

Error estimate of a fourth-order Runge-Kutta method with only one initial derivative evaluation

by A. S. CHAI

Hybrid Computer Laboratory
University of Wisconsin
Madison, Wisconsin

INTRODUCTION

In the numerical solution of differential equations it is desirable to have estimates of the local discretization (or truncation) errors of solutions at each step. The estimate may be used not only to provide some idea of the errors, but also to indicate when to adjust the step size. If the magnitude of the estimate is greater than the preassigned upper bound, the step size is reduced to achieve smaller local errors. If the magnitude of the estimate is less than the preassigned lower bound, the step size is increased to save the computing time.

The 4th-order Runge-Kutta method has the advantage that it provides an easy way to change the step size, but it does not provide as simple a way to get error estimates as does Milne's predictor-corrector method.¹ Several methods^{2,3,4,5} for achieving error estimates have been derived and are briefly as follows:

1) One-step method

The one-step method provides all the information for the error estimate in one step. The important one-step method is Sarafyan's pseudo-iterative formula² which is a 5th-order Runge-Kutta formula imbedded in a 4th-order Runge-Kutta formula as follows:

$$\begin{aligned} k_0 &= hf(x_n, y_n) \\ k_1 &= hf(x_n + h/2, y_n + k_0/2) \\ k_2 &= hf(x_n + h/2, y_n + (k_0 + k_1)/4) \\ k_3 &= hf(x_n + h, y_n - k_1 + 2k_2) \\ k_4 &= hf(x_n + 2h/3, y_n + (7k_0 + 10k_1 + k_3)/27) \\ k_5 &= hf(x_n + 2h/10, y_n + (28k_0 - 125k_1 + 546k_2 \\ &\quad + 54k_3 - 378k_4)/625) \end{aligned}$$

The 4th-order formula is

$$y_{n+1} = y_n + (k_0 + 4k_2 + k_3)/6$$

and the 5th-order formula is

$$\bar{y}_{n+1} = y_n + (14k_0 + 35k_3 + 162k_4 + 125k_5)/336$$

The estimate is

$$E_{n+1} = y_{n+1} - \bar{y}_{n+1}$$

The work of Luther and Konen⁶ (Legendre-Gauss) and of Luther⁷ (Newton-Cotes, the 2nd formula, and Lobatto) also yield suitable pseudo-iterative formulas.

Really, pseudo-iterative formulas are 5th-order Runge-Kutta integration schemes used to estimate the error of the 4th-order Runge-Kutta integration.

The pseudo-iterative formula can be used to estimate the error at the first step, which can be done by no other method. But it requires about 50% more computing time for the additional derivative evaluations.

Merson's and Scraton's methods with five derivative evaluations belong to one-step pseudo-iterative method, but they are only applicable in particular cases.^{4,8}

2) Two-step method^{4,5}

This method requires the computation of y_{n+1} and y_{n+2} with a step size h , and then the recomputation of y_{n+2}^* with a doubling of the step size. The error estimate is

$$E_{n+2} = (y_{n+2}^* - y_{n+2})/30. \quad (1.1)$$

Since the error is of the order h^5 , we can let

$$y_{n+2} = \bar{y}_{n+2} + \epsilon_{n+1} + \epsilon_{n+2}$$

where the two error terms relate to the two steps, and

$$y_{n+2}^* = \bar{y}_{n+2} + 32\epsilon_{n+1} + O(h^6)$$

Hence, Formula (1.1) really provides an estimate of the error,

$$\epsilon_{n+29/30} = (31\epsilon_{n+1} - \epsilon_{n+2})/30 + O(h^6)$$

which is close to the error, ϵ_{n+1} , in y_{n+1} . If the estimate is for the error, ϵ_{n+2} , in y_{n+2} , it requires that the errors,

ϵ_{n+1} and ϵ_{n+2} , in y_{n+1} and y_{n+2} be approximately equal.

This method can be used to estimate the error at each two steps starting at y_2 and requires about 37.5% more computing time than fourth-order Runge-Kutta integration without estimates.

