# Approximate Computation of Errors in Numerical Integration of Ordinary Differential Equations by One-step Methods 

Hisayoshi Shintani<br>(Received March 10, 1965)

## 1. Introduction

Given a differential equation

$$
\begin{equation*}
y^{\prime}=f(x, y) \tag{1.1}
\end{equation*}
$$

where $f(x, y)$ is assumed to be a sufficiently smooth function.
In numerous papers $[1-16]^{1)}$, various methods are obtained for bounding or approximating the errors in numerical integration of (1.1) by one-step methods with the aids of the functions that bound or approximate the function $f_{y}(x, y)$, the truncation error and so on.

To avoid the use of such functions for practical purposes, in this paper, $n$ steps of integration with a fixed step-size are considered as one step and a simple method is obtained for approximating the errors without computing explicitly any function other than $f(x, y)$. The method is illustrated by two numerical examples.

Since usually the step-size is not changed so often and the estimate of the error is not always necessary for each step of integration, it will not be a serious restriction to fix the step-size for the $n$ steps of integration, and this method may be used as an integration method with a check on the accuracy of the numerical solution.

## 2. Preliminaries

In this paragraph and the next, we state five lemmas and one theorem without proof, but they are proved in paragraph 6.

In the sequel, for simplicity, we assume that $f(x, y)$ is defined, continuous and bounded in the strip

$$
S: x_{0} \leqq x \leqq x_{0}+n h_{0}, \quad|y|<\infty,
$$

and that the partial derivatives of $f(x, y)$ up to the necessary order, say $s$,

[^0]exist and are continuous and bounded in $S$, where $n$ is a positive integer and $h_{0}$ is a positive number. ${ }^{2)}$

For any $u$ and $v$ such that $x_{0} \leqq u \leqq x_{0}+(n-1) h_{0}$ and $|v|<\infty$, let $y(x ; u, v)$ be the solution of (1.1) such that $y(u ; u, v)=v$, and $\Phi(u, v ; h)$ be the increment function of any one-step method of order $p(p \geqq 1)$ for approximating $y(u+h ; u, v)\left(0<h \leqq h_{0}\right)[6]$. Then $y(u+h ; u, v)$ can be written as follows:

$$
\begin{align*}
y(u+h ; u, v) & =v+h \Phi(u, v ; h)-T(u, v ; h)  \tag{2.1}\\
& =v+h \Delta(u, v ; h),
\end{align*}
$$

where $T(u, v ; h)$ is a truncation error,

$$
\begin{equation*}
T(u, v ; h)=O\left(h^{p+1}\right)^{3)} \tag{2.2}
\end{equation*}
$$

and $\Delta(u, v ; h)$ is the exact relative increment function [6].
We are concerned with the solution

$$
\begin{equation*}
y(x)=y\left(x ; x_{0}, y_{0}\right) \tag{2.3}
\end{equation*}
$$

and consider the case where the equation (1.1) is integrated numerically from $x_{0}$ to $x_{0}+n h$ with step-size $h\left(0<h \leqq h_{0}\right)$. Hence we put

$$
\begin{align*}
& x_{j+1}=x_{0}+(j+1) h,  \tag{2.4}\\
& p_{j+1}=\Phi\left(x_{j}, y_{j} ; h\right),  \tag{2.5}\\
& y_{j+1}=y_{j}+h p_{j+1} \quad(j=0,1, \cdots, n-1),  \tag{2.6}\\
& d_{i}=y_{i}-y\left(x_{i}\right), \tag{2.7}
\end{align*}
$$

and

$$
\begin{equation*}
f_{i}=f\left(x_{i}, y_{i}\right) \quad(i=0,1, \cdots, n) . \tag{2.8}
\end{equation*}
$$

To take into consideration the propagation of error, along with $y(x)$ we consider a neighboring solution

$$
\begin{equation*}
y(x ; e)=y\left(x ; x_{0}, y_{0}-e\right), \tag{2.9}
\end{equation*}
$$

and put

$$
\begin{equation*}
c(x)=y(x)-y(x ; e) \tag{2.10}
\end{equation*}
$$

[^1]Then it is seen that, if $e$ is an approximate value of the error of $y_{0}$, then $c\left(x_{i}\right)$ $+d_{i}$ provides an approximation to the error of $y_{i}$ and that $d_{i}$ is the local error of $y_{i}$. For these reasons we try to obtain the approximate values of $c\left(x_{i}\right)+d_{i}$ and $d_{i}(0<i \leqq n)$.

