change algorithm in [14] with respect to that provided by (3.5) for the range of tolerances used for testing the pairs of group 4.

We additionally tested on the D class of the DETEST problems [15], a 8(6) Nystrom pair by Dormand, El-Mikkawy and Prince [7], when both the local error estimation of (3.2) and (3.5) are used. For a range of requested tolerances $10^{-5},\ldots,$ 10^{-11}, we observed the second of these implementations to be 0.6% more efficient. While the estimation (3.2) gives a value of $\bar{E} = 1.1255$ (see Table 4.2), (3.5) yields $\bar{E} = 0.981$, which, as expected, suggests that tolerance proportionality in the latter case holds. Hence, this type of implementation seems to be more preferable in the case of $p\,(p-2)$ order Nystrom pairs (mainly for reasons of tolerance proportionality, than for those of efficiency gain).

If we intent to propose pairs of orders $p\,(p-3)$ or $p\,(p-4)$ as a serious alternative to the traditional ones, we need further numerical evidence about how the local error in the higher order formula relates to the local error in the lower order formula. Consequently, according to the tests developed by Sharp [27], we proceed with a second reliability check, in order to test the accuracy of the local error estimate. For this reason we selected the D-class test problems from DETEST, which imposes a severe test on the estimator.

For each problem, tolerance, and step-size we count the number of accepted steps for which

$$h^\beta \frac{\overline{LE}}{TOL} \le 2^{-5}, 2^{-5} \le h^\beta \frac{\overline{LE}}{TOL} \le 2^{-4}, \ldots, 2^{j-1} \le h^\beta \frac{\overline{LE}}{TOL} \le 2^{j}, \ldots, 2^{5} \le h^\beta \frac{\overline{LE}}{TOL},$$

where $\overline{LE}$ is the maximum norm of the true local error in the lower order formula. This data can be arranged in a histogram of twelve intervals. As in [27] we chose to present the cumulative percentage of each histogram. So, according to Table 4.3, the estimation of the local error of the 6th order formula of the pair NEW8(6), for the problem D3 and for tolerance 10^{-3}, indicates that 41% of the steps lay in the interval $[-\infty, 1/32]$ and 19% of the steps lay in the interval $[1/32, 1/16]$. This explains the number 60 under the -4 column. For this method we also observe that 12% of the steps lay in the interval $[1/16, 1/8]$, 16% in the interval $[1/8, 1/4]$, 6% in the interval $[1/4, 1/2]$ and finally 6% of the steps are in the interval $[1/2, 1]$. We observe that no steps were propagated with the true local error of the lower order formula being greater than $TOL\ /\ h^\beta$ (here, $\beta = 1$).

A reliable estimator requires all 100's to be in the 0-column and all numbers in -1 or -2 columns to be as small as possible. Great numbers in the left columns of these

tables, or even worse 100's to the right of 0-column, indicates possible poor reliability for the pairs under consideration. In order to measure this factor of reliability, we introduce the reliability index. If for a specific problem and tolerance, r_{-5}, r_{-4}, ..., r_6 are the cumulative percentages under the respective columns, then the corresponding reliability index is evaluated as

$$ri = r_{-5}\left(1-2^{-5}\right) + r_{-4}\left(1-2^{-4}\right) + \cdots + r_6(2^6-1) = \sum_{i=-5}^{i=6} \left|2^i - 1\right| r_i.$$

So the reliability index of a method with only one 100 under the zero-column is zero which is the best possible value. For example the RI of DP8(7) for $TOL = 10^{-3}$ and for the problem D1 is estimated to be RI$= \frac{31}{32}\cdot 7 + \frac{15}{16}\cdot 7 + \frac{7}{8}\cdot 25 + \frac{3}{4}\cdot 50 + \frac{1}{2}\cdot 88 + 0\cdot 100 = 116.72 \simeq 117$, as we see in the corresponding RI-column of Table 4.3.

Interpreting the RI values of Tables 4.3, 4.4, we conclude that the average reliability index of NEW8(6) is 138, while the corresponding value of PD87 is 187. These tests suggest that the estimators of $p(p-2)$ pairs seem to be at least as reliable as those of the conventional pairs. Furthermore the performance of HA10(6) motivates us to suggest this method for general use, since its RI seems to be no worse than that of famous methods, like FE54 or DP54 (see [27]).

The tables 4.3 and 4.4, involve information for accepted steps only. Therefore high percentage of rejected steps may not affect the RI values, even if rejections mean poor error estimation. Anyway, RI values were in direct proportionality with the number of rejected steps, in any test we carried out. On the other hand, re-evaluation of steps, surely decreases efficiency. So in order to avoid presenting more tables, we refer to the tables concerning efficiency for an indication about this matter.

From all the other tests we conducted concerning the efficiency of the new methods with respect to the older ones we include here Tables 4.6, 4.7, 4.8, and 4.9. Alternatively we include for all the RK methods of our study in Figures 4.1, 4.2, 4.3, 4.4, and 4.5 the graphical representation of the geometric mean of the maximum global errors over the whole integration interval, for all problems in each case $\left(\prod_{i=1}^{5} ge_{i,TOL}\right)^{1/5}$, against the geometric mean of the cost in function evaluations $\left(\prod_{i=1}^{5} fe_{i,TOL}\right)^{1/5}$ for each tolerance. This type of interpretation of DETEST comparisons is compatible with the tabular format used here and it is preferred over other types of graphical display for reasons explained in [16] and [19].

