CHAPTER 4

A FOURTH ORDER EMBEDDED RUNGE-KUTTA RKACeM(4,4) METHOD BASED ON ARITHMETIC AND CENTROIDAL MEANS WITH ERROR CONTROL

The Runge-Kutta method is too laborious for tabulating many steps of a numerical solution unless a computing machine is used.

- James B. Scarborough

4.1 INTRODUCTION

Many Runge-Kutta (RK) codes for the numerical solution of initial value problems in ODEs are based on embedded pairs of RK formulas. Butcher [17] introduced a modern approach with the special method RK (4,5). The subroutine DVERK produced by Hull et.al. [77] is based on a pair of formulae of orders 5 and 6 due to Verner. Bogacki and Shampine [13] derived a RK formula of orders 4 and 5. Thus embedded methods are actually two methods built into one. The first method is of order p and the second has order p+1. The difference between these methods provides an error estimate for the first method with order p. Error estimates by these methods have been explained in Henrici [71] and Lambert [83]. Evans and Yaakub [56, 149] introduced a new embedded Runge-Kutta RK(4,4) method which is actually two different RK methods but of the same order p = 4. This embedded method has been developed using Runge-Kutta methods based on arithmetic mean (RKAM) and Contraharmonic Mean (RKCoM).
In the third chapter of this thesis, it has been demonstrated that RK methods based on AM and CeM suit well for the system of IVPs. Hence, in this chapter, similar to RK(4,4), a new combination based on RKAM and RKCeM, both of which are of order $p = 4$, is formed and written as RKACeM(4,4). In order to study the effectiveness of the new method RKACeM (4, 4), we have taken ten problems based on linear as well as non-linear IVPs and the results are compared with Runge-Kutta Fehlberg Method RKF(4,5), Runge-Kutta Merson method (RK Merson) and RK(4,4) method developed by Yaakub and Evans [56, 149].

4.2 THE EMBEDDED RKACeM (4, 4) METHOD

Consider the initial value problem in ordinary differential equation

$$y'(x) = f(x, y(x)), x \geq x_0$$

$$y(x_0) = y_0$$

where $f: \mathbb{R} \times \mathbb{R}^n$ is assumed to be sufficiently differentiable in a neighborhood of the exact solution $(x, y(x)), x \in [a, b]$.

For the effective treatment of (4.1) an embedded explicit Runge-Kutta (RK) pair may be applied. A general $s$ – stage RK pair is characterized by the extended Butcher tableau of parameter

$$
\begin{array}{c|c}
C & A \\
\hline
b^T & \tilde{b}^T \\
\hline
E^T & \\
\end{array}
$$

(4.2)

where $b^T, \tilde{b}^T, C \in \mathbb{R}^s$ and $A \in \mathbb{R}^s \times \mathbb{R}^s$ is for explicit pairs, strictly lower triangular. A procedure utilizing such a pair, advances the integration from $(x_n, y_n)$ to $x_{n+1} = x_n + h$, computing at each step, two approximations $y_{n+1}$ and $\hat{y}_{n+1}$ of orders $p$ and $q$ to determine $y(x_{n+1})$ are given by
The fourth order classical Runge-Kutta methods (RKAM) can be written in the Butcher array form as

\[
\begin{array}{c|cccc}
0 & 1 & 2 & 2 & 2 \\
1 & 0 & 1 & 2 & 3 \\
2 & 0 & 0 & 1 & 3 \\
3 & 0 & 0 & 1 & 3 \\
\end{array}
\]

\[
y_{n+1} = y_n + \frac{h}{3} \left[ k_1 + k_2 + k_3 + k_4 \right]
\]

with the vectors of column \( C, b, \) and the \((s \times s)\) matrix \( A \) (aij).

From the embedded form we can obtain the local truncation error (LTE) in the \( p^{th} \) order formula, i.e., LTE = \( y_{n+1} + \hat{y}_{n+1} \) which may be used to control the step size \( h \).
\begin{equation}
\begin{array}{l}
y_{n+1} = y_n + \frac{h}{3} \left[ \sum_{i=1}^{3} \frac{k_i + k_{i+1}}{2} \right] \\
\end{array}
\tag{4.4}
\end{equation}

where
\begin{align*}
k_1 &= f(x_n, y_n) \\
k_2 &= f(x_n + \frac{h}{2}, y_n + \frac{hk_1}{2}) \\
k_3 &= f(x_n + \frac{h}{2}, y_n + \frac{hk_2}{2}) \\
k_4 &= f(x_n + h, y_n + hk_3) \\
\end{align*}

RKAM with Butcher array can be written in the new form as
\begin{center}
\begin{tabular}{c|cccc}
0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 \\
2 & 0 & 1 & 2 & 1 \\
2 & 0 & 1 & 2 & 1 \\
1 & 0 & 0 & 1 & 0 \\
3 & 0 & 1 & 1 & 1 \\
3 & 0 & 1 & 1 & 1 \\
3 & 0 & 1 & 1 & 1 \\
\end{tabular}
\end{center}