We note that $c(x)$ is the solution of the differential equation

$$
\begin{equation*}
c^{\prime}=F(x, y(x), c) \tag{2.11}
\end{equation*}
$$

satisfying the initial condition

$$
\begin{equation*}
c\left(x_{0}\right)=e \tag{2.12}
\end{equation*}
$$

and that it can be written as

$$
\begin{equation*}
c(x)=e \cdot O(1) \tag{2.13}
\end{equation*}
$$

where

$$
\begin{equation*}
F(x, y, u)=f(x, y)-f(x, y-u) . \tag{2.14}
\end{equation*}
$$

Let $r$ be a positive integer not greater than $n$, and let

$$
\begin{equation*}
m=2 n-r+1, \tag{2.15}
\end{equation*}
$$

$$
\begin{gather*}
T(x, y ; h)=h^{p}\left[\sum_{i=1}^{m} h^{i} \varphi_{i}(x, y)+O\left(h^{m+1}\right)\right]^{4)},  \tag{2.16}\\
h \Delta(x, y ; h)=\sum_{i=1}^{m} h^{i} \Delta_{i}(x, y)+O\left(h^{m+1}\right), \tag{2.17}
\end{gather*}
$$

and

$$
\begin{equation*}
d_{j}=h^{p} e_{j} \quad(j=0,1, \ldots, n) . \tag{2.18}
\end{equation*}
$$

Then we have the following
Lemma 1. $e_{j}$ and $F\left(x_{j}, y_{j}, d_{j}\right)$ can be written as follows:

$$
\begin{equation*}
e_{j}=\sum_{i=1}^{m} h^{i} P_{i}(j)+O\left(h^{m+1}\right), \tag{2.19}
\end{equation*}
$$

and

$$
\begin{equation*}
F\left(x_{j}, y_{j}, d_{j}\right)=h^{p}\left[\sum_{i=1}^{m} h^{i} Q_{i}(j)+O\left(h^{m+1}\right)\right] \quad(j=0,1, \ldots, n), \tag{2.20}
\end{equation*}
$$

4) Here it is assumed that $s \geqq p+m$.
where $P_{i}(x)$ and $Q_{i}(x)$ are polynomials in $x$ of degree at most $i$,

$$
\begin{equation*}
P_{i}(0)=0 \tag{2.21}
\end{equation*}
$$

and

$$
\begin{equation*}
Q_{i}(0)=0 \quad(i=1,2, \cdots, m) \tag{2.22}
\end{equation*}
$$

Lemma 2. There exist constants $a_{k j}$ and $b_{k j}(k, j=0,1, \ldots, n)$ that satisfy the conditions

$$
\begin{equation*}
l \sum_{j=0}^{n} j^{l-1} a_{k j}+\sum_{j=0}^{n} j^{l} b_{k j}=k^{l} \quad(l=1,2, \ldots, m) \tag{2.23}
\end{equation*}
$$

and

$$
\begin{equation*}
\sum_{j=0}^{n} j^{l} b_{k j}=0 \quad(l=0,1, \ldots, r) \tag{2.24}
\end{equation*}
$$

Lemma 3. For $y(x)$ it is valid that

$$
\begin{equation*}
y\left(x_{k}\right)-y\left(x_{0}\right)=h \sum_{j=0}^{n} a_{k j} y^{\prime}\left(x_{j}\right)+\sum_{j=0}^{n} b_{k j} y\left(x_{j}\right)+O\left(h^{m+1}\right) \tag{2.25}
\end{equation*}
$$

and

$$
\begin{equation*}
\sum_{j=0}^{n} b_{k j} y\left(x_{j}\right)=O\left(h^{r+1}\right) \quad(k=0,1, \cdots, n) \tag{2.26}
\end{equation*}
$$

where $a_{k j}$ and $b_{k j}(k, j=0,1,2, \ldots, n)$ are constants satisfying (2.23) and (2.24).
Corresponding to (2.25), we put

$$
\begin{align*}
S_{k}=y_{k}-y_{0}-h \sum_{j=0}^{n} a_{k j} f_{j}-h \sum_{j=1}^{n} b_{k j} \sum_{q=1}^{j} p_{q},  \tag{2.27}\\
g_{k}=d_{k}-S_{k} \quad(k=0,1, \cdots, n), \tag{2.28}
\end{align*}
$$

and define $\delta_{k}$ by the formula

$$
\delta_{k}= \begin{cases}1, & \text { if } \sum_{j=0}^{n} j^{r+1} b_{k j}=0  \tag{2.29}\\ 0, & \text { otherwise }\end{cases}
$$