By studying all these tables and figures we notice that the new maximal order (equivalently minimal-stage) pairs are more competitive both than the pairs proposed in [28] and other $p\,(p-1)$ pairs from the literature. It seems that in the construction of RK pairs of orders $p(q)$ every effort should be made for obtaining pairs with as high a value of p as possible. The numerical results presented here seem to confirm the dominant role played by $\left\|T^{(p+1)}\right\|_2$ over $\left\|T^{(p+2)}\right\|_2$ on the performance of a pth order RK method. We may also state that in all these cases the non-defective pairs with a greater number of stages seem to compete better than those with fewer stages (provided they are tested at suitably stringent tolerances), something that is already clear from the numerical results presented in [28] or even earlier[1]. Of course this is in accordance with the known result that higher order methods outperform those of a

[1] The one exception concerns HA10(6) and HA11(10). However, this is exclusively due to the inefficient way that the latter pair was constructed (for illustration purposes).

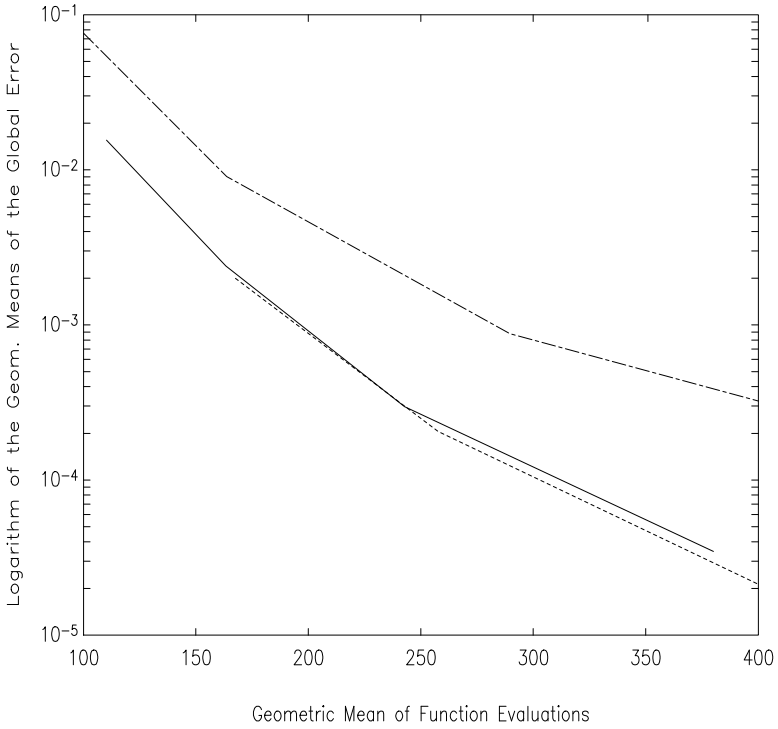

FIG. 4.1. —, *NEW*4(2); - - -, *NEW*4(3); *and* — - —, *SS*3(2).

TABLE 4.1

#group	Range of Tolerances
1	$10^{-2}, 10^{-3}, \ldots, 10^{-5}$
2	$10^{-3}, 10^{-4}, \ldots, 10^{-9}$
3	$10^{-5}, 10^{-6}, \ldots, 10^{-11}$
4	$10^{-5}, 10^{-6}, \ldots, 10^{-11}$
5	$10^{-10}, 10^{-12}, \ldots, 10^{-24}$

lower order (the former methods also require more stages) and it seems that it might be more appropriate to classify RK pairs according to the number of their stages and not just their order. The pairs 4(2) (or 4(3)), 6(4), 7(5), and 8(6) exhibit an appreciated performance for mild to more stringent accuracies and are suggested as good candidates for use by code developers. Runge-Kutta pairs of orders exceeding 8 seem to show their advantages when applied at requested tolerances higher than 10^{-14} and it seems that their use should be limited in practice on these situations only. Even the 10(6) pair must be taken seriously since according our results, seems to be an interesting alternative. We consider these efficiency and reliability tests as a starting point for a wider application of pairs selected according to the prominent criterion of utilizing, for a given order, a minimal number of stages, because of a reduced-cost embedded error estimator attained this way.

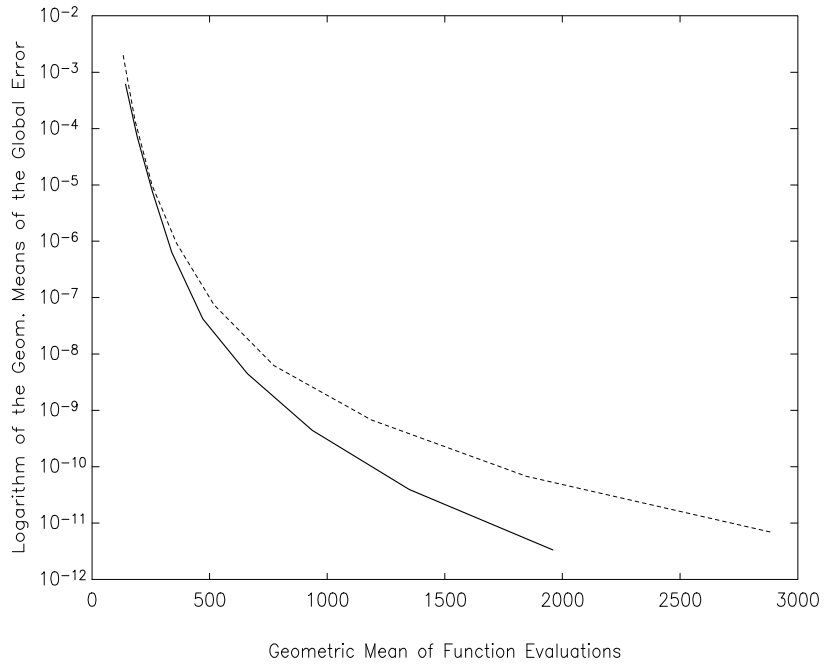

FIG. 4.2. —, $NEW6(4)$; *and* - - -, $SS5(4)$.