The fourth order Runge-Kutta Centroidal Mean RKCeM is
\begin{equation}
\begin{align*}
y_{n+1} &= y_n + \frac{2h}{9} \left[ \frac{k_1^2 + k_1k_2 + k_2^2}{k_1 + k_2} + \frac{k_2^2 + k_2k_3 + k_3^2}{k_2 + k_3} + \frac{k_3^2 + k_3k_4 + k_4^2}{k_3 + k_4} \right] \\
y_{n+1} &= y_n + \frac{h}{3} \left[ \frac{2}{3} \sum_{i=1}^{3} \left( \frac{k_i^2 + k_i^2 + k_i k_{i+1}}{k_i + k_{i+1}} \right) \right] \\
\end{align*}
\tag{4.5}
\end{equation}

where
\begin{align*}
k_1 &= f(x_n, y_n) \\
k_2 &= f(x_n + \frac{h}{2}, y_n + \frac{hk_1}{2}) \\
k_3 &= f(x_n + \frac{h}{2}, y_n + \frac{1}{24}hk_1 + \frac{11}{24}hk_2) \\
k_4 &= f(x_n + h, y_n + \frac{1}{12}hk_1 - \frac{25}{132}hk_2 + \frac{73}{66}hk_3) \\
\end{align*}
The Butcher array form for RKCeM is

\[
\begin{array}{ccc}
0 & & \\
1/2 & 1/2 & \\
1/2 & 11/24 & 13/24 \\
1/12 & 25/132 & 73/66 \\
1/3 & 1/3 & 1/3 \\
\end{array}
\]

Combination of RKAM and RKCeM (Eqs. (4.4) and (4.5)) is referred as RKACeM (4, 4), and can be formulated as:

\[
k_1 = f(x_n, y_n) = k_1^*
\]

\[
k_2 = f\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_1\right) = k_2^*
\]

\[
k_3 = f\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_2\right)
\]

\[
k_4 = f(x_n + h, y_n + hk_3)
\]

\[
k_3^* = f\left(x_n + \frac{1}{2}h, y_n + \frac{1}{24}hk_1 + \frac{11}{24}hk_2\right)
\]

\[
k_4^* = f\left(x_n + \frac{1}{2}h, y_n + \frac{1}{12}hk_1 - \frac{25}{132}hk_2 + \frac{73}{66}hk_3^*\right)
\]

\[
y_{n+1} = y_n + \frac{h}{3} \left[ \sum_{i=1}^{3} \frac{k_i + k_{i+1}}{2} \right]
\]

\[
y_{n+1}^* = y_n + \frac{h}{3} \left[ \sum_{i=1}^{3} \frac{2k_i^2 + k_{i+1}^2 + k_i^*k_{i+1}^*}{3(\frac{k_i}{k_i^*} + \frac{k_{i+1}}{k_{i+1}^*})} \right]
\]
The Butcher array form for embedded RKACeM(4,4) is

\[
\begin{array}{c|cccc}
0 & 1 & 1 & 2 & 1 \\
1 & 1 & 0 & 1 & 2 \\
2 & 0 & 0 & 1 & 2 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 1 & 11 & 24 & 24 \\
2 & 24 & 24 & \frac{-25}{12} & 73 \\
1 & \frac{-25}{12} & 73 & 66 & 66 \\
\end{array}
\]

Hence by Butcher array (4.2)

\[
b^T = y_n^{AM} = y_n + \frac{h}{3} \left( \sum_{i=1}^{3} \frac{k_i + k_{i+1}}{2} \right)
\]

\[
\hat{b}^T = y_n^{CeM} = y_n + \frac{h}{3} \left[ \frac{2}{3} \sum_{i=1}^{3} \frac{k_i^2 + k_{i+1}^2 + k_i k_{i+1}}{k_i + k_{i+1}} \right]
\]

and the estimation of the local truncation error, \( E^T = |b^T - \hat{b}^T| \) In this RKACeM(4,4) method, four stages are required to obtain the approximate solutions, which share the same set of vectors \( k_1 \) and \( k_2 \) using \( b^T \) and \( \hat{b}^T \), but \( k_3 \) and \( k_4 \) use \( b^T \), while \( k_3^* \) and \( k_4^* \) use \( \hat{b}^T \).

### 4.3 ERROR CONTROL IN RKACeM (4,4)

Error estimate for the four-stage explicit (AM – CeM) method of order four is obtained by implementing the local truncation error for the RKAM and RKCeM methods. Now we will discuss the local truncation error (LTE) and global truncation error (GTE).
Definition 4.1 Local Truncation Error: The local truncation error at the point \( x_{n+1} \) is the difference between the computed value \( y_{n+1} \) and the value at the point \( x_{n+1} \) on the solution curve that goes through the point \((x_n, y_n)\).

Definition 4.2 Global Truncation Error: The global truncation error at the point \( x_{n+1} \) is defined as \( y_{n+1} - y(x_{n+1}) \), where \( y(x) \) denotes the solution of the given initial value problem.

4.3.1 Local Truncation Error for RKACeM(4,4)

From eq. (4.6) we can obtain an estimate LTE for the RKACeM (4,4) as \( LTE = y_{n+1} - y_{n+1}^* \) which may be used to control the step size \( h \). Lotkin [90] and Ralston [112] have provided an error estimate for classical fourth order RK scheme as \( |\psi(x_n, y_n ; h)| \leq \frac{73}{720} ML^4 \) where \( M \) and \( L \) are positive constants.

The local truncation error for classical fourth order RK method (RKAM) is

\[ y_{n+1}^{AM} = y_n + LTE_{AM} \]

and for RK based on centroidal mean (RKCeM) is

\[ y_{n+1}^{CeM} = y_n + LTE_{CeM} \]

where \( y_{n+1}^{AM} \) and \( y_{n+1}^{CeM} \) are the numerical approximations at \( x_{n+1} \) obtained by AM and CeM respectively, and \( LTE_{AM} \) and \( LTE_{CeM} \) are the local truncation errors in RKAM and RKCeM.