3) Multi-step method^{3,4}

Ceschino and Kuntzmann (Ref. 3, pp. 305-310) collected several multi-step formulas for the error estimate which was based on Refs. 9, 10. One of them is (also in Ref. 4);

$$E_{n+2} = [10\Delta y_{n-1} + 19\Delta y_n + \Delta y_{n+1} - h(3f_{n-1} + 18f_n + 9f_{n+1})]/30 \quad (1.2)$$

Strictly speaking, (1.2) estimates the error

$$\epsilon_{n+7/10} = (10\epsilon_n + 19\epsilon_{n+1} + \epsilon_{n+2})/30.$$

This method can estimate the error at each step. But this method cannot be used until the completion of the third step (i.e., y_3). The estimate is close to the error at the last step, because

$$\epsilon_{n+7/10} \cong \epsilon_{n+1}.$$

If the estimate is for the error, ϵ_{n+1} , at the last step, and if the estimate causes the step size to change, four additional derivative evaluations are needed. If the estimate is for the error, ϵ_{n+2} , at the present step, it requires that the errors at three successive steps, ϵ_n , ϵ_{n+1} , and ϵ_{n+2} , be approximately equal. This requirement may not be satisfied if the errors change rapidly.

In general, the derivative evaluations need most of the computing time. The 4th-order Runge-Kutta method already requires more derivative evaluations than the other methods, e.g., Milne's method;¹ hence the extra time for the additional derivative evaluations for the error estimate is too expensive, and should be avoided as much as possible.

The suggested method

Ceschino and Kuntzmann (Ref. 3, p. 308) showed the following formula

$$E_{n+2} = \frac{11\Delta y_{n+1}}{30} + \frac{19\Delta y_n}{30} - h \left(\frac{1f_{n+2} + 19f_{n+1}}{9} + \frac{8f_n}{30} - \frac{1f_{n-1}}{90} \right) \quad n > 0 \quad (2.1)$$

for estimating the local discretization error in y_{n+2} in each step in a fourth-order Runge-Kutta integration. The author has found that (2.1) has advantages over the other methods in Section 1 and shows this below. Also, (2.1) can be extended to $n = 0$, because we can form

$$y_{-1} = y_0 + 10\Delta y_1 + 19\Delta y_0 - 3h(f_2 + 6f_1 + 3f_0) \quad (2.2)$$

which has an error of $O(h^6)$, and then evaluate

$$f_{-1} = f(x_{-1}, y_{-1}) \quad (2.3)$$

Then, using (2.1) for computing the estimate E_2 , Equations (2.2) and (2.3) can be employed when (2.1) is just started or when the step size changes.

Equation (2.1) requires f_{n+2} , which has to be computed for k_0 in the next step if the step size does not change; hence, no additional derivative evaluations in each step are needed except for E_2 where one additional evaluation for f_{-1} by (2.3) is needed. This method requires about 12.5% more computing time for evaluating f_{-1} , when f_{-1} is not available, but no additional time if $n > 0$. Hence, this method has an advantage in computing time over the one-step and two-step methods.

Equations (2.1-3) can estimate the error at each step after the first. Equation (2.1) estimates the error

$$\epsilon_{n+41/30} = (11\epsilon_{n+2} + 19\epsilon_{n+1})/30$$

which is closer to the error, ϵ_{n+2} , than the estimates in the two-step and multi-step methods. Hence, this method has another advantage over the two-step and multi-step methods. The departure of the estimate from the local discretization error in y_{n+2} will be shown in the next section.

The derivation of (2.1) was shown in Ref. 3 and, as well as the derivation of (2.2), is briefly shown in Appendix 1.

Equation (2.1) can be employed to be a corrector to get errors in y of the order h^6 . The convergence theorem and the experiment result are shown in Appendix 2.

The departure of the estimate

The departure of the estimate from the local discretization error is (Ref 3, pp. 306-308):

$$E_{n+2} - \epsilon_{n+2} = -\frac{1}{2}hf_{y_{n+1}}\epsilon_{n+1} - \frac{19h}{30}\epsilon'_{n+1} - \frac{11}{5400}h^6f''_{n+1} + O(h^7) \quad (3.1)$$

To give the reader some picture of the departure, let us consider a differential equation:

$$y' = ay, y_0 = y(x_0)$$

Where a is a constant and is not equal to zero.