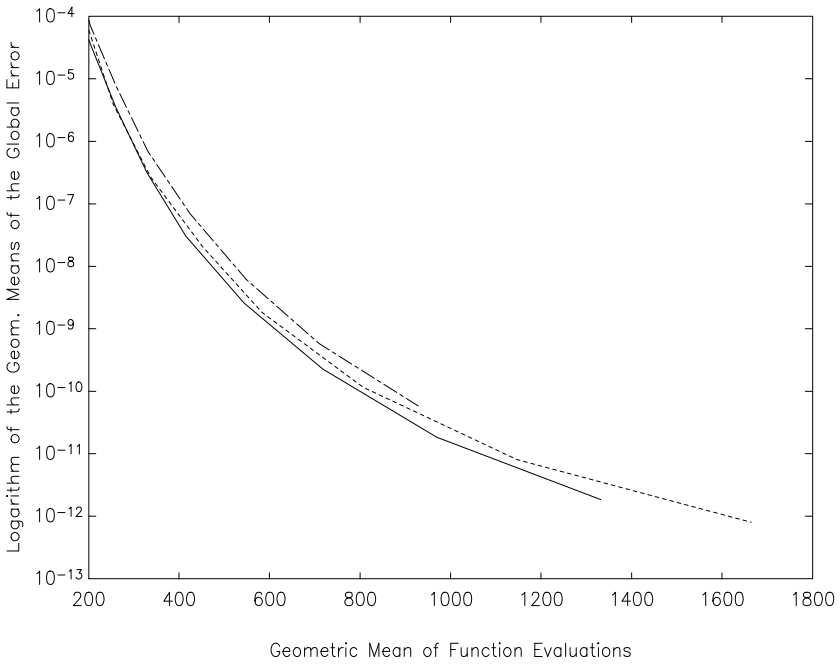

FIG. 4.3. —, $NEW7(5)$; - - -, $SS6(5)$; *and* — - —, $NEW7(6)$.

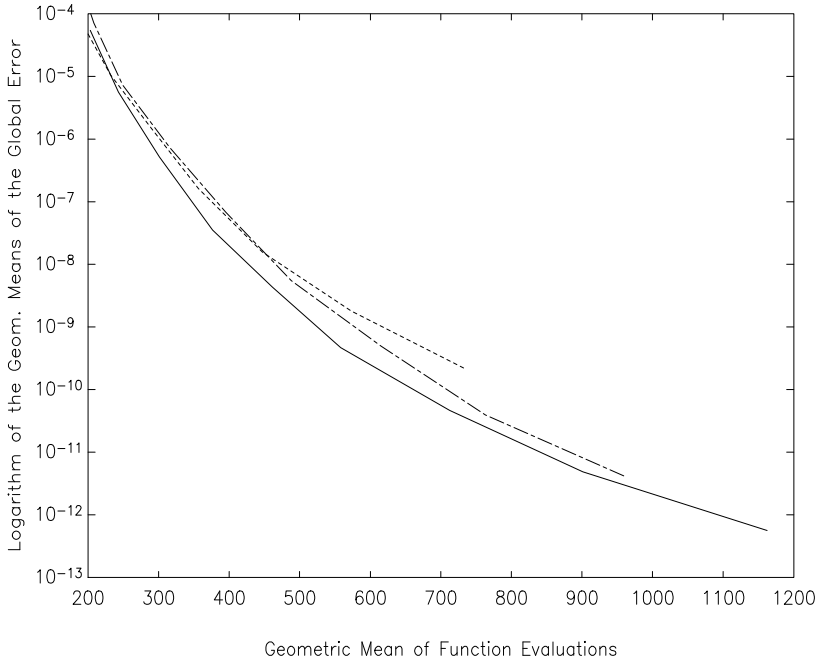

FIG. 4.4. —, $NEW8(6)$; — - —, $PD8(7)$; *and* - - -, $PHNW8(5)(3)$.