The difference between the RKAM and RKCeM gives an error estimate for the numerical approximation at \( x_{n+1} \) by

\[ y_{n+1}^{AM} - y_{n+1}^{CeM} = LTE_{AM} - LTE_{CeM} \]

The local truncation error involves the \( y \) derivatives of RKAM and is given by

\[
LTE_{AM} = \frac{h^5}{2880} \left[ -24 f f_y f_y + f^4 f_{yyy} + 2 f^3 f_y f_{yyy} - 6 f^3 f_y^2 f_{yy} + 36 f^2 f_y^2 f_{yy} \right] \quad (4.7)
\]
While the local truncation error of the RKCeM is given by

\[
\text{LTE}_{\text{CeM}} = \frac{h^5}{69120} \left[ -762 f_{x}^4 + 8f_y^4 + 36 f^3 f_{x} f_{yy} - 744 f^3 f_{yy}^2 + 273 f^2 f_{y}^2 f_{yy} \right]
\]  

(4.8)

The absolute difference between LTE_{AM} and LTE_{CeM} is given by

\[
|\text{LTE}_{\text{AM}} - \text{LTE}_{\text{CeM}}| = \frac{h^5}{69120} \left[ 186 f f_{x}^4 + 16f_y^4 f_{yy} 
+ 12 f^3 f_x f_{yy} + 600 f^3 f_{yy}^2 + 591 f^2 f_{y}^2 f_{yy} \right]
\]  

(4.9)

By following an argument suggested by Lotkin [90], if we assume that the following bounds for \( f \) and its partial derivatives, hold for \( x \in [a, b] \) and \( y \in (-\infty, \infty) \), we have

\[
|f(x, y)| < Q, \quad \left| \frac{\partial^{i+j} f(x, y)}{\partial x^i \partial y^j} \right| < \frac{P^{i+j}}{Q^{j-1}}, \quad i + j \leq p
\]  

(4.10)

where \( P \) and \( Q \) are positive constants and \( p \) is the order of the method.

In this case, we have \( p = 4 \). Hence using (4.10), we obtain

\[
\left\{ \begin{array}{l}
|f_x| < P \\
|f_y| < 2PQ \\
|f_f| < \frac{Q^4P^4}{Q^2} \\
|f^4 f_{yy}| < \frac{Q^4P^4}{Q^2} \\
|f^3 f_x f_{yy}| < Q^3 \frac{P^3}{Q^2} \\
|f^3 f_{yy}^2| < Q^3 \left( \frac{P^2}{Q} \right)^2 \\
|f^2 f_y^2 f_{yy}| < Q^2 \frac{P^2}{Q}
\end{array} \right\} < P^4Q
\]

From eqs. (4.9) and (4.10) we obtain

\[
|\text{LTE}_{\text{AM}} - \text{LTE}_{\text{CeM}}| \leq \frac{281}{13824} P^4 Q h^5
\]

Hence, \( \left| y_{AM}^{n+1} - y_{CeM}^{n+1} \right| \leq \frac{281}{13824} P^4 Q h^5 \)  

(4.11)
Suppose that the tolerance $TOL = 0.00001$, then by setting $|y_{n+1}^A - y_{n+1}^C| \leq TOL$

The error control and step size selection can be determined by eq. (4.11) to give the formula

$$\frac{281}{13824} p^4 q h^5 < TOL \quad \text{or} \quad h < \left[\frac{49.2 \times TOL}{p^4 q} \right]^{1/5} \quad (4.12)$$

**4.3.2 Global Truncation Error for RKACE (4,4)**

The LTE for the RKAM and RKCeM are given in eqs. (4.7) and (4.8) respectively and both are of order 5.

**GTE for RKAM**

The Taylor expansion series up to $h^5$ at $x = x_n$ can be written as

$$y(x_n + h) = y(x_n) + hf + \frac{h^2}{2} f_y(x_n, y_n) + \frac{h^3}{6} f_x^2(x_n, y_n) + \frac{h^4}{27} f_y^3(x_n, y_n) + \frac{h^5}{120} f_y^4(x_n, y_n)$$

Subtract Eq. (4.13) from Eq. (4.4) to obtain

$$\epsilon_{n+1} = \epsilon_n + h \left[ f(x_n, y_n) - f(x_n, y(x_n)) \right]$$

$$+ \frac{h^2}{2} \left[ f_y(x_n, y_n) - f_y(x_n, y(x_n)) \right]$$

$$+ \frac{h^3}{6} \left[ f_x^2(x_n, y_n) - f_x^2(x_n, y(x_n)) \right]$$

$$+ \frac{h^4}{24} \left[ f_y^4(x_n, y_n) - f_y^4(x_n, y(x_n)) \right] + \frac{h^5}{2880} \left( \xi \right)$$

$$\leq \epsilon_n + hL \epsilon_n + \frac{h^2}{2} L \epsilon_n + \frac{h^3}{6} L \epsilon_n + \frac{h^4}{24} L \epsilon_n + \frac{h^5}{2880} \epsilon_n \left( \xi \right)$$

$$\epsilon_{n+1} \leq \left[ 1 + hL + \frac{h^2}{2} L + \frac{h^3}{6} L - \frac{h^4}{24} L \right] |\epsilon_n| + \frac{h^5}{2880} M \left( \xi \right)$$

$$\leq [1 + C] |\epsilon_n| + B$$
where \[ C = hL \left( \sum_{p=1}^{4} \frac{h^{p-1}}{p!} \right), \quad A = 1 + C, \]
\[ B = \frac{h^5}{2880} M \quad \text{and} \quad |y'(x)| \leq M \]