The formula of the local discretization error of the order h^5 in y_{n+2} by several known 4th-order Runge-Kutta formulas can be found in (Ref. 3, p. 81). In this example, the local discretization error of the order h^5 is

$$\begin{aligned} \epsilon_{n+2} &= -\frac{h^5}{120} \left(f_{y_{n+2}} \right)^3 f'_{n+2} \\ &= -\frac{h^5}{120} a^3 y''_{n+2}. \end{aligned}$$

Since $y = y_0 e^{ax}$,
this gives

$$\epsilon_{n+1} = -\frac{h^5}{120} y_0 a^5 e^{ax_{n+1}},$$

$$\epsilon'_{n+1} = -\frac{h^5}{120} y_0 a^6 e^{ax_{n+1}},$$

$$f_{n+1}^v = y_0 a^6 e^{ax_{n+1}},$$

and

$$f_y = a.$$

Hence the departure of the estimate is

$$E_{n+2} - \epsilon_{n+2} = \frac{1}{135} h^6 a^6 y_0 e^{ax_{n+1}}.$$

The relative error is

$$\frac{E_{n+2} - \epsilon_{n+2}}{\epsilon_{n+2}} = \frac{8ha}{9} = .889ha$$

of which the magnitude is less than 10% if $|ha| \leq 0.1$.

To verify the theory an experiment was run to solve

$$y' = y, y_0 = 1, h = 0.1 \text{ and } x = 0 \text{ to } 10.$$

The local discretization error at x_{n+2} is $y_{n+2} - y_{n+1}e^h$. The experimental results for the relative errors of the estimates for $x = .2$ to 10, were between .08211 and .08225. Then the equation was changed to

$$y' = -y$$

with the same parameters. The local discretization error at x_{n+2} is $y_{n+2} - y_{n+1}e^{-h}$. The experimental results for the relative errors were between -.09620 and -.09637, or slightly less than 10%. Hence the experimental result agrees with the theory.

If Equations (2.2) and (2.3) are employed for computing f_{-1} , the departure, which is derived in Appendix 3, is

$$E_2 - \epsilon_2 = -\frac{1}{6} h f_y \epsilon_1 - \frac{19}{30} h \epsilon_1' - \frac{11}{5400} h^6 f_y + O(h^7)$$

As before, the relative error for E_2 in

$$y' = ay, a \neq 0$$

is approximately equal to

$$\frac{5}{9} h a = .556 h a$$

which is about 5% if $|ha| \approx 0.1$.

The experimental results for E_2 in

$$y' = y \text{ and in } y' = -y,$$

are .0518 and -.0593 respectively, where $h = 0.1$.

Another experiment is to solve a system of non-linear differential equations

$$\begin{aligned} y'(x) &= z(x) \\ z'(x) &= (2y(x) - 1)z(x) \\ y(0) &= 0.5, z(0) = y'(0) = -0.25 \end{aligned} \quad (3.2)$$

The step size, h , was 0.1 and x ran from 0 to 5.

The local discretization error at y_{n+1} is

$$\epsilon_{n+2} = y_{n+2} - \frac{\alpha(y_{n+1} - \beta) - \beta(y_{n+1} - \alpha)e^{(\alpha-\beta)h}}{y_{n+1} - \beta - (y_{n+1} - \alpha)e^{(\alpha-\beta)h}}$$

where

$$\frac{\alpha}{\beta} = \frac{1 \pm \sqrt{1 + 4(y_{n+1}^2 - y_{n+1} - z_{n+1})}}{2}$$

The local discretization error ϵ_{n+2} and the relative error of the estimate are shown in Table 3.1. The values of the relative errors are about 0.1 in general except when the curve of ϵ_{n+2} approaches zero rapidly. Hence the estimate is in general suitable for practical purpose.