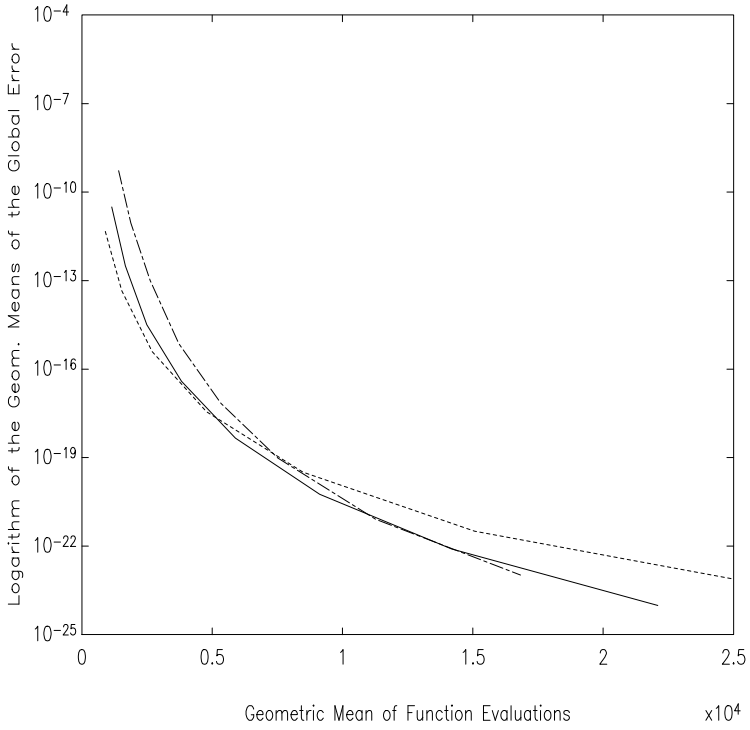

FIG. 4.5. —, $HA10(6)$; — - —, $HA11(10)$; *and* - - -, $NEW8(6)$.

TABLE 4.2
DETEST reliability statistics for the methods tested in this article.

	Method	$\bar{E}$	$\sum_{i=1}^{25} \frac{\lvert E_i - 1\rvert}{25}$	$\sum_{i=1}^{25} \frac{\lvert E_i - \bar{E}\rvert}{25}$	av (rms)	max $\left(h^\beta \frac{LE}{TOL}\right)$
	NEW4(3)	1.0089	0.0652	0.0638	0.0329	0.47
#1	SS3(2)	0.9945	0.1157	0.1153	0.0894	6.1
	NEW4(2)	0.8963	0.1423	0.1101	0.0636	0.1
#2	SS5(4)	1.0970	0.1255	0.0963	0.2329	2.1
	NEW6(4)	1.0458	0.0929	0.1025	0.1361	2.8
	NEW7(6)	1.0254	0.0896	0.0865	0.1213	5.9
#3	SS6(5)	1.0977	0.1284	0.1095	0.1680	0.05
	NEW7(5)	1.0578	0.0858	0.0849	0.1068	0.4
	PD8(7)	1.0572	0.0852	0.0746	0.1385	1.7
#4	NEW8(7)	0.9840	0.1133	0.1127	0.2033	2.9
	NEW8(6)	0.9931	0.1241	0.1227	0.1440	0.1
	NEW8(5)	1.0075	0.0813	0.0804	0.1170	1.8
#4	PHNW8(5)(3)	0.9514	0.1000	0.0914	0.1916	> 10
	PHNW8(6)	0.9923	0.0590	0.0590	0.1186	0.4
	NEW8(6)	1.0032	0.0342	0.0340	0.1318	0.03
#5	NEW8(4)	0.9912	0.0138	0.0134	0.0632	1.75
	HA10(6)	0.9659	0.0482	0.0497	0.2066	0.12
	HA11(10)	0.9933	0.0414	0.0407	0.2161	2.18
#3	NY8(6), eq. (3.2)	1.1255	0.1255	0.0335	0.1016	—
	NY8(6), eq. (3.5)	0.9810	0.0526	0.0488	0.1079	—

The various measures presented here concern the good behavior of the pairs under the present test conditions, and NOT their efficiency. More details may be found in [10] and [27].
LE is the global error of the local problem for the propagation formula.(see [15]).
The quantity max $\left(h^\beta \frac{LE}{TOL}\right)$ seems to be in direct proportionality with the numbers of *steps deceived* and *steps bad deceived* of DETEST (see [10]).

TABLE 4.3
Cumulative percentages histograms for Prince-Dormand 8(7) *method in the left and for NEW*8(6) *method in the right hand of the table.*

-5	-4	-3	-2	-1	0	1	RI	problem	TOL	-5	-4	-3	-2	-1	0	1	RI
7	7	25	50	88	100		117		10^{-3}			5	32	68	100		63
4	4	4	25	79	100		69	D1	10^{-6}				6	63	100		36
2	2	2	5	69	100		44		10^{-9}					77	100		39
34	39	45	56	84	95	100	293		10^{-3}		32	40	72	88	100		163
	8	31	62	100	100		131	D2	10^{-6}				45	89	100		78
			31	79	100		63		10^{-9}				7	63	100		37
63	67	71	71	84	88	100	381		10^{-3}	41	60	72	88	95	100		278
8	40	54	76	88	100		194	D3	10^{-6}			36	55	97	100		121
		9	48	89	100		88		10^{-9}				38	79	100		68
73	73	76	82	88	100		311		10^{-3}	60	73	78	88	95	100		308
18	48	62	84	92	100		226	D4	10^{-6}	5	29	47	71	95	100		174
		22	49	94	100		103		10^{-9}		1	2	46	91	100		83
85	88	90	93	97	100		362		10^{-3}	75	75	82	88	93	100		327
27	54	75	89	99	100		259	D5	10^{-6}	1	47	60	88	94	100		211
		33	54	95	100		117		10^{-9}	1	1	2	53	96	100		91

TABLE 4.4
*Cumulative percentages histogram for Hairer*10(6) *method.*

−5	−4	−3	−2	−1	0	1	*RI*	problem	*TOL*
			26	74	100		57		10^{-6}
			5	98	100		53	D1	10^{-9}
				99	100		45		10^{-12}
		20	76	94	100		70		10^{-6}
		1	39	72	100		66	D2	10^{-9}
			22	75	100		54		10^{-12}
	26	58	81	96	100		184		10^{-6}
		22	55	95	100		108	D3	10^{-9}
			39	75	100		67		10^{-12}
28	48	71	83	95	100		244		10^{-6}
	8	44	64	95	100		142	D4	10^{-9}
		3	49	81	100		80		10^{-12}
59	77	82	90	96	99	100	417		10^{-6}
18	39	52	85	98	100		210	D5	10^{-9}
		33	52	95	100		115		10^{-12}

REFERENCES

[1] J. C. Butcher, *On the attainable order of Runge-Kutta methods*, Math. Comp. **19** (1965), 408–417.