A simple induction proof gives,
\[ |\varepsilon_n| \leq A^n |\varepsilon_0| + \left( \sum_{k=0}^{n-1} A^k \right) B \]  \hspace{1cm} (4.14)

i.e., for \( A \neq 1 \), where \( A = (1 + C) \) since \( \varepsilon_0 = 0 \) then from the geometric series we have
\[ |\varepsilon_n| \leq \left( \frac{A^n - 1}{A - 1} \right) B \]
if we use the inequality \( 1 + x \leq e^x \)
we now get \( A^n = (1 + C)^n = \left( 1 + hL \left( \sum_{p=1}^{4} \frac{h^{p-1}}{p!} \right) \right)^n \leq e^{Cn} = e^{hL \left( \sum_{p=1}^{4} \frac{h^{p-1}}{p!} \right)} = e^{DL(x_n - x_0)} \)
where \( D = \sum_{p=1}^{4} \left( \frac{h^{p-1}}{p!} \right) \).

By inserting the \( A^n \) into the inequality (4.14) for \( \varepsilon \), we finally get
\[ |\varepsilon_n| \leq \frac{h^4}{2880LD} M \left( e^{DL(x_n - x_0)} - 1 \right) \]
and therefore the GTE for the fourth order arithmetic mean (AM) method is \( O(h^4) \).

**GTE for RKCeM**

By using the procedure to evaluate the GTE for the RKAM and by subtracting eq. (4.13) from eq. (4.5), we can obtain
\[ \varepsilon_{n+1} \leq \varepsilon_n + hL \varepsilon_n + \frac{h^2}{2} L \varepsilon_n + \frac{h^3}{6} L \varepsilon_n + \frac{h^4}{24} L \varepsilon_n + \frac{h^5}{69120} y'(\xi) \]
where \( 0 < \xi < 1 \).
\[ |\varepsilon_{n+1}| \leq \left[ 1 + hL + h^2L \frac{2}{2} + h^3L \frac{6}{6} + h^4L \frac{24}{24} \right] |\varepsilon_n| + \frac{h^5}{69120} M \]
\[ \leq \left[ 1 + C \right] |\varepsilon_n| + B \]
where \[ C = hL \left( \sum_{p=1}^{4} \frac{h^{p-1}}{p!} \right) \]

\[ A = 1 + C \]

\[ B = \frac{h^5}{69120} M \quad \text{and} \quad |y^r(\infty)| \leq M. \]

A simple induction proof gives \[ |e_n| \leq A^n |e_0| + \left( \sum_{k=0}^{n} A^k \right) B \]

i.e., for \( A \neq 1 \), since \( e_0 = 0 \) then from the geometric series, we have

\[ |e_n| \leq \left( \frac{A^n - 1}{A - 1} \right) B \quad (4.15) \]

we use the inequality \( 1 + x \leq e^x \).

We now get \[ A^n = (1+C)^n = 1 + hL \left( \sum_{p=1}^{4} \frac{h^{p-1}}{p!} \right) \leq e^{Cn} \]

\[ \leq e^{Lh_n \left( \sum_{p=1}^{4} \frac{h^{p-1}}{p!} \right)} = e^{DL(x_n - x_0)} \]

where \[ D = \sum_{p=1}^{4} \frac{h^{p-1}}{p!} \]

Therefore, the inequality (4.15) becomes \[ |e_n| \leq \frac{h^4}{69120 LD} M \left( e^{DL(x_n - x_0)} - 1 \right) \] and hence the GTE for the fourth order centroidal mean method is \( O(h^4) \). Also in Chapter - 3, page 72, it is briefly explained that for the system of IVPs the GTE is of order \( 'r' \), where \( r \) is the order of the RK method.

From the above discussion, we can conclude that if the LTE of a numerical method is \( O(h^{p+1}) \) then the GTE is \( O(h^p) \). Thus the estimates of the GTE cannot be used for practical error estimation or error control because, it is less accurate than the LTE.
4.3.3 Error estimation and automatic step size selection

We choose the error estimates as the difference between the RKAM and RKCeM. From eq. (4.11), the error estimate is

\[ \text{Err - Est} = |Y_{AM} - Y_{CeM}| \times \frac{281}{13824} \]  \hspace{1cm} (4.16)

For an automatic step size selection, a facility has been introduced in the software to decrease as well as increase the step size according to the error tolerance. This is mainly introduced to estimate an error as well as to optimize the number of computations with respect to the desired accuracy of results. Due to the above, in the error estimation, any of the following is to be executed during the computation, at a time.

Step (a): If the estimated truncation error Err-Est given by eq. (4.16) exceeds a certain pre-set tolerance \( \varepsilon = 0.00001 \) then the routine halves the step size, recomputes Err-Est and tests again.

Step (b): If Err-Est is smaller than \( 2^4 \varepsilon \), then the routine doubles the step size, recomputes Err-Est and tests again (since it is of \( O(h^4) \) method, \( 2^4 \) is used instead of \( 2^5 \)).

It is also to be noted that if the desired accuracy is attained, the step size is neither decreased nor increased. We can make the test more stringent by increasing the Err-Est in eq. (4.16), from fourth order to fifth order, by multiplying eq. (4.16) with a constant \( `h` \) (\( l < 1 \)), and is in the form

\[ \text{Err - Est} = |Y_{AM} - Y_{CeM}| \times h \times \frac{281}{13824} \] \hspace{1cm} (4.17)

by using (4.17) also we can maintain the higher accuracy of results.
4.4 COMPARISON OF RKACeM(4,4), RK(4,4), RKF(4,5) AND RK MERSON METHODS

To study the effectiveness of the RKACeM (4, 4) method, results are compared with RK(4, 4), RKF(4, 5) and RK Merson.

