NEW RUNGE–KUTTA ALGORITHMS FOR NUMERICAL SIMULATION IN DYNAMICAL ASTRONOMY

J. R. DORMAND and P. J. PRINCE

Department of Mathematics and Statistics, Teesside Polytechnic, Middlesbrough, Cleveland TS1 3BA, U.K.

(Received 10 October, 1977)

Abstract. Some new Runge–Kutta and Runge–Kutta–Nystrom algorithms are presented for the solution of ordinary differential equations of the initial value type. The methods are compared with others in integrating the equations of motion of the two body problem and are shown to offer advantages in efficiency. It is also demonstrated that the new methods can be 'tuned' to achieve some measure of global error control.

1. Introduction

Runge–Kutta embedding is now a popular method for the numerical solution of ordinary differential equations. Being a single step procedure it is relatively stable and hence particularly suitable for the simulation of long-period evolution in dynamical astronomy. We present here some formulae of the Runge–Kutta (RK) and Runge–Kutta–Nystrom (RKN) types which seem to have advantages in efficiency over those currently in use.

For the first order initial value problem

$$\dot{x} = f(t, x), \qquad x(t_0) = x_0, \tag{1}$$

the embedded RK $p(p+1)$ (or RK $p+1(p)$) algorithm has the form

$$\left.\begin{aligned} X(t_0 + h) &= x_0 + \sum_{i=1}^{v_p} b_i k_i, \\ \hat{X}(t_0 + h) &= x_0 + \sum_{i=1}^{v_{p+1}} \hat{b}_i k_i, \end{aligned}\right\} \tag{2}$$

where

$$k_i = hf\left(t_0 + c_i h, x_0 + \sum_{j=1}^{i-1} a_{ij} k_j\right), \qquad i = 1, 2, \ldots, v_{p+1}.$$

If $x(t_0 + h)$ represents the true value of x after step h, then

$$x(t_0 + h) = X + O(h^{p+1}) = \hat{X} + O(h^{p+2}), \tag{3}$$

and so the difference $(\hat{X} - X)$ represents an approximation for the leading error term of the lower order formula which can be used to estimate a new or subsequent step size for local error control. Fehlberg (1968, 1969) has established a number of formulae of type (2) ($p \leqslant 8$). The coefficients b_i, $\hat{b}_i$, a_{ij}, and c_i must satisfy equations of condition that may be obtained from Taylor expansions; these are given by

Celestial Mechanics **18** (1978) 223–232. *All Rights Reserved.*
Copyright © 1978 by D. Reidel Publishing Company, Dordrecht, Holland.

Butcher (1964) for $p \leqslant 8$. To obtain embedded formulae of type (2) one must solve two such sets of equations simultaneously.

Fehlberg presents his formulae as pth order methods, i.e., the value of $x(t)$ is estimated by X. We prefer to use $\hat{X}$ to estimate $x(t)$ hence implementing the formulae in $(p + 1)$th order mode where possible.

Fehlberg's formulae have been developed to give very small leading truncation error terms for the pth order formula. This can result in an underestimate of the actual truncation error since the higher order terms could be dominant for moderate h. For example the error of the lower order formula is given by

$$E = h^{p+1} \sum_{i=1}^{n_{p+1}} \alpha_i^{(p+1)} F_i^{(p+1)} + h^{p+2} \sum_{i=1}^{n_{p+2}} \alpha_i^{(p+2)} F_i^{(p+2)} + O(h^{p+3})$$

where the $F_i^{(p+1)}$ are the elementary differentials of order $p + 1$ of f and $\alpha_i^{(p+1)}$ are the leading truncation coefficients (Stetter, 1971). If the $\alpha_i^{(p+1)}$ are minimized then the $\alpha_i^{(p+2)}$ could become dominant if h is not small. To take a specific case, the Fehlberg RKF4 (Fehlberg, 1969) where $p = 4$, we have

$$\|\alpha_t^{(6)}\|_2 / \|\alpha^{(5)}\|_2 = 3.15.$$

Assuming that the elementary differentials $F_i^{(5)}$ and $F_i^{(6)}$ do not differ in magnitude then unless $h < 1/3.15$ the principal error term is not dominant.

It seems more appropriate to minimize the leading truncation terms for the higher order formula thus making $\hat{X}$ a better estimate of $x(t)$. This may not, however, make $(\hat{X} - X)$ a more realistic error estimate.

In order to provide flexibility in the new methods the coefficient b_{vp} has been left free. The leading error term of the lower order formula can be made proportional to this coefficient (see (6)) and therefore the value of $(\hat{X} - X)$ can be tuned to give reliable step-size and error control.

The above comments are also applicable to RKN formulae which may be used directly to solve the second order equation

$$\ddot{x} = f(t, x), \quad \text{with} \quad x(t_0) = x_0 \quad \text{and} \quad \dot{x}(t_0) = \dot{x}_0. \tag{4}$$

The embedded RKN $p + 1(p)$ algorithm may be written in the form

$$\left.\begin{aligned} X(t_0 + h) &= x_0 + \dot{x}_0 h + h^2 \sum_{i=1}^{v_p} b_i f_i \\ \hat{X}(t_0 + h) &= x_0 + \dot{x}_0 h + h^2 \sum_{i=1}^{v_{p+1}} \hat{b}_i f_i \\ \hat{\dot{X}}(t_0 + h) &= \dot{x}_0 + h \sum_{i=1}^{v_{p+1}} \hat{\dot{b}}_i f_i \end{aligned}\right\} \tag{5}$$

where

$$f_i = f(t_0 + c_i h, x_0 + \dot{x}_0 c_i h + h^2 \sum_{j=1}^{i-1} a_{ij} f_j), \quad i = 1, \ldots, v_{p+1}.$$

Using (3) we note that $(\hat{X} - X)$ can be used for error estimation and step-size control. In addition to (5) (or alternatively) the embedding could be achieved for $\dot{x}$. The equations of condition for formulae of type (5) have been presented by Fehlberg (1972) who has also derived similar formulae up to order 8.