[2] ———, *The non-existence of ten stage eighth order explicit Runge-Kutta methods*, BIT **25** (1985), 521–540.

[3] ———, *The numerical analysis of ordinary differential equations*, John Wiley and Sons, Chichester, 1987.

[4] M. Calvo, J. I. Montijano, and L. Randez, *A new embedded pair of Runge-Kutta formulas of order* 5 *and* 6, Comput. Math. Appl. **20** (1990), 15–24.

[5] A. R. Curtis, *An eighth order Runge-Kutta process with eleven function evaluations per step*, Numer. Math. **16** (1970), 268–277.

[6] ———, *High order explicit Runge-Kutta formulae, their uses and their limitations*, J. Inst. Math. Appl. **16** (1975), 35–55.

[7] J. R. Dormand, M. E. El-Mikkawy, and P. J. Prince, *High order embedded Runge-Kutta-Nystrom formulae*, IMA J. Num. Analysis **7** (1987), 423–430.

[8] J. R. Dormand, M. R. Lockyer, N. E. McCorrigan, and P. J. Prince, *Global error estimation with Runge-Kutta triples*, J. Comput. Appl. Math. **7** (1989), 835–846.

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

[10] W. H. Enright and J. D. Pryce, *Two FORTRAN packages for assessing initial value methods*, ACM Trans. Math. Software **13** (1987), 1–27.

[11] E. Fehlberg, *Classical fifth, sixth, seventh, and eighth order Runge-Kutta formulas with stepsize control*, TR R–287, NASA, 1968.

[12] ———, *Low order classical Runge-Kutta formulas with stepsize control and their application to some heat-transfer problems*, TR R–315, NASA, 1969.

[13] E. Hairer, *A Runge-Kutta method of order* 10, J. Inst. Math. Appl. **21** (1978), 47–59.

[14] E. Hairer, S. P. Norsett, and G. Wanner, *Solving ordinary differential equations I*, second ed., Springer, Berlin, 1993.

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

[16] G. Papageorgiou, Ch. Tsitouras, and S. N. Papakostas, *Runge-Kutta pairs for periodic initial value problems*, Computing **51** (1993), 151–163.

[17] S. N. Papakostas, *On a class of families of high order Runge-Kutta methods and pairs*, submitted (1996).

[18] S. N. Papakostas and G. Papageorgiou, *A family of fifth order Runge-Kutta pairs*, Math. Comp. **65** (1996), 1165–1181.

TABLE 4.5
Cumulative per DETEST problem efficiency gains for all the methods tested here.

Method vs	Method	Total	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
NEW4(2)	NEW4(3)	1.6%	0	−4	−2	−4	1	4	0	−1	1	0	1	0	0	0	0	0	0	1	1	−1	0	2	0	2	3
NEW4(2)	SS3(2)	50%	4	4	3	2	12	11	2	1	−4	8	1	0	1	1	10	12	13	7	10	5	2	3	2	7	5
NEW6(4)	SS5(4)	15.7%	3	0	4	4	−2	2	2	2	−2	4	3	1	2	2	0	0	1	1	0	0	4	0	−1	6	2
NEW7(5)	NEW7(6)	13.1%	2	−1	3	1	−1	2	1	1	1	3	1	1	2	2	1	1	1	1	0	−1	1	2	3	2	5
NEW7(5)	SS6(5)	7.4%	0	−2	0	3	−2	3	1	1	5	−1	2	1	2	1	2	0	0	−1	−2	−2	2	2	1	1	1
NEW8(7)	PD8(7)	3.5%	−2	2	0	−1	0	0	0	0	0	1	−1	−1	0	0	−1	1	1	1	1	1	−2	1	3	2	3
NEW8(6)	NEW8(7)	7.5%	2	−2	4	2	0	1	1	1	1	0	1	1	1	1	1	1	2	0	0	−1	1	0	−1	0	1
NEW8(6)	NEW8(5)	−3.4%	−3	−1	1	−2	5	−2	−3	0	−1	−2	−2	−2	−2	−2	1	1	2	1	1	2	−2	1	0	−1	1
NEW8(6)	PHNW8(5)(3)	12.8%	−1	−1	7	2	1	2	−1	3	1	−1	0	1	0	1	1	2	2	1	0	0	−1	3	2	1	7
NEW8(6)	PHNW8(6)	6.5%	−1	2	2	1	3	0	−2	1	2	1	−1	−3	−1	−1	1	1	3	2	1	1	0	0	2	0	2
HA10(6)	NEW8(6)	2.9%	0	−4	3	9	1	2	−3	−1	3	2	−3	−6	−4	−4	0	−1	−1	−1	0	−2	3	1	4	7	3
HA10(6)	NEW8(4)	19.1%	−3	2	3	6	13	2	−5	1	1	2	−2	−7	−4	−4	1	1	1	3	4	4	0	3	4	7	14
HA10(6)	HA11(10)	22.8%	4	−1	3	3	−1	2	3	3	1	3	3	1	3	3	2	3	3	3	1	−1	4	3	2	3	2

Unity represents 10% and each number is rounded to the nearest integer. Positive numbers mean, that the first method is better.