Then we have the following
Lemma 4. $\quad g_{k}$ can be written as follows:

$$
\begin{equation*}
g_{k}=h^{p}\left[\sum_{i=1}^{m} h^{i+1} R_{i+1}(k)+\sum_{i=r+1}^{m} h^{i} c_{k i}\right]+O\left(h^{m+1}\right) \quad(k=0,1, \ldots, n) \tag{2.30}
\end{equation*}
$$

where

$$
\begin{equation*}
R_{i+1}(u)=\int_{0}^{u} Q_{i}(t) d t \quad(i=1,2, \ldots, m) \tag{2.31}
\end{equation*}
$$

$c_{k i}$ 's are constants,

$$
\begin{equation*}
c_{0 i}=0 \tag{2.32}
\end{equation*}
$$

and

$$
\begin{equation*}
c_{k r+1}=0 \quad \text { when } \quad \delta_{k}=1 \tag{2.33}
\end{equation*}
$$

## 3. Approximate error formulas

From lemma 4 it is readily seen that, if

$$
\begin{equation*}
m \geqq p+1 \tag{3.1}
\end{equation*}
$$

then

$$
\begin{equation*}
g_{k}=O\left(h^{p+2}\right) \tag{3.2}
\end{equation*}
$$

Hence we have
Theorem 1. Under the condition (3.1), the local error of $y_{k}$ satisfies

$$
\begin{equation*}
y_{k}-y\left(x_{k}\right)=S_{k}+O\left(h^{p+2}\right) \quad(k=1,2, \cdots, n) . \tag{3.3}
\end{equation*}
$$

To obtain better approximation to the error, in the sequel, we consider the case where the condition

$$
\begin{equation*}
m \geqq p+r+1 \tag{3.4}
\end{equation*}
$$

namely the condition

$$
\begin{equation*}
2(n-r) \geqq p \geqq 1 \tag{3.5}
\end{equation*}
$$

is satisfied.
Let

$$
\begin{equation*}
x=x_{u}=x_{0}+u h \quad(0 \leqq u \leqq n) \tag{3.6}
\end{equation*}
$$

and, corresponding to (2.19) and (2.30), we define the functions $e(x)$ and $g(x)$ by the formulas

$$
\begin{equation*}
e(x)=\sum_{i=1}^{m} h^{i} P_{i}(u)+h^{m+1} a(u) \tag{3.7}
\end{equation*}
$$

and

$$
\begin{equation*}
g(x)=h^{p}\left[\sum_{i=1}^{m} h^{i+1} R_{i+1}(u)+\sum_{i=r+1}^{m} h_{i}^{i} c_{i}(u)\right]+h^{m+1} b(u), \tag{3.8}
\end{equation*}
$$

where $a(u), c_{i}(u)$ and $b(u)$ are interpolation polynomials of degree at most $n$ such that

$$
\begin{equation*}
e\left(x_{k}\right)=e_{k}, \quad c_{i}(k)=c_{k}, \quad g\left(x_{k}\right)=g_{k} \quad(k=0,1, \ldots, n) . \tag{3.9}
\end{equation*}
$$

Making use of these functions and $y(x)$, we define the functions $d(x), v(x)$ and $S(x)$ as follows:

$$
\begin{equation*}
d(x)=h^{p} e(x) \tag{3.10}
\end{equation*}
$$

$$
\begin{equation*}
v(x)=y(x)+d(x) \tag{3.11}
\end{equation*}
$$

and

$$
\begin{equation*}
S(x)=d(x)-g(x) . \tag{3.12}
\end{equation*}
$$

Then evidently

$$
\begin{equation*}
d\left(x_{k}\right)=d_{k}, \quad v\left(x_{k}\right)=y_{k}, \quad S\left(x_{k}\right)=S_{k}, \tag{3.13}
\end{equation*}
$$

and we have the following
Lemma 5. Let $w(x ; e)$ be the solution of the differential equation

$$
\begin{equation*}
w^{\prime}=F(x, v(x), S(x)+w) \tag{3.14}
\end{equation*}
$$

satisfying the initial condition

$$
\begin{equation*}
w\left(x_{0} ; e\right)=e . \tag{3.15}
\end{equation*}
$$

Then, under the condition (3.5), it is valid that

$$
\begin{equation*}
y_{k}-y\left(x_{k} ; e\right)=S_{k}+w\left(x_{k} ; e\right)+O\left(h^{p+r+\delta_{k}+1}\right) \quad(k=0,1, \ldots, n) . \tag{3.16}
\end{equation*}
$$

Since $S_{k}(k=0,1, \ldots, n)$ are computable, this lemma shows that, for the estimation of the error of $y_{k}$, we have only to obtain the approximate value of $w\left(x_{k} ; e\right)$. For this purpose we use some one-step method. It is desired, however, to use such a one-step method as requires only the values of the
derivative for equidistant abscissas, because the values of $S(x)$ and $v(x)$ are known only for $x=x_{k}(k=0,1, \ldots, n)$. Among such methods are Euler, improved Euler, modified Euler, Kutta, Heun, and Runge-Kutta methods and so on.