2. Some New RK and RKN Formulae

In Tables I and II we reproduce fourth and fifth order embedded RK formulae. The parameter λ may be chosen arbitrarily but suitable values in this case are found

TABLE I

Runge–Kutta 4(3)T

c_i	a_{ij}				$\hat{b}_i$	b_i
0					$\frac{1}{6}$	$\frac{1}{6}$
$\frac{1}{2}$	$\frac{1}{2}$				$\frac{1}{3}$	$\frac{1}{3}$
$\frac{1}{2}$	0	$\frac{1}{2}$			$\frac{1}{3}$	$\frac{1}{3}$
1	0	0	1		$\frac{1}{6}$	$\frac{1}{6}-\lambda$
1	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	0	λ

TABLE II

Runge–Kutta 5(4)T

c_i	a_{ij}					$\hat{b}_i$	b_i
0						$\frac{17}{192}$	$\frac{17}{192}$
$\frac{1}{8}$	$\frac{1}{8}$					0	0
$\frac{1}{4}$	0	$\frac{1}{4}$				$\frac{64}{231}$	$\frac{64}{231}$
$\frac{4}{9}$	$\frac{196}{729}$	$-\frac{320}{729}$	$\frac{448}{729}$			$\frac{2187}{8960}$	$\frac{2187}{8960}$
$\frac{4}{5}$	$\frac{836}{2875}$	$\frac{64}{575}$	$-\frac{13\,376}{20\,125}$	$\frac{21\,384}{20\,125}$		$\frac{2875}{8448}$	$\frac{2875}{8448}$
1	$-\frac{73}{48}$	0	$\frac{1312}{231}$	$-\frac{2025}{448}$	$\frac{2875}{2112}$	$\frac{1}{20}$	$\frac{1}{20}-\lambda$
1	$\frac{17}{192}$	0	$\frac{64}{231}$	$\frac{2187}{8960}$	$\frac{2875}{8448}$	$\frac{1}{02}$	0 $\quad$ λ

TABLE III

Runge–Kutta–Nystrom 7(6)T

c_i	a_{ij}							$\hat{b}_i$	b_i	$\hat{\dot{b}}_i$
0								$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$
$\frac{1}{10}$	$\frac{1}{200}$							0	0	0
$\frac{1}{5}$	$\frac{1}{150}$	$\frac{1}{75}$						0	0	0
$\frac{3}{8}$	$\frac{171}{8192}$	$\frac{45}{4096}$	$\frac{315}{8192}$					0	0	0
$\frac{1}{2}$	$\frac{5}{288}$	$\frac{25}{528}$	$\frac{25}{672}$	$\frac{16}{693}$				$\frac{8}{45}$	$\frac{8}{45}$	$\frac{16}{45}$
$\frac{7-\sqrt{21}}{14}$	$\frac{1003-205\sqrt{21}}{12348}$	$\frac{-25}{90552}(751-173\sqrt{21})$	$\frac{25}{43218}(624-137\sqrt{21})$	$\frac{-128}{237699}(361-79\sqrt{21})$	$\frac{3411-745\sqrt{21}}{24696}$			$\frac{7}{360}(7+\sqrt{21})$	$\frac{7}{360}(7+\sqrt{21})$	$\frac{49}{180}$
$\frac{7+\sqrt{21}}{14}$	$\frac{793+187\sqrt{21}}{12348}$	$\frac{-25}{90552}(331+113\sqrt{21})$	$\frac{25}{43218}(1044+247\sqrt{21})$	$\frac{-128}{9745659}(14885+3779\sqrt{21})$	$\frac{3327+797\sqrt{21}}{24696}$	$\frac{-(581+127\sqrt{21})}{1722}$		$\frac{7}{360}(7-21)$	$\frac{7}{360}$	$(7-\sqrt{21})\frac{49}{180}$
1	$\frac{-(157-3\sqrt{21})}{378}$	$\frac{25(143-10\sqrt{21})}{2772}$	$\frac{-25}{3969}(876+55\sqrt{21})$	$\frac{1280}{596673}(913+18\sqrt{21})$	$\frac{-1}{2268}(1353+26\sqrt{21})$	$\frac{7}{4428}(1777+377\sqrt{21})$	$\frac{7(5-\sqrt{21})}{36}$	0	$-\lambda$	$\frac{1}{20}$
1	$\frac{1}{20}$	0	0	0	$\frac{8}{45}$	$\frac{7(7+\sqrt{21})}{360}$	$\frac{7(7-\sqrt{21})}{360}$	0	$+\lambda$	0

to be 1/10 for the RK4(3)T and 1/60 for the RK5(4)T. In each case the final function evaluation at the nth step is the same as the first at the $(n + 1)$th step. Thus the effective number of function evaluations per step are four and six respectively. The basis of the RK4(3)T is the familiar classic RK4 of Kutta. This idea has been used by Fehlberg (1972) and Bettis (1973) with regard to RKN methods. Zonneveld (1970) also uses an extra function evaluation in this manner for calculation of the last Taylor series term which is then used for step-size control.

Apart from λ, the RK5(4)T can be expressed in terms of c_2 and c_3, and these have been chosen to give small error terms for $\hat{x}$. Thus for $p = 5$, $\|\boldsymbol{\alpha}^{(6)}\|_2 = 1.12 \times 10^{-3}$ compared with 3.36×10^{-3} for the Fehlberg RKF4. For $p = 4$, $\|\boldsymbol{\alpha}^{(6)}\|_2/\|\boldsymbol{\alpha}^{(5)}\|_2 = 1.05$ and so the principal error term in the lower order formula will generally be dominant.

Table III shows a seventh order RKN formula; $\lambda = 1/20$ meets our requirements in this case. The leading local truncation error estimates in the lower order formulae are as follows:

$$\left.\begin{aligned} &\text{RK4(3)}T: && T = \lambda(k_4 - k_5) \\ &\text{RK5(4)}T: && T = \lambda(k_6 - k_7) \\ &\text{RKN7(6)}T: && T = \lambda h^2(f_8 - f_9). \end{aligned}\right\} \tag{6}$$

Following Hull *et al.* (1972) the nth step-size h_n is given by

$$h_n = 0.9h_{n-1}(\varepsilon/T_{\max})^{1/p}, \tag{7}$$

where p is the order of the lower order formula, ε is the tolerated local error per *unit step*, and $T_{\max}$ is given by (6). If it is required to control error per *step* the exponent in (7) should be $1/(p + 1)$. The step would be rejected if $T_{\max} > \varepsilon$ and recomputed with h given by (7).