TABLE 3.1

x	Local error	Rel. error
.2	.7558E-08	-.6193E-02
.3	.7047E-08	-.4302E-02
.4	.6301E-08	-.1157E-01
.5	.5371E-08	-.2127E-01
.6	.4320E-08	-.3440E-01
.7	.3223E-08	-.5614E-01
.8	.2145E-08	-.9597E-01
.9	.1144E-08	-.1944E 00
1.0	.2710E-09	-.8569E 00
1.1	-.4447E-09	.5242E 00
1.2	-.9877E-09	.2258E 00
1.3	-.1352E-08	.1517E 00
1.4	-.1544E-08	.1189E 00
1.5	-.1586E-08	.9782E-01
1.6	-.1502E-08	.8015E-01
1.7	-.1317E-08	.6420E-01
1.8	-.1059E-08	.4473E-01
1.9	-.7522E-09	.1596E-01
2.0	-.4229E-09	-.5341E-01
2.1	-.9004E-10	-.6064E 00
2.2	.2301E-09	.3657E 00
2.3	.5275E-09	.2065E 00
2.4	.7931E-09	.1637E 00
2.5	.1021E-08	.1448E 00
2.6	.1211E-08	.1329E 00
2.7	.1362E-08	.1254E 00
2.8	.1477E-08	.1195E 00
2.9	.1556E-08	.1157E 00
3.0	.1605E-08	.1126E 00
3.1	.1626E-08	.1103E 00
3.2	.1624E-08	.1083E 00
3.3	.1603E-08	.1068E 00
3.4	.1565E-08	.1054E 00
3.5	.1516E-08	.1043E 00
3.6	.1457E-08	.1031E 00
3.7	.1391E-08	.1025E 00
3.8	.1320E-08	.1018E 00
3.9	.1247E-08	.1012E 00
4.0	.1172E-08	.1005E 00
4.1	.1087E-08	.1005E 00
4.2	.1024E-08	.9993E-01
4.3	.9525E-09	.9958E-01
4.4	.8836E-09	.9921E-01

4.5	.8176E-09	.9892E-01
4.6	.7548E-09	.9876E-01
4.7	.6954E-09	.9867E-01
4.8	.6395E-09	.9861E-01
4.9	.5872E-09	.9860E-01
5.0	.5385E-09	.9827E-01

Local discretization error and relative error of estimate at y in $y' = z$ and $z' = (2y - 1)z$

ACKNOWLEDGMENTS

The author is indebted to Professors C.A. Ranous and V. C. Rideout for aid in expression in writing this paper and Professor C. W. Cryer for discussion of convergence theory. The author particularly thanks his former advisor Professor H. J. Wertz who suggested this problem and has often given encouragement.

The author is also indebted to the referees for their constructive comments.

All the experimental results were obtained on an SDS 930 computer (38 bit = 11.4 digits in mantissa) in the Hybrid Computer Laboratory at the University of Wisconsin, Madison, Wisconsin.

Appendix 1

Derivation of (2.1) [3, pp 305-308] and (2.2)

It is easy to derive

$$y(x_{n+2}) = -\frac{8}{11}y(x_{n+1}) + \frac{19}{11}y(x_n) + h\left(\frac{10}{33}\bar{f}_{n+2}\right) + \frac{19}{11}\bar{f}_{n+1} + \frac{8}{11}\bar{f}_n - \frac{1}{33}\bar{f}_{n-1} - \frac{1}{180}h^6\bar{f}_{n+1}^{(6)} + O(h^7) \quad (A1.1)$$

Substitute

$$y(x_n) = y(x_{n+1}) - \Delta y(x_{n+1}) - \epsilon_{n+2}$$

and

$$y(x_{n+2}) = y(x_{n+1}) + \Delta y(x_{n+1}) - \epsilon_{n+2}$$

then, after simplification, we get

$$\begin{aligned} \epsilon_{n+41/30} &= \frac{11\epsilon_{n+2} + 19\epsilon_{n+1}}{30} + O(h^7) \\ &= \frac{11}{30}\Delta y(x_{n+1}) + \frac{19}{30}\Delta y(x_n) - h\left(\frac{1}{9}\bar{f}_{n+2}\right) + \frac{19}{30}\bar{f}_{n+1} \\ &\quad + \frac{8}{30}\bar{f}_n - \frac{1}{90}\bar{f}_{n-1} - \frac{11}{5400}h^6\bar{f}_{n+1}^{(6)} + O(h^7) \end{aligned}$$

Now, therefore, the estimate is (2.1).