Let $\psi(t, w ; h)$ be the increment function of such a one-step method of order $q(1 \leqq q \leqq p+r+1)$ for approximating $w(x ; e)$ that does not require the evaluations of the derivative for $x$ other than $x=t+\frac{1}{k} j h(j=0,1, \ldots, k)$, where $k$ is a positive integer not greater than $n$. Let $l$ be an integral multiple of $k$ not greater than $n, l$ be the greatest integer not exceeding $n / l$ and put $\nu=t l$.

We consider the case where the equation (3.14) is integrated numerically from $x_{0}$ to $x_{\nu}$ with step-size $l h$, and put

$$
\begin{align*}
& q_{j+l}(e)=\psi\left(x_{j}, w_{j}(e) ; l h\right)  \tag{3.17}\\
& w_{j+l}(e)=e+l h z_{j+l}(e) \quad(j=0, l, 2 l, \cdots,(t-1) l), \tag{3.18}
\end{align*}
$$

where

$$
\begin{gather*}
z_{j+l}(e)=z_{j}(e)+q_{j+l}(e)  \tag{3.19}\\
w_{0}(e)=e \tag{3.20}
\end{gather*}
$$

and

$$
\begin{equation*}
z_{0}(e)=0 \tag{3.21}
\end{equation*}
$$

Then we have the following

Theorem 2. Under the conditions (3.5) and

$$
\begin{equation*}
p \geqq n-r-1, \quad 1 \leqq q \leqq p+r+1, \tag{3.22}
\end{equation*}
$$

it is valid that

$$
\begin{equation*}
y_{j}-y\left(x_{j} ; e\right)=T_{j}(e)+e \cdot O\left(h^{q+1}\right)+O\left(h^{p+l_{j}+1}\right) \quad(j=l, 2 l, \cdots, t l) \tag{3.23}
\end{equation*}
$$

where

$$
\begin{equation*}
T_{j}(e)=S_{j}+w_{j}(e), \tag{3.24}
\end{equation*}
$$

and

$$
\begin{equation*}
l_{j}=\min \left(q, r+\delta_{j}\right) \tag{3.25}
\end{equation*}
$$

$T_{j}(e)$ is the desired approximate error formula for $y_{j}$, where $e$ is an approximate value of the error of $y_{0}$.

Remark 1. In the case where $n=\nu, y\left(x_{n}\right)$ can be written as follows:

$$
\begin{equation*}
y\left(x_{0}+n h\right)=y\left(x_{0}\right)+n h \Theta\left(x_{0}, y_{0} ; n h\right)+O\left(h^{p+l_{n}+1}\right) \tag{3.26}
\end{equation*}
$$

where

$$
\begin{equation*}
\Theta\left(x_{0}, y_{0} ; h\right)=\frac{1}{n}\left[\sum_{j=0}^{n} a_{n j} f_{j}+\sum_{j=1}^{n} b_{n j} \sum_{q=1}^{j} p_{q}-l z_{n}(0)\right] . \tag{3.27}
\end{equation*}
$$

Hence $\Theta(x, y ; h)$ can be considered as the increment function of a one-step method of order $p+l_{n}$. In this way we can generate one-step methods of higher orders from those of lower orders.

Remark 2. The assumptions stated at the beginning of paragraph 2 can be replaced by the assumption that there exists a positive number $M$ that satisfies the following conditions:
$1^{\circ}$. $f(x, y)$ is defined and continuous in the domain

$$
G: x_{0} \leqq x \leqq x_{0}+n h_{0}, \quad|y| \leqq M
$$

and the partial derivatives of $f(x, y)$ up to the order $s$ exist and are continuous in $G$, where $s=\max (p+m, q+1)$;
$2^{\circ}$. $y(x)$ and $y(x ; e)$ exist over the interval $\left[x_{0}, x_{n}\right] ;$
$3^{\circ} . \Phi\left(x_{i}, y_{i} ; h\right)(i=0,1, \ldots, n-1)$ are defined;
$4^{\circ}$. $w(x)$ and $w(x ; e)$ exist over $\left[x_{0}, x_{n}\right]$;
$5^{\circ} . \psi\left(x_{j}, w_{j}(e) ; l h\right)(j=0, l, 2 l, \ldots,(t-1) l)$ are defined.

## 4. Round-off errors

In this paragraph, we take into consideration round-off errors and denote by $\hat{a}$ the computed value of $a$.