3. Some Numerical Comparisons

As a suitable test problem we have considered the equations of two-body motion

$$\left.\begin{aligned} &\ddot{x} = -x/r^3, \quad x(0) = 1 - e, \quad \dot{x}(0) = 0, \\ &\ddot{y} = -y/r^3, \quad y(0) = 0, \quad \dot{y}(0) = \{(1 + e)/(1 - e)\}^{1/2}, \end{aligned}\right\} \tag{8}$$

where

$$r^2 = x^2 + y^2,$$

which are satisfied by an elliptic orbit of eccentricity e and major semi axis unity. Five values of e have been used, namely 0.1 (0.2) 0.9, which will be referred to as cases 1 to 5. For the application of RK methods equations (8) were reduced to four first order equations. In every case equations (8) were integrated from $t = 0$ to

$t = 20$, which is just over 3 times the orbital period. The analytical solution of (8), viz:

$$x = \cos u - e, \quad y = \sqrt{1 - e^2}\sin u,$$

$$\dot{x} = \frac{-\sin u}{1 - e\cos u}, \quad \dot{y} = \frac{\sqrt{1 - e^2}\cos u}{1 - e\cos u},$$

where u is given by $u - e\sin u = t$, was used to determine the global error at all stages of the numerical computation.

In addition to the algroithms described above, three RK methods given by Fehlberg (1968, 1969) were applied to the test problem. These were RKF4, RKF5 and RKF7 which require 6, 8, and 13 function evaluations per step respectively. Also tested were RKNF7 and RKNF8 (Fehlberg, 1972), together with RKN5(4) of Bettis (1973). This last method is unusual in giving 6th order local accuracy for x and only 5th for $\dot{x}$. These RKN formulae require 9, 11 and 5 function evaluations per step respectively.

We have solved the problem (8) for a wide range of tolerances and the results are presented as two series of efficiency curves (Figures 1 and 2) for the RK and RKN methods respectively. The maximum global error is taken over all the variables and all steps. The presentation of function evaluations rather than computer processing time makes the results independent of machine and programming techniques. To justify our preference of the higher order mode where possible we present results for the Fehlberg RKF7 in both modes. The other Felhberg formulae have generally similar behaviour. We do not present any results for the RK4(3)T since it has been found to be inferior to other formulae even with ε as high as 10^{-3}.