Equation (2.2) is easy to be derived by

$$y_{-1} = 10y_2 + 9y_1 - 18y_0 - 3h(f_2 + 6f_1 + 3f_0) + O(h^6)$$

Appendix 2

A convergence theorem of (2.1)

Equation (A1.1) can be rewritten to

$$\begin{aligned} y_{n+1} &= -\frac{19}{11}\bar{y}_{n+1} + \frac{30}{11}\bar{y}_n + \frac{19}{11}\Delta y_n - \frac{19}{11}\Delta y_{n-1} \\ &\quad + h\left(\frac{10}{33}\bar{f}_{n+1} + \frac{19}{11}\bar{f}_n + \frac{8}{11}\bar{f}_{n-1} - \frac{1}{33}\bar{f}_{n-2}\right) \end{aligned}$$

Then

$$\begin{aligned} \bar{y}_{n+1} &= \bar{y}_n + \frac{19}{30}(\Delta y_n - \Delta y_{n-1}) + h\left(\frac{1}{9}\bar{f}_{n+1} \right. \\ &\quad \left. + \frac{19}{30}\bar{f}_n + \frac{8}{30}\bar{f}_{n-1} - \frac{1}{90}\bar{f}_{n-2}\right) = y_n + \theta(x_n, y_n) \end{aligned}$$

The following statement of a convergence theorem is similar to Byrne and Lambert¹¹.

Assume that (1) $f(x, y)$ is continuous for $x \in I$ and $\|y\| < \infty$, (2) f satisfies a Lipschitz condition and (3) $0(x_n, y_n)$ is defined for all h such that $x + h$ and $x - 2h \in I$.

(A2.1) is consistent because

$$\lim_{h \rightarrow 0} (x, y)/h = f(x, y), \quad x \in I, \quad \|y\| < \infty.$$

A convergence theorem is

Let y_0 be the initial value and

$$\|y_1 - y(x_1)\| + \|y_2 - y(x_2)\| \leq hL$$

where L is a non-negative constant and let there exist three non-negative numbers L_1 , L_2 , and L_3 such that

$$\begin{aligned} \|0(x_n, w_n^*) - 0(x_n, w_n)\| &\leq h(L_1 \|w_n^* - w_n\| + L_2 \|w_{n-1}^* \\ &\quad - w_{n-1}\| + L_3 \|w_{n-2}^* - w_{n-2}\|) \end{aligned}$$

hold for the vectors w_n^* and w_n . Under these conditions, the hypotheses (1-3), and the consistence property, (A2.1) is convergent.

The proof can be similarly employed as in Byrne and Lambert¹¹ except that one more term,

$$h L_3 z_{n-2} = h L_3 \|y(x_{n-2}) - y_{n-2}\|$$

is added in the right-hand side of the inequality

$$z_{n+1} \leq z_n(1 + hL_1) + hz_{n-1}L_2 + hg(h)$$

and thereafter. This proof does not include the point x_{-1} , so we have to assume that $x_{-1} \in I$ if (2.2) is employed.

Hence (A2.1) can be used as a corrector to get errors to $O(h^6)$ except at y_1 . However y_2 should be corrected by twice the estimate at x_2 .

Table A2.1 shows the accumulative errors of non-corrected and corrected solutions in the experiment to solve the system of non-linear equations (3.2). The theoretical solution is

$$y(x) = \frac{1}{1 + e^x}$$

The accumulative error was obtained by subtracting $y(x_n)$ from y_n (non-corrected) or y_n (corrected by (2.1)).

In Table A2.1 the errors of the non-corrected solutions are greater than those of the corrected.

x	Non-corrected	Corrected
.2	1.531×10^{-8}	2.929×10^{-10}
.3	2.228×10^{-8}	5.402×10^{-10}
.5	3.364×10^{-8}	1.759×10^{-9}
1.0	4.327×10^{-8}	6.483×10^{-9}
2.0	2.837×10^{-8}	6.314×10^{-9}
3.0	2.555×10^{-8}	-5.898×10^{-10}
4.0	2.104×10^{-8}	-2.966×10^{-9}
5.0	1.352×10^{-8}	-2.749×10^{-9}

TABLE A2.1
Errors in non-corrected and corrected
solutions of y in Equation (3.2).