The computed values $\hat{y}_{i}, \hat{w}_{j}(e)$ and $\hat{T}_{j}(e)$ can be written as follows:

$$
\begin{array}{ll}
\hat{y}_{i}=\hat{y}_{i-1}+h \hat{p}_{i}-r_{i}^{\prime} & (i=1,2, \cdots, n) \\
\hat{w}_{j}(e)=e+l h \hat{z}_{j}(e)-s_{j} & (j=l, 2 l, \cdots, t l) \tag{4.2}
\end{array}
$$

and

$$
\begin{equation*}
\hat{T}_{j}(e)=\hat{S}_{j}+\hat{w}_{j}(e)-t_{j}, \tag{4.3}
\end{equation*}
$$

where $r_{i}^{\prime}, s_{j}$ and $t_{j}$ are round-off errors,

$$
\begin{equation*}
\hat{y}_{0}=y_{0}, \quad \hat{w}_{0}=e \tag{4.4}
\end{equation*}
$$

From these we have

$$
\begin{equation*}
\hat{y}_{i}-y_{i}=\hat{y}_{i-1}-y_{i-1}+h\left(\hat{p}_{i}-p_{i}\right)-r_{i}^{\prime} \tag{4.5}
\end{equation*}
$$

and

$$
\begin{equation*}
\hat{w}_{j}(e)-w_{j}(e)=\operatorname{lh}\left[\hat{z}_{j}(e)-z_{j}(e)\right]-s_{j} . \tag{4.6}
\end{equation*}
$$

Let

$$
\begin{gather*}
\mu=\max \left(\left|\hat{p}_{i}-p_{i}\right|,\left|\hat{z}_{j}(e)-z_{j}(e)\right|,\left|\hat{f}_{k}-f_{k}\right|\right)  \tag{4.7}\\
(i=1,2, \cdots, n ; j=l, 2 l, \cdots, t l ; k=0,1, \cdots, n)
\end{gather*}
$$

Then $\hat{S}_{k}, \hat{w}_{j}(e)$ and $\hat{T}_{j}(e)$ can be written as follows:

$$
\begin{align*}
& \hat{S}_{k}=\hat{y}_{k}-y_{k}+S_{k}+O(h \mu)-r_{k} \quad(k=1,2, \cdots, n),  \tag{4.8}\\
& \hat{w}_{j}(e)=w_{j}(e)+O(h \mu)-s_{j} \quad(j=l, 2 l, \cdots, t l) \tag{4.9}
\end{align*}
$$

and

$$
\begin{equation*}
\hat{T}_{j}(e)=\hat{y}_{j}-y_{j}+T_{j}(e)+O(h \mu)-r_{j}-s_{j}-t_{j}, \tag{4.10}
\end{equation*}
$$

where $r_{k}$ 's are round-off errors.
From these we obtain the following
Theorem 3. Under the condition (3.1), the local error of $\hat{y}_{k}$ satisfies

$$
\begin{equation*}
\hat{y}_{k}-y\left(x_{k}\right)=\hat{S}_{k}+O\left(h^{p+2}\right)+O(h \mu)+r_{k} \quad(k=1,2, \cdots, n) \tag{4.11}
\end{equation*}
$$

and, under the conditions (3.5) and (3.22), it is valid that

$$
\begin{gather*}
\hat{y}_{j}-y(x ; e)=\hat{T}_{j}(e)+e \cdot O\left(h^{q+1}\right)+O\left(h^{p+l_{j}+1}\right)+O(h \mu)+r_{j}+s_{j}+t_{j}  \tag{4.12}\\
(j=l, 2 l, \cdots, t l)
\end{gather*}
$$

where $\mu$ is defined by (4.7) and $r_{k}, s_{j}$ and $t_{j}$ are round-off errors in computing $S_{k}, w_{j}(e)$ and $\hat{S}_{j}+\hat{w}_{j}(e)$ respectively.

From this result it is seen that round-off errors other than $r_{k}, s_{j}$ and $t_{j}$ appear in (4.11) and (4.12) in the form multiplied by $h$ so that $S_{k}, w_{j}(e)$ and $\hat{S}_{j}+\hat{w}_{j}(e)$ should be computed minutely.