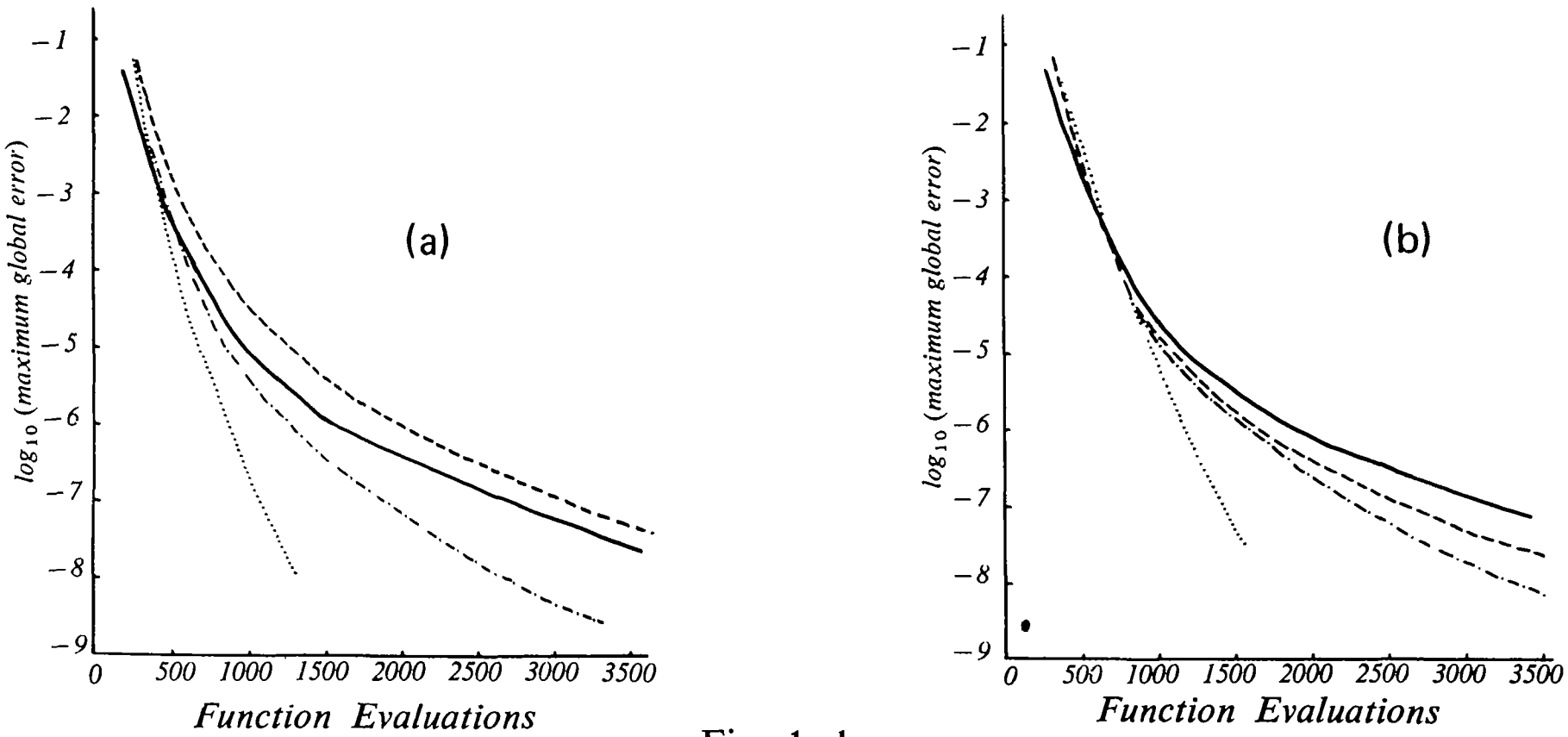
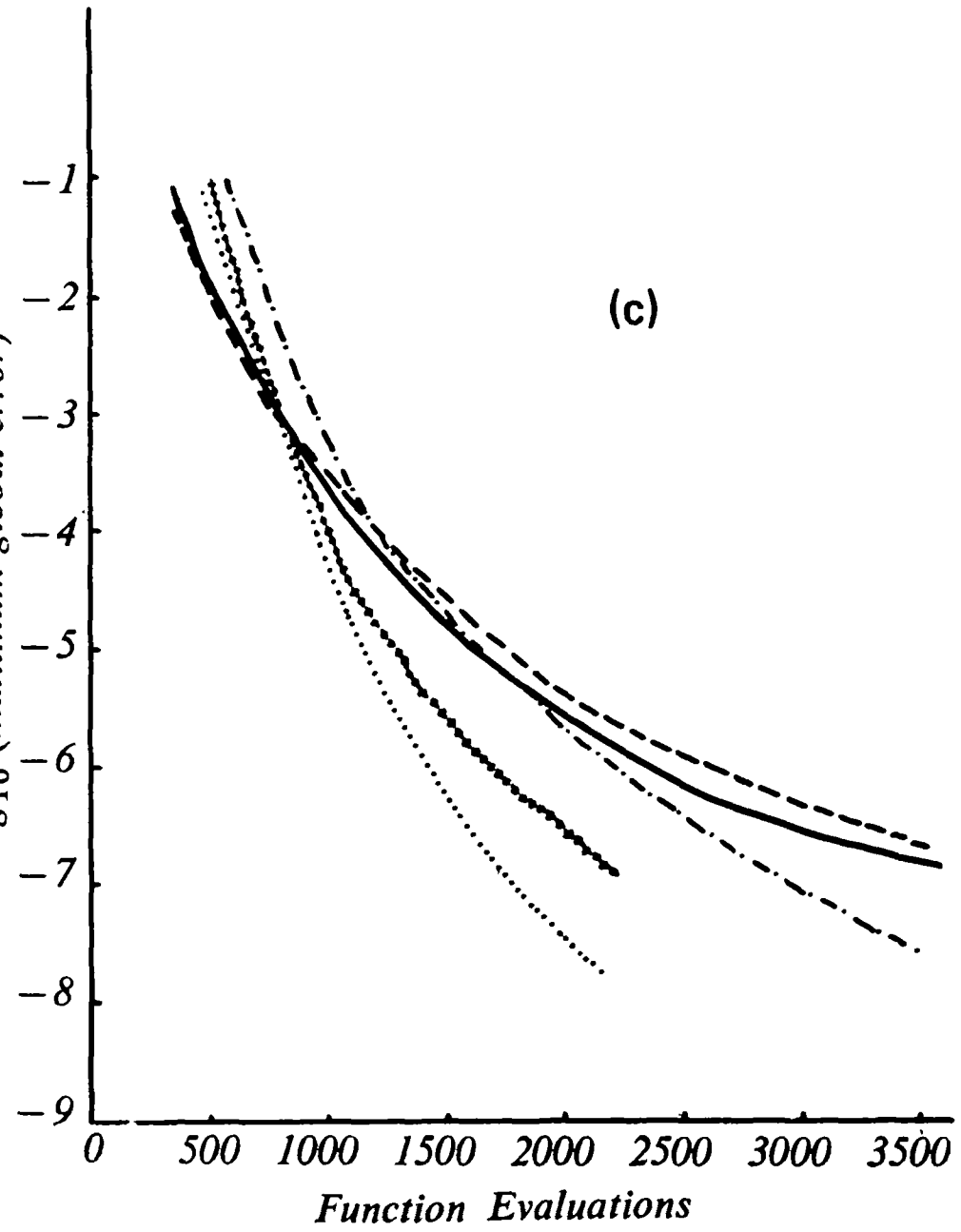
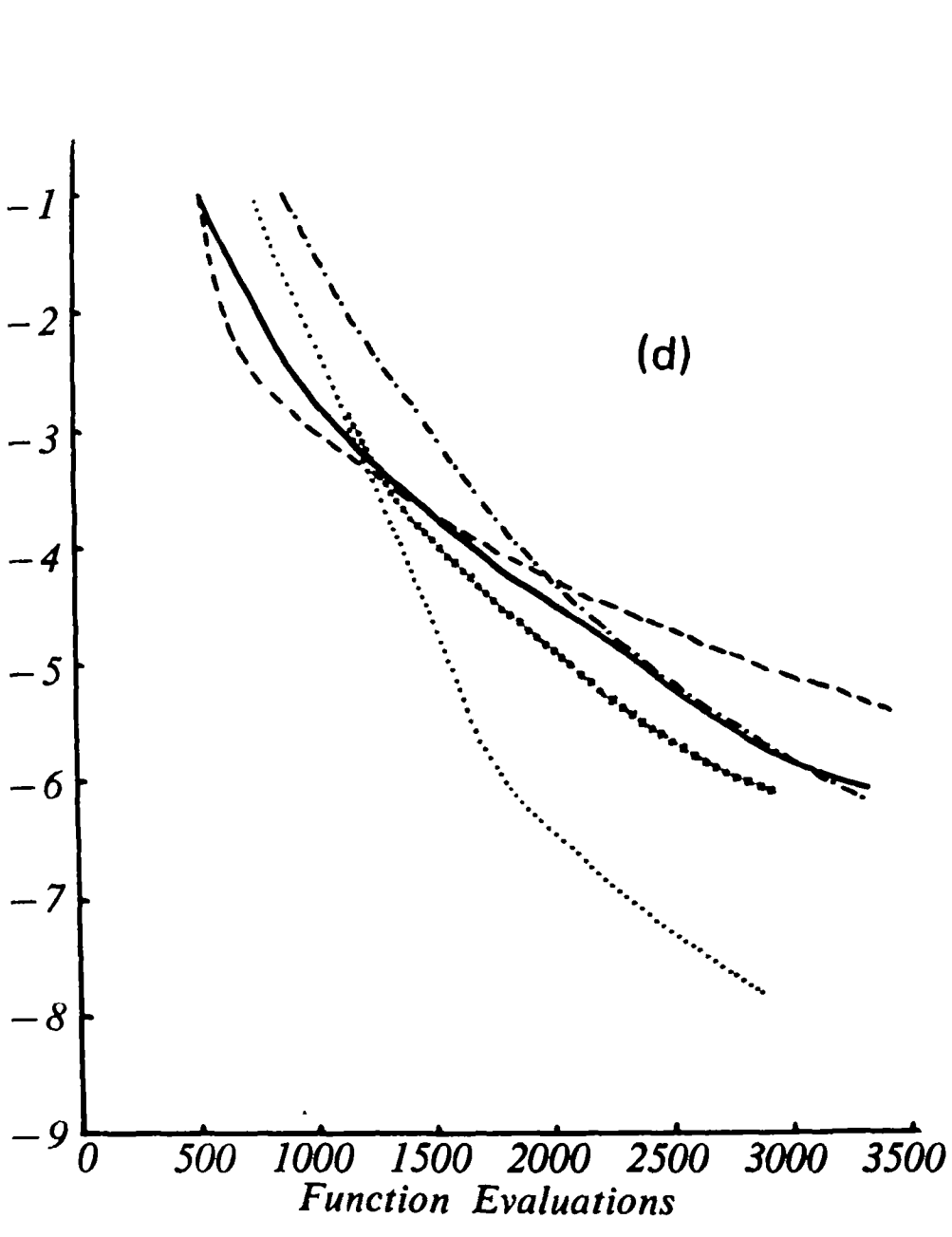