4.4.1 RK(4,4)

Evans and Yaakub (1995, 1999) introduced a new Runge-Kutta method based on Arithmetic Mean and Contraharmonic Mean [55, 154]. The RK (4,4) method in Butcher array form is

\[
\begin{align*}
0 & | & 1 & 1 \\
1/2 & | & 1/2 & 0 & 1 \\
1/2 & | & 0 & 1/2 & 1 \\
1 & | & 1/3 & 8/3 \\
2 & | & 1 & -3 & 3 \\
4 & | & 1/4 & 3/4 \\
E^T & | & 1 & 1 & 1 \\
& | & 3 & 3 & 3 \\
& | & 1 & 1 & 1 \\
& | & 3 & 3 & 3 \\
\end{align*}
\]

(4.18)

In this method, six stages are required to obtain the approximate solution. The estimate of the local truncation error LTE is given by

\[
\text{LTE}_\varepsilon = \frac{h^5}{23040} \left[ 86 f_y^4 + 16 f_y^4 f_{yy}^2 + 12 f_y^2 f_{yy} + 600 f_y^2 f_{yy}^2 + 591 f_y^2 f_{yy} f_{yyy} \right]
\]
4.4.2 RKF(4,5)

The most popular embedded method is RKF(4,5), which is one of the class of methods developed by Fehlberg [17, 59]. This method, in Butcher array form, is written as

\[
\begin{array}{c|ccc}
0 & & & \\
1 & 1 & & \\
\frac{1}{4} & & & \\
3 & 3 & 9 & \\
\frac{8}{32} & & & \\
12 & 1932 & -7200 & 7296 \\
13 & 2197 & 2197 & 2197 \\
1 & 439 & -8 & 3680 & -845 \\
\frac{216}{2} & & & 513 \quad 4104 \\
1 & -8 & 2 & -3544 & 1859 & -11 \\
\frac{27}{2} & & & 2565 \quad 4104 \quad 40 \\
25 & 0 & 1408 & 2197 & -1 \\
\frac{216}{2} & & & 2565 \quad 4104 \quad 5 \\
16 & 0 & 6656 & 28561 & -9 \quad 2 \\
\frac{135}{2} & & & 12825 \quad 56430 \quad 50 \quad 55 \\
1 & 0 & -128 & -2197 & 1 \quad 2 \\
\frac{360}{2} & & & 4275 \quad 75240 \quad 50 \quad 55 \\
\end{array}
\]

(4.19)

which share the same set of vectors \(\{k_i\}\) and five stages are required to obtain the solution and six stages are required when the error estimate is used. The last row represents an estimate of the local truncation error

\[
\text{LTE} = h \left| \frac{k_1}{360} - \frac{128k_3}{4275} - \frac{2197k_4}{75240} + \frac{k_5}{50} + \frac{2k_6}{55} \right|
\]

4.4.3 RK Merson Method

An early example of a Runge-Kutta method with an error estimate in terms of computed values \(k_i\) was proposed by Merson [17, 71]. It is defined in the Butcher array as
which is a 5 - stage method of order 4. Merson proposed the principal local truncation error (LTE) as

\[ LTE = \frac{h}{30} \left[ -2k_1 + 9k_3 - 8k_4 + k_5 \right] \]

It is clear that, from eqs. (4.6), (4.18) and (4.19), the methods RKACeM(4,4), RK(4,4) and RKF(4,5) call the function f(x,y) six times per iteration, while the RK Merson method in eq.(4.20) calls only 5 times per iteration.

4.5 NUMERICAL EXPERIMENTS

The following is a list of sample system of IVPs used in the numerical experiments.

<table>
<thead>
<tr>
<th>Problem</th>
<th>System</th>
<th>Initial Condition</th>
<th>Analytical Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>y'+y = 0</td>
<td>y(0) = 1</td>
<td>y = exp(-t)</td>
</tr>
<tr>
<td>2</td>
<td>y'+y - t - 1 = 0</td>
<td>y(0) = 1</td>
<td>y = t + exp(-t)</td>
</tr>
<tr>
<td>3</td>
<td>y'+y - t^2 -1 = 0</td>
<td>y(0) = 1</td>
<td>y = -2 exp(-t) + t^2 - 2t + 3</td>
</tr>
<tr>
<td>4</td>
<td>y' + 3t^2y = 0</td>
<td>y(0) = 1</td>
<td>y = exp(-t^3)</td>
</tr>
<tr>
<td>5</td>
<td>An oscillatory problem y' - ycos(t) = 0</td>
<td>Y(0) = 1</td>
<td>y = exp(sin(t))</td>
</tr>
</tbody>
</table>

Continued...
| 6 | $y'' = -y$  
    ie. $y_1' = y_2$  
    $y_2' = -y_1$  
| 7 | $y_1' = -1/y_2$  
    $y_2' = -1/y_1$  
| 8 | Non-Linear Singular system  
    $y_1(0) = 0$  
    $y_2(0) = 0$  
| 9 | Stiff system  
    $y_1(t) = \sin(t) + \cos(t)$  
    $y_2(t) = \cos(t) - \sin(t)$  

**Problem 10: State-space system of Electronic Circuit**

Consider the physical model of an electronic circuit discussed on page 83, of Chapter 3.