TABLE 4.6
Efficiency gains of NEW4(2) relative to SS3(2), for the range of tolerances 10^{-2}, ..., 10^{-5}.

log global error	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
0																		4	6	5					
−1																8	9	10	13						
−2			1			11			−4	4		−1				15	17				1	3			
−3			3	2	11		1	0		11	0	−1	1	0	10						2	3	2		
−4	4	4	5	3	14		4	1			1	0	2	2							4		3	7	5
−5		5						2			3														
50%	4	4	3	2	12	11	2	1	−4	8	1	0	1	1	10	12	13	7	10	5	2	3	2	7	5

The final row, gives the mean value of efficiency gain for all tolerances in a problem. Empty places in the tables are due to the unavailability of data for the respective accuracies. Final row's first decimal number is the average efficiency gain for all problems in units of 1%.

TABLE 4.7
Efficiency gains of NEW6(4) relative to SS5(4), for the range of tolerances 10^{-3}, ..., 10^{-9}.

log global error	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
0																				−2					
−1																			−2	−1					
−2																		−1	−1	0					
−3						3						0				1	1	0	0	0					
−4			2		−1	2	1		−3	1		1	0	0	0	0	1	1	1	1	−1	1	−1		
−5	3	−1	3		−1	2	2	2	−3	2	1	1	1	1	0	0	1	1	2		0	0	−1		
−6	3	−2	4	3	−1	2	2	2	−2	4	2	1	1	1	0	0	1	2	2		2	0	−1		2
−7	3	−2	5	4	−2	3	3	2	−1	6	2	2	2	2	0	0	0	2			4	0	−1	4	1
−8	4	0	6	4	−2		3	2		8	3	2	3	3	0	0	0				7	0	−1	6	3
−9	4	1		4	−2		3	2			5		4	4	0						10			7	2
−10	3	2					4	2					5	5										8	3
15.7%	3	0	4	4	−2	2	2	2	−2	4	3	1	2	2	0	0	1	1	0	0	4	0	−1	6	2

[19] S. N. Papakostas, Ch. Tsitouras, and G. Papageorgiou, *A general family of explicit Runge-Kutta pairs of orders* 6(5), SIAM J. Numer. Anal. **33** (1996), 917–936.

[20] P. J. Prince and J. R. Dormand, *High order embedded Runge-Kutta formulae*, J. Comput. Appl. Math. **7** (1981), 67–75.

[21] L. F. Shampine, *Local extrapolation in the solution of ordinary differential equations*, Math. Comp. **27** (1973), 91–97.

[22] ———, *Quadrature and Runge-Kutta formulas*, Appl. Math. Comput. **2** (1976), 161–171.

[23] ———, *Local error estimation by doubling*, Computing **34** (1985), 179–190.

[24] ———, *Some practical Runge-Kutta formulas*, Math. Comp. **46** (1986), 135–150.

[25] ———, *Numerical solution of ordinary differential equations*, Chapman and Hall, New York, 1994.

[26] B. E. Shanks, *Solutions of differential equations by evaluations of functions*, Math. Comp. **20** (1966), 21–38.

[27] P. Sharp, *Numerical comparisons of some explicit runge-Kutta pairs of orders* 4 *through* 8, ACM Trans. Math. Software **17** (1991), 387–409.

[28] P. W. Sharp and E. Smart, *Explicit Runge-Kutta pairs with one more derivative evaluation than minimum*, SIAM J. Sci. Statist. Comput. **14** (1993), 338–348.

[29] J. H. Verner, *Explicit Runge-Kutta methods with estimates of the local truncation error*, SIAM J. Numer. Anal. **15** (1978), 772–790.

[30] ———, *A contrast of some Runge-Kutta formula pairs*, SIAM J. Numer. Anal. **27** (1990), 1332–1344.

[31] ———, *Some Runge-Kutta formula pairs*, SIAM J. Numer. Anal. **28** (1991), 496–511.

[32] ———, *Strategies for deriving new explicit Runge-Kutta pairs*, Annals of Numerical Mathematics **1** (1994), 225–244.

TABLE 4.8
Efficiency gains of NEW7(5) relative to SS6(5), for the range of tolerances 10^{-5}, ..., 10^{-11}.

log global error	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
−5																		−1	−2	−3					
−6						2										0	1	0	−1	−1					
−7			0			2	0		1	0		1	1	1		0	0	0	0		1	1			
−8	0		0	3	−3	2	1	−1	2	0	0	1	2	1	0	−1	−1	−1	−3		1	2	0		
−9	0		0	2	−2	3	0	0	4	−1	1	1	1	1	1	−1	−1	−2			2	2	1	0	3
−10	0	−3	0	3	−3	3	1	1	6	−1	2	1	2	2	2						3	3	1	1	1
−11	1	−2	0	3	−2	4	1	1	8	−1	2	2	2	2	2						3	3	2	2	0
−12	1	−2		4	−1		2	2	11		3		2	2	3						4		3	3	0
7.4%	0	−2	0	3	−2	3	1	1	5	−1	2	1	2	1	2	0	0	−1	−2	−2	2	2	1	1	1

TABLE 4.9
Efficiency gains of NEW8(6) relative to PHNW8(5)(3), for the range of tolerances 10^{-5}, ..., 10^{-11}.