Fig. 1a-b.

Fig. 1a-e. Efficiency curves for Runge–Kutta methods applied to the two-body problem. (a) $e = 0.1$, (b) $e = 0.3$, (c) $e = 0.5$, (d) $e = 0.7$, (e) $e = 0.9$.

Key: ———— RK5(4)T — — — — — — — — RKF4
—.—.—.—.—.—.—.—. RKF5 . . . . . RKF7 (Eighth order mode)
+ + + + + + + + + + + + + + + + + + RKF7 (Seventh order mode)


(c)
log10 (maximum global error)
−1
−2
−3
−4
−5
−6
−7
−8
−9
0
500
1000
1500
2000
2500
3000
3500
Function Evaluations


(d)
−1
−2
−3
−4
−5
−6
−7
−8
−9
0
500
1000
1500
2000
2500
3000
3500
Function Evaluations

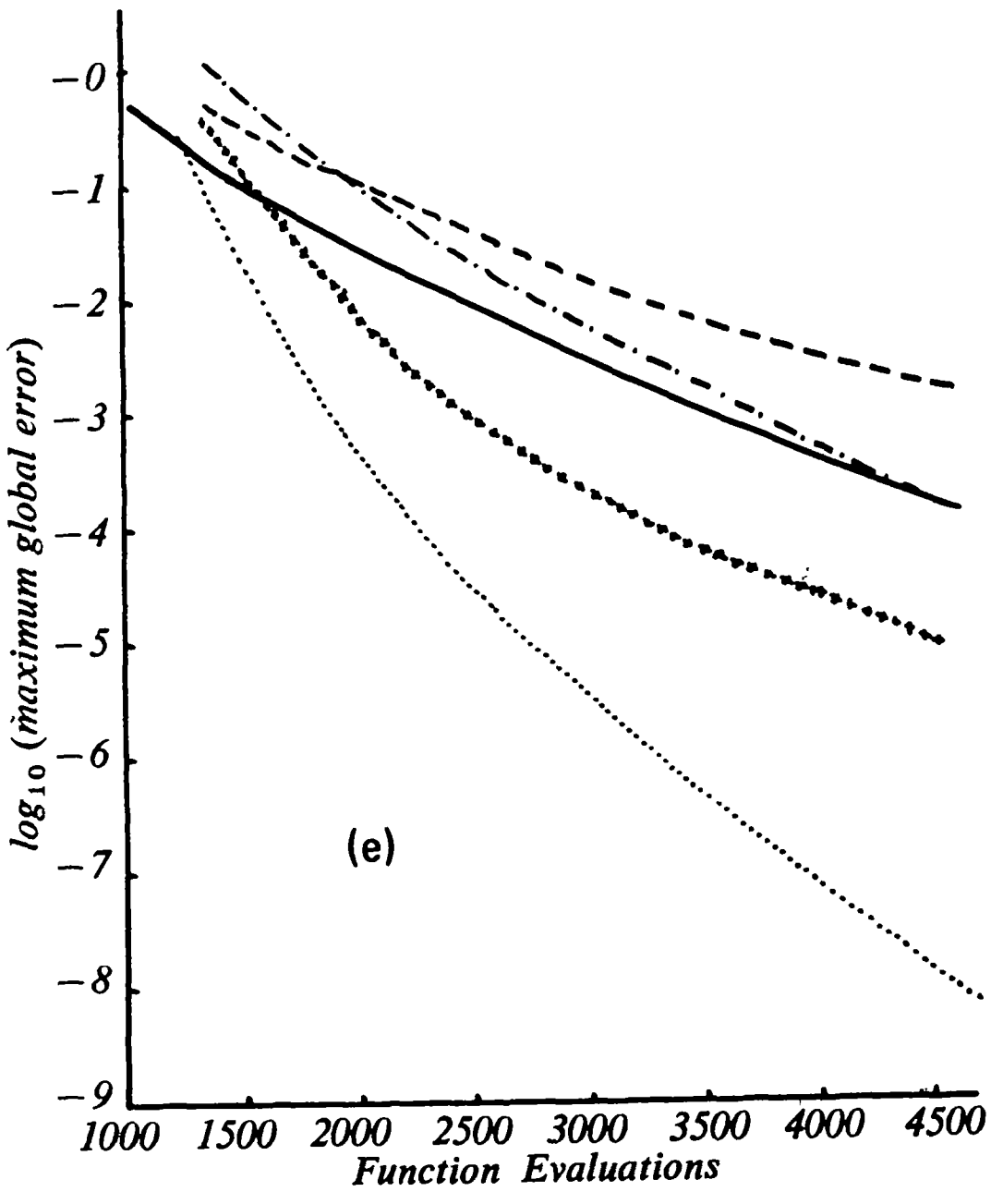
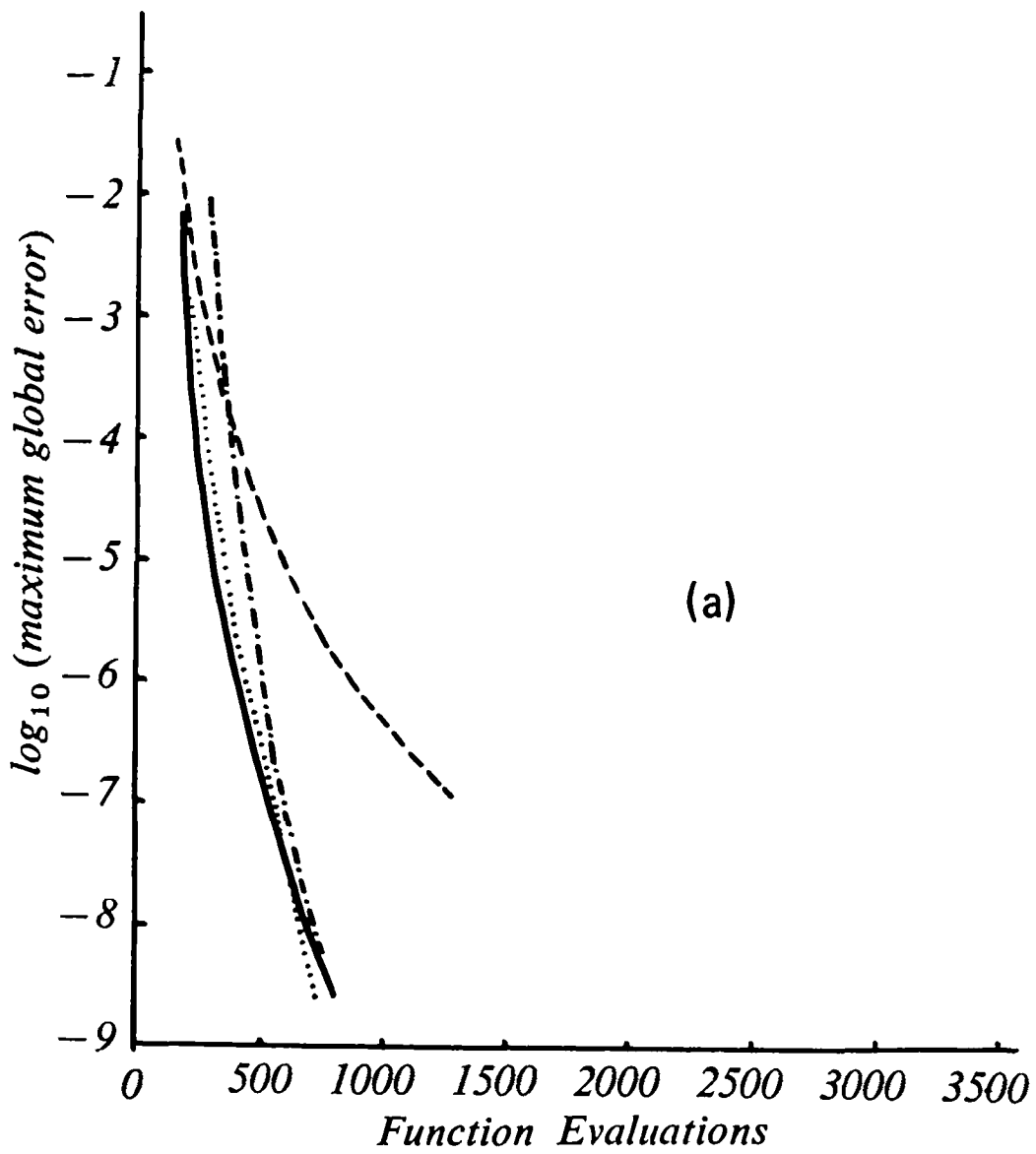
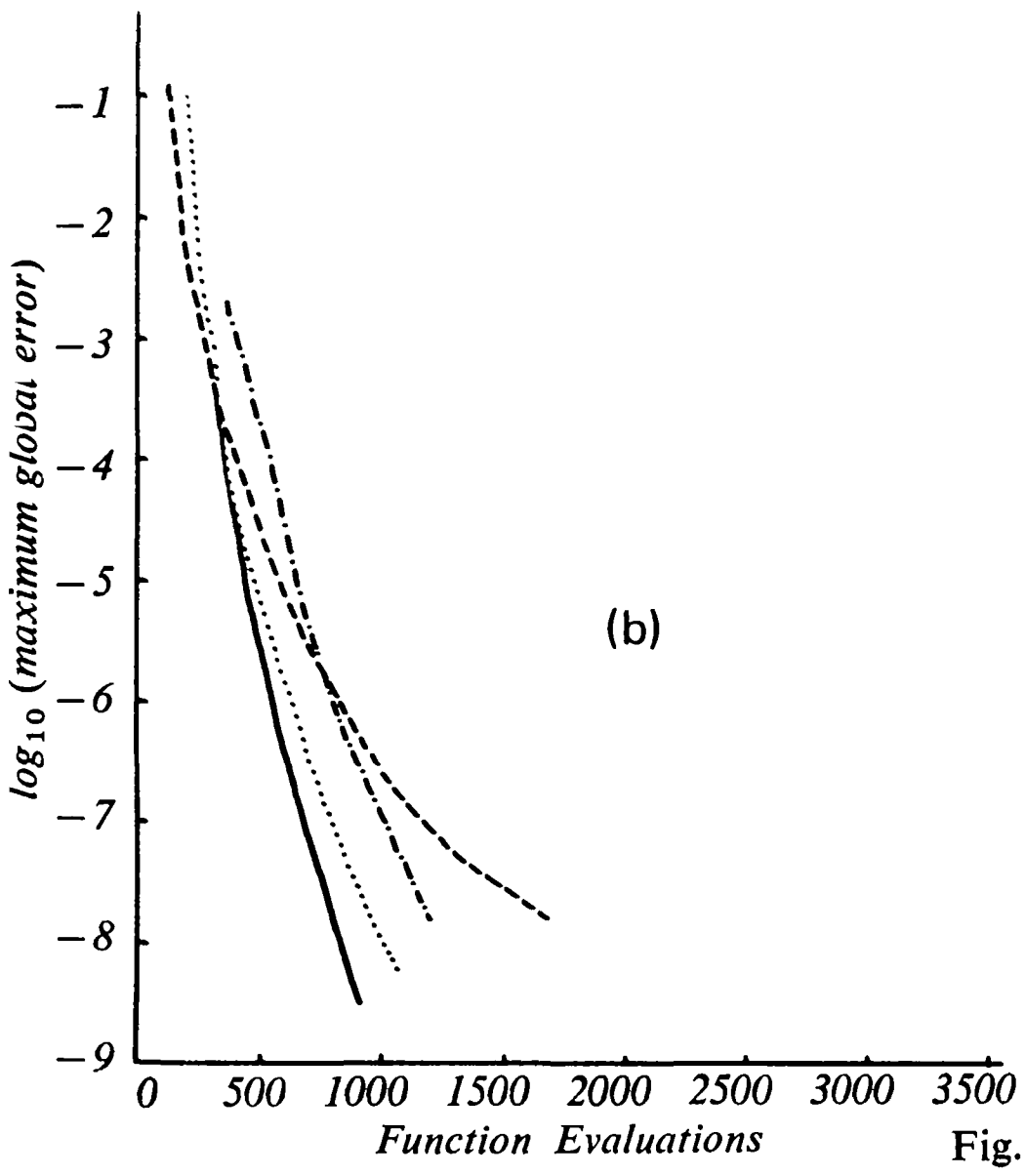