The generalized state-space systems equation is

$$
\begin{bmatrix}
2 & 2 & 0 & 0 \\
0 & 0 & 2 & 2 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
v_1' \\
v_2' \\
i_3' \\
i_4'
\end{bmatrix}
= 
\begin{bmatrix}
0 & 0 & -1 & 0 \\
-1 & 0 & 1 & 0 \\
1 & -1 & 0 & 0 \\
0 & 0 & 1 & -1
\end{bmatrix}
\begin{bmatrix}
v_1 \\
v_2 \\
i_3 \\
i_4
\end{bmatrix}
+ 
\begin{bmatrix}
E_a \\
J_b
\end{bmatrix}
$$

and by taking

$$
E_a = 1 + t + \frac{t^2}{2} + \frac{t^3}{3} \\
J_b = 1 + t + t^2
$$

the corresponding exact solution is

$$
v_1(t) = \frac{-93}{2} (1 - \sqrt{5}) \exp \left( \frac{1 + \sqrt{5}}{8} t \right) - \frac{93}{2} (1 + \sqrt{5}) \times \exp \left( \frac{1 - \sqrt{5}}{8} t \right) - 27t + \frac{3t^2}{2} - \frac{t^3}{3} + 163
$$

$$
v_2(t) = \frac{-93}{2} (1 - \sqrt{5}) \exp \left( \frac{1 + \sqrt{5}}{8} t \right) - \frac{93}{2} (1 + \sqrt{5}) \times \exp \left( \frac{1 - \sqrt{5}}{8} t \right) - 26t + 26t + 2t^2 + 164
$$
$$i_2(t) = -39 \exp \left( \frac{1+\sqrt{5}}{8} \right) t - 39 \exp \left( \frac{1-\sqrt{5}}{8} \right) t - 14t + 2t^3 + 106$$

$$i_4(t) = -39 \exp \left( \frac{1+\sqrt{5}}{8} \right) t - 39 \exp \left( \frac{1-\sqrt{5}}{8} \right) t - 15t + t^2 + 105$$

with initial conditions.

$$\begin{bmatrix} v_1(0) & v_2(0) & i_3(0) & i_4(0) \end{bmatrix}^T = \begin{bmatrix} 70 & 71 & -80 & -81 \end{bmatrix}^T$$

### 4.5.1 Experimental results for RKACeM(4,4)

The following Table 4.1 presents the numerical results of testing RKACeM(4,4) with error control i.e., using the Err-Est (4.16) and (4.17) for automatic step size for the sample problem 1:

$$y'+y = 0 \quad \text{with} \quad y(0) = 1 \quad \text{and the exact solution is} \quad y = \exp (-t)$$

There is no significant difference in using the Err-Est (4.16) and (4.17) for the procedure of automatic step size selection.

**Table 4.1 Solution for Problem – 1**

<table>
<thead>
<tr>
<th>Time</th>
<th>Solution by</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exact</td>
<td>RKACeM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h = 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>νh = 0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>6.0653E-01</td>
<td>6.0649E-01</td>
</tr>
<tr>
<td>1.00</td>
<td>3.6788E-01</td>
<td>3.6800E-01</td>
</tr>
<tr>
<td>1.50</td>
<td>2.2313E-01</td>
<td>2.2329E-01</td>
</tr>
<tr>
<td>2.00</td>
<td>1.3534E-01</td>
<td>1.3549E-01</td>
</tr>
<tr>
<td>2.50</td>
<td>8.2085E-02</td>
<td>8.2209E-02</td>
</tr>
<tr>
<td>3.00</td>
<td>4.9787E-02</td>
<td>4.9882E-02</td>
</tr>
<tr>
<td>3.50</td>
<td>3.0197E-02</td>
<td>3.0267E-02</td>
</tr>
<tr>
<td>h = 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.50</td>
<td>1.1109E-02</td>
<td>1.0916E-02</td>
</tr>
<tr>
<td>5.50</td>
<td>4.0868E-03</td>
<td>4.0935E-03</td>
</tr>
<tr>
<td>6.50</td>
<td>1.5034E-03</td>
<td>1.5350E-03</td>
</tr>
<tr>
<td>7.50</td>
<td>5.5308E-04</td>
<td>5.7564E-04</td>
</tr>
<tr>
<td>8.50</td>
<td>2.0347E-04</td>
<td>2.1587E-04</td>
</tr>
<tr>
<td>9.50</td>
<td>7.4852E-05</td>
<td>8.0950E-05</td>
</tr>
<tr>
<td>10.50</td>
<td>2.7536E-05</td>
<td>3.0356E-05</td>
</tr>
</tbody>
</table>
The following Tables, 4.2 and 4.3, contain the results of testing RKACeM(4,4) using Err-Est (4.16) and (4.17) for the oscillatory problem 5, depicting the significant difference between the two error estimates, relevant to step halving for automatic step-size selection, in which one is of $O(h^4)$ and the other is $O(h^5)$.