log global error	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
−2																				−2					
−3																			−1	−1					
−4																			0	0					
−5																		1	0	0					
−6			6			1			0							2	2	1	0	0					
−7			8		1	2	−2		1	−2		0	−1	−1		2	2	2	0		−1	3	3		
−8	0	−2	9	1	1	3	−1	4	1	−1	−1	2	0	0	2	2	3	2	1		−1	2	2		6
−9	−1	−1		2	1		0	3		0	0		2	2	1						0	3	1	1	7
−10	−1	−1		2	1		−1	3		0	0				1							3	1	1	
−11	−2	0																							
−12		1																							
12.8%	−1	−1	7	2	1	2	−1	3	1	−1	0	1	0	1	1	2	2	1	0	0	−1	3	2	1	7

5. Appendix: The remaining detailed tables concerning the cumulative results of Table 4.5. ——

TABLE 5.1
Efficiency gains of NEW4(2) relative to NEW4(3), for the range of tolerances 10^{-2}, ..., 10^{-5}.

log global error	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
0																				−1					
−1																		1	1						
−2						4			1							0	0	1							
−3			−2		0					1					0	0	0				0	2	1		
−4			−3	−4	1		0	0		0	1	0	0	0	0						0		0	1	3
−5	0			−4	2		0	−1			1		0	0	0									2	3
−6		−4																						1	
1.6%	0	−4	−2	−4	1	4	0	−1	1	0	1	0	0	0	0	0	0	1	1	−1	0	2	0	2	3

TABLE 5.2
Efficiency gains of NEW7(5) relative to NEW7(6), for the range of tolerances 10^{-5}, ..., 10^{-11}.

log global error	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
−2																				−1					
−3																			1	−1					
−4																1	1	1	0	−2					
−5						1										1	1	1	−1	−1					
−6			2			1	1		1	3		1	1	1		1	1	1	0	−1		1	2		
−7	2		3		−1	2	0	2	1	3	1	1	2	2		1	1	1	0		1	2	2		
−8	2	0	3	0	−1	2	1	1	1	4	1	1	2	1	1	1	1	0	−2		1	3	2	3	1
−9	2	−1	4	1	−1	3	1	2	1	3	1	1	1	2	1						1	3	3	2	7
−10	1	−1		1	−1		1	1	2		1	0	1	2	1						1	2	3	2	6
−11	1	−1						1			1		1	2											
−12		−2																							
13.1%	2	−1	3	1	−1	2	1	1	1	3	1	1	2	2	1	1	1	1	0	−1	1	2	3	2	5

TABLE 5.3
Efficiency gains of NEW8(7) relative to DP8(7), for the range of tolerances 10^{-5}, ..., 10^{-11}.

log global error	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
−3																				0					
−4									−1							0	0	0	0	0					
−5						0			0							2	1	0	0	1					
−6			1			0	0		0	0		−3	0	0	0	2	0	0	1	1	−1		2		
−7	0	2	1	1	0	0	0	1	1	1	−1	−3	0	0	−1	1	1	1	1	1	−1	0	2	2	4
−8	−2	2	0	0	−1	1	1	1	1	1	−1	−1	1	1	−1	1	1	1	1	2	−2	1	4	4	3
−9	−3	2	−1	−2	0	1	−1	0	1	1	−1	−1	1	1	−1	1	1	1	1		−2	1	4	0	3
−10	−3	2	−1	−1	0		−2	−1		1	−1	−1	0	0	−1		1				−2	1	4	2	4
−11		1										0			−2							1		2	1
−12		1																							
3.5%	−2	2	0	−1	0	0	0	0	0	1	−1	−1	0	0	−1	1	1	1	1	1	−2	1	3	2	3

TABLE 5.4
Efficiency gains of NEW8(6) relative to NEW8(7), for the range of tolerances 10^{-5}, ..., 10^{-11}.

log global error	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
−2																				−1					
−3																			−1	−1					
−4																			−1	−1					
−5																		0	−1	−1					
−6						1			1							2	2	0	−1	−1					
−7		−1			0	1	1		2	−1		2	1	1		1	1	0	0	−2	1	0	0		
−8	1	−3	4	2	0	1	1	2	1	0	0	2	1	1	2	1	1	0	0	0	1	−1	−1		
−9	2	−2	4	2	1	2	2	1	1	1	0	1	1	1	1	1	2	0	1		1	0	−1	1	
−10	2	−2	3	2	1		1	1		0	1	1	1	1	1	1	2				1	0	−1	0	2
−11		−1										1			1							0		0	0
−12		0																							
7.5%	2	−2	4	2	0	1	1	1	1	0	1	1	1	1	1	1	2	0	0	−1	1	0	−1	0	1

TABLE 5.5
Efficiency gains of NEW8(6) relative to NEW8(5), for the range of tolerances 10^{-5}, ..., 10^{-11}.

log global error	A1 A2 A3 A4 A5	B1 B2 B3 B4 B5	C1 C2 C3 C4 C5	D1 D2 D3 D4 D5	E1 E2 E3 E4 E5
−2				2	
−3				1 2	
−4				1 3	
−5				1 1	
−6		−1 0		1 2 1 1	
−7	4	−2 −4 0 −3	−5 −3 −3	0 1 1 1	−2 0 0
−8	−3 2 3	−3 −2 0 −1 −2	−2 −3 −3 −3 1	1 1 0	−2 0 0
−9	−3 −2 2 −1 5	−2 −2 0 −2 −1	−2 −2 −2 −1 1	1 3 1	−2 1 0 1
−10	−3 −2 1 −2 6	−2 −3 0 −2 −1	−1 −2 −1 −1 1	1 4	−2 2 0 0 −1
−11	−4 −1 1 −2 8	−3 1 −3 −1	−1 −1 −1 −1 1		−2 0 −1 −2
−12	−4 0 1 −2	0	−2 −1 2		−2 4
−3.4%	−3 −1 1 −2 5	−2 −3 0 −1 −2	−2 −2 −2 −2 1	1 2 1 1 2	−2 1 0 −1 1