To compare $S_{k}$ with round-off errors, we rewrite $S_{k}$ as

$$
\begin{equation*}
R_{k}=y_{k}-y_{0}-h \sum_{j=0}^{n} a_{k j} f_{j}-\sum_{j=0}^{n} b_{k j} y_{j} \tag{4.13}
\end{equation*}
$$

so that theoretically

$$
\begin{equation*}
v_{k}=R_{k}-S_{k}=0 \tag{4.14}
\end{equation*}
$$

Then $\hat{R}_{k}$ and $\hat{v}_{k}$ can be written as follows:

$$
\begin{equation*}
\hat{R}_{k}=\hat{y}_{k}-y_{0}-h \sum_{j=0}^{n} a_{k j} \hat{f}_{j}-\sum_{j=1}^{n} b_{k j} \sum_{q=1}^{j}\left(h \hat{p}_{q}-r_{q}^{\prime}\right)-t_{k}^{\prime}, \tag{4.15}
\end{equation*}
$$

and

$$
\begin{equation*}
\hat{v}_{k}=\sum_{j=1}^{n} b_{k j} \sum_{q=1}^{j} r_{q}^{\prime}+r_{k}-t_{k}^{\prime}-t_{k}^{\prime \prime}, \tag{4.16}
\end{equation*}
$$

where $t_{k}^{\prime}$ and $t_{k}^{\prime \prime}$ are round-off errors. Clearly $\hat{v}_{k}$ is an aggregate of round-off errors.

From these we may conclude that, if the local error of $y_{k}$ dominates round-off errors, then $\hat{v}_{k}$ must be significantly smaller in magnitude than $\hat{S}_{k}$.

## 5. Numerical examples

In the following examples, $n=l=4$ and the Runge-Kutta method is used for approximating $y(x)$ and $w(x ; e)$, so that $p=q=4 . S_{2}, S_{4}$ and $R_{4}$ are computed by the following formulas:

$$
\begin{gather*}
S_{2}=y_{2}-y_{0}-h P+\frac{1}{2} h\left(p_{4}-p_{2}+p_{3}-p_{1}\right),  \tag{5.1}\\
S_{4}=y_{4}-y_{0}-2 h P \tag{5.2}
\end{gather*}
$$

and

$$
\begin{equation*}
R_{4}=\frac{1}{21}\left[5\left(y_{4}-y_{0}\right)+32\left(y_{3}-y_{1}\right)\right]-2 h\left(2 f_{2}+\frac{4}{7} \Delta^{2} f_{1}+\frac{1}{35} \Delta^{4} f_{0}\right) \tag{5.3}
\end{equation*}
$$

where

$$
\begin{gather*}
P=2 f_{2}+\frac{4}{7} \Delta^{2} f_{1}+\frac{1}{35} \Delta^{4} f_{0}+\frac{8}{21}\left(p_{4}-p_{3}+p_{1}-p_{2}\right),  \tag{5.4}\\
m=8, \quad r=1, \quad \delta_{4}=1,
\end{gather*}
$$

and $\Delta$ is the forward difference operator.
Computation is carried out in the floating-point arithmetic with 39 bits mantissa and rounding is done by chopping. At the start of integration, $e$ and $h$ are set equal to 0 and 0.05 respectively. One step of integration in our
extended sense is performed in accordance with the following program:
(1) set parameter $a=0$;
(2) compute $y_{i}(i=1,2,3,4)$ and $S_{4}$;
(3) when the inequality

$$
\begin{equation*}
\varepsilon\left|\hat{y}_{4}\right| \geqq\left|\hat{S}_{4}\right| \quad\left(\varepsilon=5 \times 10^{-7}\right) \tag{5.6}
\end{equation*}
$$

is not satisfied, halve the step-size, set $a=1$ and go to (2);
(4) when (5.6) holds, compute $v_{4}$;
(5) when the inequality

$$
\begin{equation*}
\delta\left|\hat{S}_{4}\right| \geqq\left|\hat{v}_{4}\right| \quad\left(\delta=5 \times 10^{-4}\right) \tag{5.7}
\end{equation*}
$$

is not satisfied, if $a=0$, double the step-size and go to (2); if $a=1$, stop the machine (for then computation must be performed in multiple precision.);
(6) when (5.7) holds, compute $S_{2}$ and $T_{4}(e)$;
(7) replace $y_{0}, x_{0}$ and $e$ with $x_{4}, \hat{y}_{4}$ and $\hat{T}_{4}(e)$ respectively.

The criterion (5.6) is set up to control the local error and (5.7) is made so that round-off errors may not dominate the local error. By our experience the criterion (5.7) improves the accuracy of the numerical solution in some cases.

## Example 1.

$$
\begin{equation*}
y^{\prime}=2 x y, \quad y(0)=1 \tag{5.8}
\end{equation*}
$$

The solution is $y=\exp x^{2}$ and the results are shown in table 1 . The actual errors of the numerical solution computed by the above program with process (5) omitted are $-1.487-06,-1.218-04$ and $-3.226-02$ for $x=1.0,2.0$, and 3.0 respectively. Comparing these results with those in table 1, we can understand the significance of the criterion (5.7). The only difference in the computation procedure is that the step-size used in the first and the second step of integration in our sense is 0.1 and 0.05 for the program with process (5) but 0.05 and 0.1 for the program without process (5).