The oscillatory problem 5:

$$y' - y \cos(t) = 0 \quad \text{with } y(0) = 1$$
and the exact solution is $$y = \exp(\sin(t))$$

**Table 4.2  Err-Est By (4.16) for Problem – 5**

<table>
<thead>
<tr>
<th>Time</th>
<th>Solution by</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exact</td>
<td>RKACeM Absolute</td>
</tr>
<tr>
<td>h = 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h = 0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5000</td>
<td>1.6152E+00</td>
<td>1.6144E+00 7.4434E-04</td>
</tr>
<tr>
<td>1.0000</td>
<td>2.3198E+00</td>
<td>2.3194E+00 3.4332E-04</td>
</tr>
<tr>
<td>h = 0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1250</td>
<td>2.4652E+00</td>
<td>2.4653E+00 7.0334E-05</td>
</tr>
<tr>
<td>1.2500</td>
<td>2.5831E+00</td>
<td>2.5832E+00 1.2994E-04</td>
</tr>
<tr>
<td>1.3750</td>
<td>2.6668E+00</td>
<td>2.6671E+00 2.4176E-04</td>
</tr>
<tr>
<td>h = 0.125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4375</td>
<td>2.6943E+00</td>
<td>2.6943E+00 5.2452E-05</td>
</tr>
<tr>
<td>1.5000</td>
<td>2.7115E+00</td>
<td>2.7116E+00 8.9884E-05</td>
</tr>
<tr>
<td>1.5625</td>
<td>2.7182E+00</td>
<td>2.7185E+00 2.7514E-04</td>
</tr>
<tr>
<td>h = 0.0625</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.578125</td>
<td>2.7182E+00</td>
<td>2.7182E+00 4.5300E-06</td>
</tr>
<tr>
<td>1.593750</td>
<td>2.7176E+00</td>
<td>2.7176E+00 1.0252E-05</td>
</tr>
<tr>
<td>1.609375</td>
<td>2.7163E+00</td>
<td>2.7163E+00 4.7684E-06</td>
</tr>
<tr>
<td>1.625000</td>
<td>2.7143E+00</td>
<td>2.7143E+00 3.0994E-06</td>
</tr>
<tr>
<td>1.640625</td>
<td>2.7117E+00</td>
<td>2.7117E+00 2.3842E-06</td>
</tr>
<tr>
<td>1.656250</td>
<td>2.7084E+00</td>
<td>2.7084E+00 1.6689E-06</td>
</tr>
<tr>
<td>1.671875</td>
<td>2.7044E+00</td>
<td>2.7044E+00 1.4305E-06</td>
</tr>
<tr>
<td>1.687500</td>
<td>2.6999E+00</td>
<td>2.6999E+00 9.5367E-07</td>
</tr>
<tr>
<td>1.703125</td>
<td>2.6946E+00</td>
<td>2.6946E+00 9.5367E-07</td>
</tr>
<tr>
<td>1.718750</td>
<td>2.6888E+00</td>
<td>2.6888E+00 7.1526E-07</td>
</tr>
<tr>
<td>1.734375</td>
<td>2.6822E+00</td>
<td>2.6822E+00 7.1526E-07</td>
</tr>
<tr>
<td>1.750000</td>
<td>2.6849E+00</td>
<td>2.6849E+00 -2.1674E-08</td>
</tr>
<tr>
<td>Time</td>
<td>Solution by</td>
<td>Error</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td></td>
<td>Exact</td>
<td>RKACeM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h = 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h = 0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5000</td>
<td>1.6152E+00</td>
<td>1.6144E+00</td>
</tr>
<tr>
<td>1.0000</td>
<td>2.3198E+00</td>
<td>2.3194E+00</td>
</tr>
<tr>
<td>h = 0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2500</td>
<td>2.5831E+00</td>
<td>2.5839E+00</td>
</tr>
<tr>
<td>h = 0.125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3750</td>
<td>2.6668E+00</td>
<td>2.6671E+00</td>
</tr>
<tr>
<td>1.5000</td>
<td>2.7115E+00</td>
<td>2.7120E+00</td>
</tr>
<tr>
<td>1.6250</td>
<td>2.7143E+00</td>
<td>2.7136E+00</td>
</tr>
<tr>
<td>1.7500</td>
<td>2.6751E+00</td>
<td>2.6744E+00</td>
</tr>
<tr>
<td>1.8750</td>
<td>2.5963E+00</td>
<td>2.5960E+00</td>
</tr>
<tr>
<td>2.0000</td>
<td>2.4826E+00</td>
<td>2.4824E+00</td>
</tr>
<tr>
<td>2.1250</td>
<td>2.3404E+00</td>
<td>2.3403E+00</td>
</tr>
<tr>
<td>2.2500</td>
<td>2.1773E+00</td>
<td>2.1772E+00</td>
</tr>
<tr>
<td>2.3750</td>
<td>2.0011E+00</td>
<td>2.0011E+00</td>
</tr>
<tr>
<td>2.5000</td>
<td>1.8193E+00</td>
<td>1.8193E+00</td>
</tr>
<tr>
<td>2.6250</td>
<td>1.6387E+00</td>
<td>1.6387E+00</td>
</tr>
<tr>
<td>2.7500</td>
<td>1.4647E+00</td>
<td>1.4647E+00</td>
</tr>
<tr>
<td>2.8750</td>
<td>1.3014E+00</td>
<td>1.3014E+00</td>
</tr>
<tr>
<td>3.0000</td>
<td>1.1516E+00</td>
<td>1.1516E+00</td>
</tr>
</tbody>
</table>

Using the strategy, discussed in section 3.3 for an automatic step-size selection for the problems 1 to 5, we determine the initial step-size for the Err-Est given in Eqs. (4.16) and (4.17), presented in Table 4.4 and the time taken for all the discussed methods RKACeM(4,4), RK (4,4), RKF (4,5) and RK Merson is given in Table 4.5. (As Pentium II processor gives the same time 0.05 seconds for all the methods, and in order to identify the execution time for each method separately, much lower speed processor 386 is used).
Table 4.4  Initial step size using the error estimate (4.16) and 4.17)