(e)
log10 (maximum global error)
−0
−1
−2
−3
−4
−5
−6
−7
−8
−9
1000
1500
2000
2500
3000
3500
4000
4500
Function Evaluations

Fig. 1c-e.





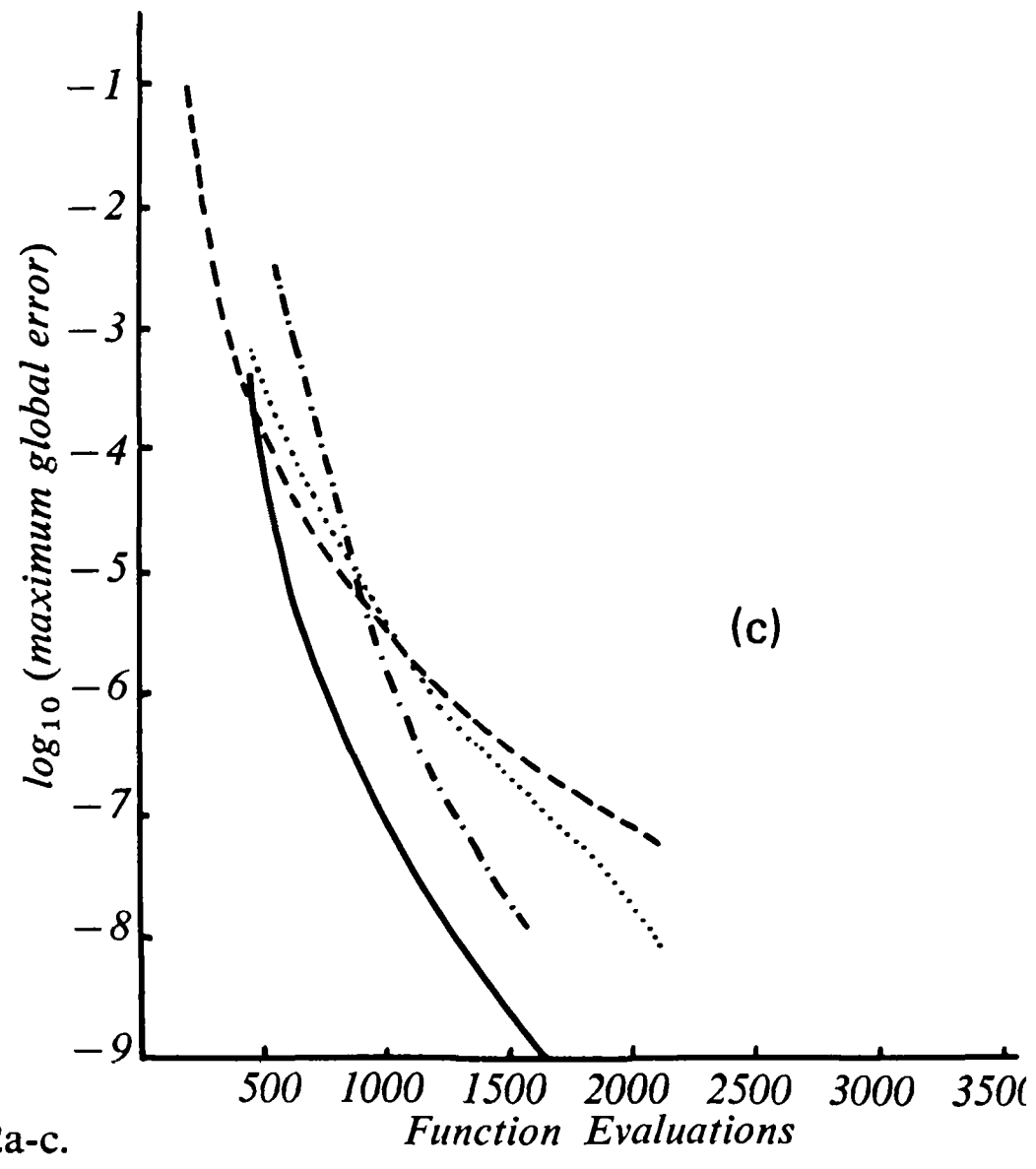


Fig. 2a-c.

Fig. 2. Efficiency curves for Runge–Kutta–Nyström methods applied to the two-body problem. (a) $e = 0.1$, (b) $e = 0.3$, (c) $e = 0.5$, (d) $e = 0.7$, (e) $e = 0.9$.