Example 2.

$$
\begin{equation*}
y^{\prime}=12 x^{3}-\frac{8 y}{x}, \quad y(-1)=1 \tag{5.9}
\end{equation*}
$$

The solution is $y=x^{4}$ and the results are given in table 2. For this equation the origin is a singular point. Since the general solution is $y=x^{4}+C x^{-8}$, even a small error will cause a great trouble. In fact, integrating (5.9) by the

Runge-Kutta method with a uniform step-size $2^{-n} / 10(n=1,2, \ldots, 9)$, we could not have even one significant figure for $x=-0.1$. Without knowing the actual solution, however, we can readily see from table 2 that even one significant figure is not obtained for $x=-0.3,-0.2$ and -0.1 .

Table 1

| $\begin{array}{ll} \hline & \text { error } \\ \mathbf{x} & \end{array}$ | computed | actual |
| :---: | :---: | :---: |
| 1.0 | -8.361-07 | $-8.720-07$ |
| 2.0 | -9.946-05 | -9.941-05 |
| 3.0 | -3.057-02 | -3.039-02 |
| 4.0 | $-6.386+01$ | $-6.343+01$ |
| 5.0 | $-9.764+05$ | $-9.687+05$ |

Table 2

| x | error | computed |
| :---: | :---: | :---: |
| x | actual |  |
| -0.9 | $-2.374-07$ | $-2.370-07$ |
| -0.8 | $-8.889-07$ | $-8.877-07$ |
| -0.7 | $-2.925-06$ | $-2.922-06$ |
| -0.6 | $-1.007-05$ | $-1.006-05$ |
| -0.5 | $-4.331-05$ | $-4.328-05$ |
| -0.4 | $-2.581-04$ | $-2.580-04$ |
| -0.3 | $-2.575-03$ | $-2.578-03$ |
| -0.2 | $-6.599-02$ | $-6.706-02$ |
| -0.1 | $-1.688+01$ | $-1.691+01$ |

In both examples errors seem to be approximated fairly well and this method may be used as an integration method with a check on the accurracy of the numerical solution.

## 6. Proof

### 6.1 Proof of lemma 1

From (2.7), (2.6), (2.4) and (2.1), it follows that

$$
\begin{align*}
& d_{j+1}-d_{j}=h\left[\Delta\left(x_{j}, y\left(x_{j}\right)+d_{j} ; h\right)-\Delta\left(x_{j}, y\left(x_{j}\right) ; h\right)\right]+  \tag{6.1}\\
& \quad+T\left(x_{j}, y\left(x_{j}\right)+d_{j} ; h\right) \quad(j=0,1, \ldots, n-1) .
\end{align*}
$$

Since $d_{0}=0$, from (2.2) it is readily seen that

$$
\begin{equation*}
d_{i}=O\left(h^{p+1}\right) \tag{6.2}
\end{equation*}
$$

and so

$$
\begin{equation*}
e_{i}=O(h) \quad(i=0,1, \cdots, n) \tag{6.3}
\end{equation*}
$$

By (2.16), (2.17) and (2.18), (6.1) can be written as follows:

$$
\begin{align*}
& e_{j+1}-e_{j}=\sum_{i=1}^{m} h^{i} \varphi_{i}\left(x_{j}, y\left(x_{j}\right)\right)+  \tag{6.4}\\
&+\sum_{i+l+(l l i) p>m}^{i, l \geq 1} \\
& \sum \frac{1}{l!} h^{(l-1) p+i} \frac{\partial^{l}}{\partial y^{l}} \Delta_{i}\left(x_{j}, y\left(x_{j}\right)\right) e_{j}^{l}+
\end{align*}
$$

$$
+\sum_{\substack{i+l \\ i, l>l \equiv 1}} \frac{1}{l!} h^{l p+i} \frac{\partial^{l}}{\partial y^{l}} \varphi_{i}\left(x_{j}, y\left(x_{j}\right)\right) e_{j}^{l}+O\left(h^{m+1}\right) .
$$