<table>
<thead>
<tr>
<th>Problem</th>
<th>(4.16)</th>
<th>(4.17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>2</td>
<td>0.1250</td>
<td>0.0625</td>
</tr>
<tr>
<td>3</td>
<td>0.2500</td>
<td>0.1250</td>
</tr>
<tr>
<td>4</td>
<td>0.1250</td>
<td>0.1250</td>
</tr>
<tr>
<td>5</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
</tbody>
</table>

Table 4.5  Time taken in seconds for the embedded RK methods

<table>
<thead>
<tr>
<th>Problem</th>
<th>RKACeM (4,4)</th>
<th>RK (4,4)</th>
<th>RKF (4,5)</th>
<th>RK Merson</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.70</td>
<td>3.02</td>
<td>2.97</td>
<td>1.10</td>
</tr>
<tr>
<td>2</td>
<td>5.61</td>
<td>5.66</td>
<td>3.02</td>
<td>1.09</td>
</tr>
<tr>
<td>3</td>
<td>3.07</td>
<td>5.16</td>
<td>1.87</td>
<td>1.15</td>
</tr>
<tr>
<td>4</td>
<td>5.17</td>
<td>5.16</td>
<td>3.13</td>
<td>4.62</td>
</tr>
<tr>
<td>5</td>
<td>2.52</td>
<td>3.35</td>
<td>1.93</td>
<td>2.31</td>
</tr>
</tbody>
</table>

To test the effectiveness of the discussed methods, the numerical test has also been carried out for system of IVPs (problems 2 to 5). The comparison of the accuracy, between RKACeM(4, 4), RK(4, 4), RKF (4, 5) and RK Merson, are shown in Tables 4.6 to 4.7 with the time interval 0 ≤ t ≤ 3 and the error graphs are drawn and presented in Figures 4.1 to 4.3 for the problems 1 to 10, at time t = 3.

Table 4.6  Absolute Error

<table>
<thead>
<tr>
<th>Problems</th>
<th>Time</th>
<th>RKACeM(4,4)</th>
<th>RK(4,4)</th>
<th>RKF(4,5)</th>
<th>RK Merson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem 1</td>
<td>1.00</td>
<td>3.5763E-07</td>
<td>2.6822E-07</td>
<td>2.9802E-08</td>
<td>8.9407E-08</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>2.6822E-07</td>
<td>2.5332E-07</td>
<td>1.4901E-08</td>
<td>5.9605E-08</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>1.0058E-07</td>
<td>8.9407E-08</td>
<td>5.2154E-08</td>
<td>2.6077E-08</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>9.6560E-06</td>
<td>2.8491E-05</td>
<td>0.00000E+00</td>
<td>1.1921E-07</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>2.8610E-06</td>
<td>7.6294E-06</td>
<td>0.00000E+00</td>
<td>0.00000E+00</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>1.6689E-06</td>
<td>3.0994E-06</td>
<td>7.1526E-07</td>
<td>7.1526E-07</td>
</tr>
</tbody>
</table>

Continued...
### Tables 4.7, 4.8 and Figures 4.1 - 4.4 are presented in pages 121 to 123.

#### 4.6 CONCLUSIONS

From the results, we can see that the numerical solution of IVPs by the RKACeM(4,4), RK(4,4), RKF (4,5) and RK Merson methods are comparable in terms of accuracy and time. Altogether, the proposed method RKACeM(4,4) is better than the method RK(4,4) in terms of error and time taken (Table 4.5 – 4.8 and Figures 4.1 – 4.3). The analytical and discrete solutions of the state-space electronic circuit problem are studied by all the RK methods based on a variety of means in chapter 3.5, and here, in this chapter, by the newly proposed RKACeM(4,4) method. Figure 4.4 shows the efficiency of the RKACeM(4,4) method over the methods RKAM and RKCeM. Hence, this RKACeM(4,4) gives another and better alternative approach for solving IVPs.
### Table - 4.7

<table>
<thead>
<tr>
<th>Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
</tr>
<tr>
<td>time</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>2.00</td>
</tr>
<tr>
<td>3.00</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>2.00</td>
</tr>
<tr>
<td>3.00</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>2.00</td>
</tr>
<tr>
<td>3.00</td>
</tr>
</tbody>
</table>

### Table - 4.8

<table>
<thead>
<tr>
<th>Methods</th>
<th>Problem 9</th>
<th>Problem 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>$y_1(t)$</td>
<td>$y_2(t)$</td>
</tr>
<tr>
<td>RKAcM(4,4)</td>
<td>1.00</td>
<td>8.9049E-05, 9.9540E-06</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.6093E-05, 4.0174E-05</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>3.2663E-05, 2.2173E-05</td>
</tr>
<tr>
<td>RK(4,4)</td>
<td>1.00</td>
<td>2.6453E-04, 2.5705E-05</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>3.3945E-05, 1.1349E-04</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>8.3566E-05, 4.2856E-05</td>
</tr>
<tr>
<td>RKF(4,5)</td>
<td>1.00</td>
<td>1.1921E-06, 5.3048E-06</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.4037E-05, 4.6492E-06</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>1.1802E-05, 2.3246E-05</td>
</tr>
<tr>
<td>RKMerson</td>
<td>1.00</td>
<td>5.9605E-07, 5.0515E-06</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.2815E-05, 4.1723E-06</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>1.0371E-05, 2.0921E-05</td>
</tr>
</tbody>
</table>
Figure - 4.1 Error-graph for problems 1 - 5

Figure - 4.2 Error-graph for problems 6 - 8
Figure 4.3 Error-graph for problems 9 & 10

Figure 4.4 Error area blocks in RKAM, RKCeM and RKACeM(4,4) for Electronic circuit problem