TABLE 5.6
Efficiency gains of NEW8(6) relative to PHNW8(6), for the range of tolerances 10^{-5}, ..., 10^{-11}.

log global error	A1 A2 A3 A4 A5	B1 B2 B3 B4 B5	C1 C2 C3 C4 C5	D1 D2 D3 D4 D5	E1 E2 E3 E4 E5
−2				0	
−3				0 1	
−4				0 2	
−5				2 1 2	
−6		0 1		2 2 2 1	
−7	2 2	0 −2 2 0	−4 −2 −2	2 2 2 2	0 0 1
−8	−1 2 2 2 1	0 −1 1 2 0	−1 −3 −1 −1 1	1 2 2	0 0 2
−9	−1 2 2 1 3	1 −1 0 2 1	−1 −2 −1 0 0	0 4	0 1 2 1
−10	−1 2 2 1 3	−2 1 1	−1 −2 −1 0 1		1 2 1 3
−11	−2 2 2 1 3	1			0 −1
−12	2				−2 5
6.5%	−1 2 2 1 3	0 −2 1 2 1	−1 −3 −1 −1 1	1 3 2 1 1	0 0 2 0 2

TABLE 5.7
Efficiency gains of HA10(6) relative to NEW8(6), for the range of tolerances 10^{-10}, 10^{-12}, ..., 10^{-24}.

log global error	A1 A2 A3 A4 A5	B1 B2 B3 B4 B5	C1 C2 C3 C4 C5	D1 D2 D3 D4 D5	E1 E2 E3 E4 E5
−8				−5	
−10				−4 −4	
−12		−1 −7 0 −1	−8 −14 −10 −10	−4 −5 −4 −2 −2	−3 −2 −1
−14	−3 0 3 −1	0 −6 −5 1 0	−6 −10 −7 −7 −3	−2 −3 −2 −1 −1	−1 0 0
−16	−2 −6 1 5 0	1 −4 −3 2 1	−4 −7 −5 −5 −2	−1 −1 −1 1 1	0 1 2 2 0
−18	−1 −5 2 7 0	2 −2 −2 3 2	−2 −4 −3 −3 0	0 0 1 2	2 2 4 4 1
−20	0 −4 5 10 1	4 −1 −1 4 4	−1 −2 −1 −1 1	1 1 2 4	4 3 7 6 2
−22	2 −2 7 13 1	5 1 1 6 5	0 −1 0 0 2	2 3	7 5 10 8 4
−24	2 −2 17 2	2	2 3		9 11 5
−26	3				13 7
2.9%	0 −4 3 9 1	2 −3 −1 3 2	−3 −6 −4 −4 0	−1 −1 −1 0 −2	3 1 4 7 3

TABLE 5.8

*Efficiency gains of HA*10(6) *relative to NEW*8(4)*, for the range of tolerances* 10^{-10}, 10^{-12}, ..., 10^{-24}.

log global error	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
−4																				−2					
−6																				0					
−8																		−4	−1	3					
−10						−3			−3			−18				−3	−3	−1	1	5		−1			
−12			1			−1	−10		−1	−2		−11	−9	−9		−2	−1	1	3	7	−6	0	0		
−14	−8	−1	1	1	6	1	−8	−3	0	0	−7	−7	−7	−7	−2	−1	0	3	5	10	−4	2	1		
−16	−5	0	2	2	9	2	−5	−2	1	1	−5	−5	−4	−4	0	1	2	4	7		−2	3	3	1	8
−18	−3	1	3	4	12	4	−3	0	2	2	−3	−3	−2	−3	1	2	3	7	10		0	5	4	3	10
−20	−2	3	4	5	15	6	−1	1	3	4	−1	−1	−1	−1	2	3	5	9			2	6	6	6	12
−22	0	4	5	7	18	8	0	3	4	5	0	0	1	1	3	5					4	8	8	8	15
−24	1	5		9	21			4			1				5						6			10	18
−26				12																				13	22
19.1%	−3	2	3	6	13	2	−5	1	1	2	−2	−7	−4	−4	1	1	1	3	4	4	0	3	4	7	14

TABLE 5.9

*Efficiency gains of HA*10(6) *relative to HA*11(10)*, for the range of tolerances* 10^{-10}, 10^{-12}, ..., 10^{-24}.

log global error	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5
−8																		7	5	1					
−10						6			5							6	9	6	4	0		8			
−12			8			4	5		4	6	4	4	6	6		4	3	4	3	0	4	6	6		
−14	5	2	5	6	1	2	4	5	2	5	3	2	4	4	3	3	3	2	1	−1	4	4	4		
−16	5	0	3	3	0	1	3	4	1	3	3	1	3	3	2	3	2	1	−1	−4	4	2	2	5	8
−18	4	−1	2	2	−1	0	3	3	0	2	2	1	3	3	2	2	2	1	−2		4	1	1	4	7
−20	4	−1	0	1	−2	0	3	3	−1	2	2	0	2	2	1	2	1	0	−3		4	0	0	3	0
−22	3	−3	−1	1	−3		2	2	−2	1	2	0	2	2	0						4	−1	−1	2	−3
−24		−4			−4																				−4
22.8%	4	−1	3	3	−1	2	3	3	1	3	3	1	3	3	2	3	3	3	1	−1	4	3	2	3	2