Appendix 3

The departure of (2.1) when (2.3) is employed

If (2.2) is employed for y_{-1} , since y_0 is the initial value, we can assume that y_0 has no error, then

$$y_1 = y(x_1) + \epsilon_1$$

and

$$y_2 = y(x_2) + \epsilon_1 + \epsilon_2$$

Hence

$$f_i = f_1 + f_{y_1}(\epsilon_1 + \dots + \epsilon_i), i = 1, 2$$

Now

$$\Delta y_0 = \Delta y(x_0)$$

and

$$\Delta y_1 = \Delta y(x_1) + hf_{y_1}\epsilon_1$$

with

$$y_{-1} = y(x_{-1}) + 19\epsilon_1 + 10\epsilon_2$$

and

$$f_{-1} = f_1 + f_{y_{-1}}(19\epsilon_1 + 10\epsilon_2)$$

then

$$E_2 \cong \frac{11}{30}\epsilon_2 + \frac{19}{30}\epsilon_1 - \frac{1}{6}hf_{y_1}\epsilon_1 - \frac{11}{5400}h^6f_1$$

Hence the departure of the estimate E_2 from ϵ_2 is

$$E_2 - \epsilon_2 \cong -\frac{19}{30}h\epsilon_1' - \frac{1}{6}hf_{y_1}\epsilon_1 - \frac{11}{5400}h^6f_1$$

NOMENCLATURE

- $y' = f(x, y)$ - a system of differential equations where x is the independent variable and y represents the dependent variables. y, y' and f are vectors.
- h - step size.

- y_n - solution of y at $x_n = x_0 + nh$ with an error of $O(h^5)$ obtained by a fourth-order Runge-Kutta formula.
- y_n - solution of y at x_n with an error of $O(h^6)$.
- $f = f(x_n, y_n)$
- $f_n = f(x_n, y_n)$
- $y(x_n)$ - theoretical solution of y at x_n .
- Δy_n - increment function which is defined by $y_{n+1} - y_n$.
- $\Delta y(x_n)$ - similar to Δy_n , but replacing y_n in Δy_n by $y(x_n)$.
- local discretization (or truncation) error in y_n which is defined by the difference of $y(x_n)$ from y_n with considering y_{n-1} to be the initial value, e.g.
- $\epsilon_n = y(x_{n-1}) + \Delta y(x_{n-1}) - Y(x_n)$
- E_n - estimate of ϵ_n .

REFERENCES

- 1 W E MILNE
Numerical solution of differential equations
Wiley New York p 66 (1953)
- 2 D SARAFYAN
Error estimation for Runge-Kutta methods through pseudo-iterative formulas
Technical Report No 14 Louisiana State University New Orleans (May 1966)
- 3 F CESCHINO J KUNTZMANN
Problemes differentiels de conditions initiales
Dunod Paris France pp 81 305-310 (1963) English translation by D Boyanovitch Prentice-Hall Englewood Cliffs N J pp 67 247-251 (1966)
- 4 R E SCRATON
Estimation of the truncation error in Runge-Kutta and allied processes
Comp J vol 7 pp 246-248 April 1964
- 5 L COLLATZ
The numerical treatment of differential equations
Third edition second printing English translation by P G Williams Springer-Verlag New York pp 51-52 1966
- 6 H A LUTHER H P KONEN
Some fifth-order classical Runge-Kutta formulas
SIAM Review vol 7 pp 551-558 1965
- 7 H A LUTHER
Further explicit fifth-order Runge-Kutta formulas
SIAM Review vol 8 pp 374-380 1966
- 8 A R CURTIS
Correspondence to Estimation of the truncation error in Runge-Kutta and allied processes
Comp J vol 8 p 52 1965
- 9 H MOREL
Evaluation de l'erreur sur un pas dans la methode de Runge-Kutta
Comptes rendus de l'Academie des Sciences Paris vol 243 pp 1999-2002 1965
- 10 J KUNTZMANN
Evaluation de l'erreur sur un pas dans les methodes a pas separees
Chiffres vol 2 pp 97-102 1959
- 11 G D BYRNE R J LAMBERT
Pseudo-Runge-Kutta methods involving two points
J ACM vol 13 pp 114-123 1966