Key: ——————— RKN7(6)*T* — — — — — — — — RKN5(4)
—.—.—.—.—.—.—.—. RKNF8 . . . . . RKNF7

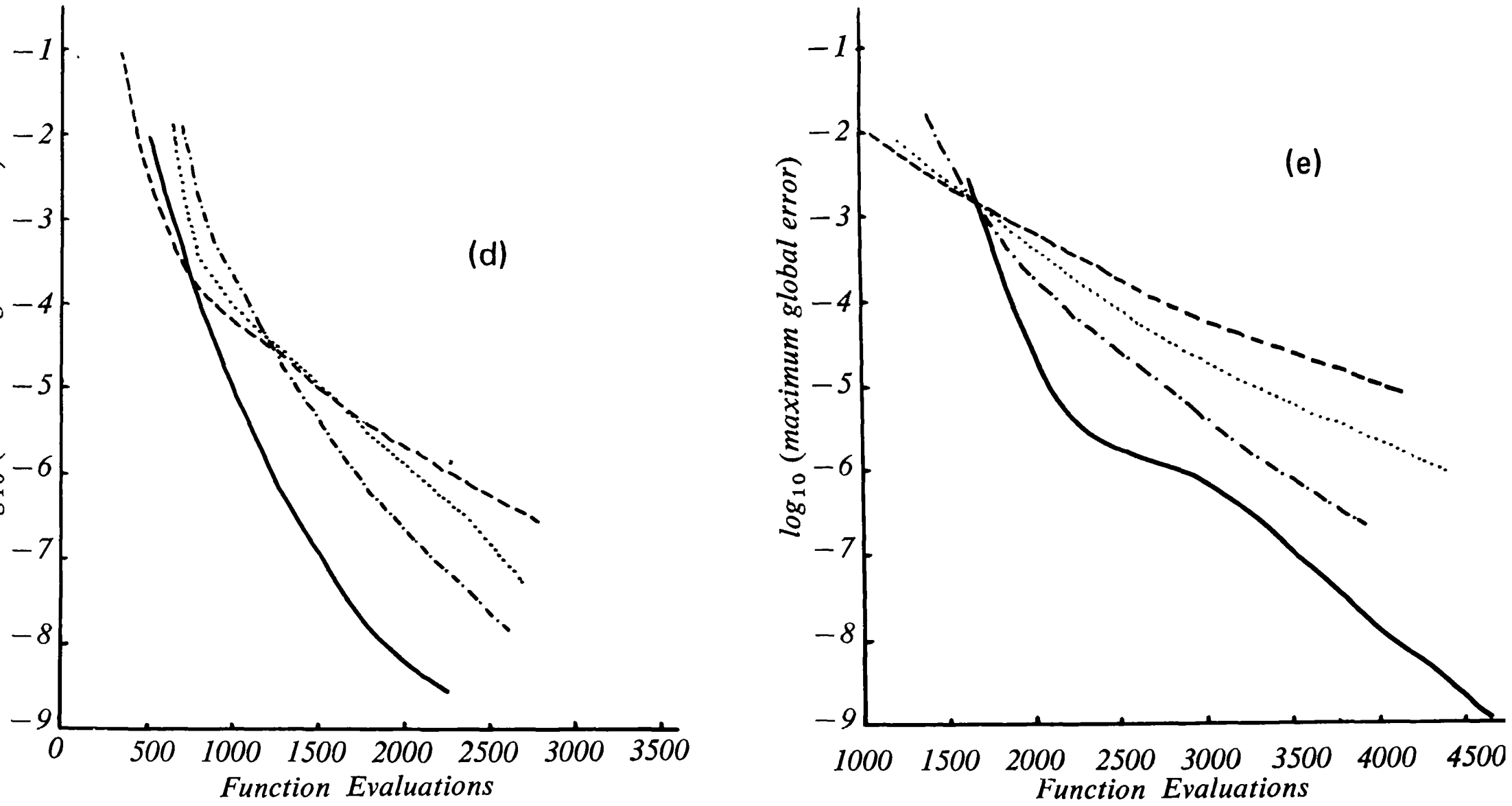


Fig. 2d-e.

It is clear from Figure 1 that the higher order mode gives substantially better accuracy in the case of RKF7. Our results consistent with those of Enright *et al.* (1974) indicate that the order of the optimum formula should be chosen with regard to the required accuracy. For low accuracy, global error $>10^{-3}$ approximately, low order methods are to be preferred over high order methods which can suffer from absolute instability due to the very large step sizes encountered when using high tolerances. These comments apply equally to RK and RKN formulae.

An examination of Figure 1 indicates that RKF7 in eighth order mode should be preferred to the others for global error less than 10^{-3} approximately otherwise either RKF4 (in fifth order mode) or RK5(4)*T* should be used. The intermediate formula RKF5 does not seem to offer advantages for any accuracy. A comparison of Figures 1 and 2 indicates that RKN methods should be preferred wherever possible. This conclusion was reach by Fehlberg (1972) with regard to a different problem. The RKN5(4) of Bettis is most efficient for low accuracy and though it is far inferior to the RKN7(6)*T* for high accuracy it is as efficient as the RKF7 even for an accuracy of 10^{-5} on problem 4. For high accuracy the RKN7(6)*T* is easily the best formula.

4. Discussion

We have presented only results for the integration of the two-body problem but it will be clear that these are applicable to many-body problems; the different cases

(1 to 5) give an indication of numerical behaviour in the simulation of interplay in N-body systems.

In view of the rather small global error range in which low order methods are useful, it may seem reasonable to disregard these formulae in celestial mechanical calculations. However we feel that they should be used in some cases when only a few integration steps are required. The resisting medium calculations of Dormand and Woolfson (1977) use a combination of a fourth order method and RKF7. In this, and many other, cosmogonical computations, the initial conditions are somewhat speculative and therefore the application of low tolerances is unreasonable.

It will be seen that in cases 4 and 5, especially the latter, the global error for a given number of function evaluations is considerably larger than in the other cases, thus illustrating the need for regularization (Bettis and Szebehely, 1971) during close encounters in the numerical analysis of the N-body problem.

The computations described above were performed in FORTRAN using 11 or 20 significant figure precision on the ICL 1905E of the Teesside Polytechnic computer unit. We are grateful to a referee for his valuable comments on an earlier version of this paper.

References

Bettis, D. G.: 1973, *Celest. Mech.* **8**, 229–233.
Bettis, D. G. and Szebehely, V.: 1971, *Astrophys. Space Sc.* **14**, 133–150.
Butcher, J. C.: 1964, *J. Austral. Math Soc.* **4**, 179–194.
Dormand, J. R. and Woolfson, M. M.: 1977, *Monthly Notices Roy. Astron. Soc.* **180**, 243–279.
Enright, W. H., Bedet, R., Farkas, I., and Hull, T. E.: 1974, Technical Report No. 68, Dept of Computer Science, University of Toronto.
Fehlberg, E.: 1968, NASA TR R287.
Fehlberg, E.: 1969, NASA TR R315.
Fehlberg, E.: 1972, NASA TR R381.
Hull, T. E., Enright, W. H., Fellen, B. M., and Sedgwick, A. E.: 1972, *SIAM J. Num. An.* **9**, 603–637.
Stetter, H. J.: 1971, *SIAM J. Num. An.* **8**, 512–523.
Zonneveld, J. A.: 1970, *Automatic Numerical Integration*, 2nd edition, Mathematisch Centrum, Amsterdam.