Making use of the fact that

$$
\begin{equation*}
y\left(x_{j}\right)=y_{0}+\sum_{q=1}^{m} \frac{1}{q!}(j h)^{q} y^{(q)}\left(x_{0}\right)+O\left(h^{m+1}\right), \tag{6.5}
\end{equation*}
$$

and expanding $\varphi_{i}\left(x_{j}, y\left(x_{j}\right)\right), \frac{\partial^{l}}{\partial y^{l}} \Delta_{i}\left(x_{j}, y\left(x_{j}\right)\right)$ and $\frac{\partial^{l}}{\partial y^{l}} \varphi_{i}\left(x_{j}, y\left(x_{j}\right)\right)$ at $x_{0}$, we can write (6.4) as follows:

$$
\begin{equation*}
e_{j+1}-e_{j}=\sum_{i=1}^{m} h^{i} a_{i-1}(j+1)+\sum_{\substack{i+l \leq m \\ i, l \leq 1}} h^{i} b_{i-1}^{(l)}(j+1) e_{j}^{l}+O\left(h^{m+1}\right) \tag{6.6}
\end{equation*}
$$

where $a_{i-1}(x)$ and $b_{i-1}^{(l)}(x)$ are polynomials in $x$ of degree at most $i-1$.
Since $e_{0}=0$, summing (2.28) from $j=0$ to $k-1$, we have

$$
\begin{gather*}
e_{k}=\sum_{i=1}^{m} h^{i} \sum_{j=1}^{k} a_{i-1}(j)+\sum_{\substack{i+l \leq m \\
i, l \geq 1}} h^{i} \sum_{j=1}^{k} b_{i-1}^{(l)}(j) e_{j-1}^{l}+O\left(h^{m+1}\right)  \tag{6.7}\\
(k=1,2, \cdots, n) .
\end{gather*}
$$

Substitution of (6.3) into (6.7) yields

$$
\begin{equation*}
e_{k}=h k a_{0}+O\left(h^{2}\right) \tag{6.8}
\end{equation*}
$$

so that (2.19) and (2.21) are valid for $m=1$ and $i=1$ respectively, where $a_{0}=a_{0}(0)$.

Let $B_{n}(x)$ be the Bernoulli polynomial of degree $n$. Then it is well known that

$$
\begin{equation*}
\sum_{j=1}^{k} j^{l}=\frac{1}{l+1}\left[B_{l+1}(k+1)-B_{l+1}(1)\right] \quad(l \geqq 0) \tag{6.9}
\end{equation*}
$$

The right hand side of (6.9) is a polynomial in $k$ of degree $l+1$ without the constant term. From this fact follow readily (2.19) and (2.21) by induction.

Although (2.19) is proved only for $j \geqq 1$, since $e_{0}=0$ and (2.21) holds, (2.19) is valid also for $j=0$.

By (2.14), (2.17) and (2.18) we have

$$
\begin{align*}
F\left(x_{j}, y_{j}, d_{j}\right) & =f\left(x_{j}, y\left(x_{j}\right)+d_{j}\right)-f\left(x_{j}, y\left(x_{j}\right)\right)  \tag{6.10}\\
& =h^{p}\left[\sum_{\substack{l+(l) 1 \\
l \geq 1}} \frac{1}{l!} h^{(l-1) p} \frac{\partial^{l}}{\partial y^{l}} f\left(x_{j}, y\left(x_{j}\right)\right) e_{j}^{l}+O\left(h^{m+1}\right)\right] .
\end{align*}
$$

Substituting (2.19) into (6.10), making use of (6.5) and expanding $\frac{\partial^{l}}{\partial y^{l}} f\left(x_{j}, y\left(x_{j}\right)\right)$ at $x_{0}$, we obtain (2.20) and (2.22). This completes the proof of lemma 1.

### 6.2 Proof of lemma 2 and lemma 3

Let $E, \Delta$ and $D$ be the shifting, forward difference and differential operators respectively [7]. Then by the well-known formulas

$$
\begin{equation*}
E=1+\Delta \tag{6.11}
\end{equation*}
$$

and

$$
\begin{equation*}
h D=\log (1+\Delta), \tag{6.12}
\end{equation*}
$$

$\Delta$ and $E^{k}-1$ can be rewritten as follows:

$$
\begin{equation*}
\Delta=\left(a_{0}+a_{1} \Delta+a_{2} \Delta^{2}+\ldots\right) h D \tag{6.13}
\end{equation*}
$$

and

$$
\begin{align*}
E^{k}-1 & =(1+\Delta)^{k}-1=\sum_{i=1}^{k}\binom{k}{i} \Delta^{i}  \tag{6.14}\\
& =\left(\gamma_{k 0}+\gamma_{k 1} \Delta+\gamma_{k 2} \Delta^{2}+\ldots\right) h D \quad(k=1,2, \cdots),
\end{align*}
$$

where

$$
\begin{equation*}
\gamma_{k i}=\sum_{j=1}^{i}(-1)^{j+1} \frac{1}{j+1} \gamma_{k i-j}\binom{k}{i+1} \quad(i=1,2, \cdots), \tag{6.15}
\end{equation*}
$$

and

$$
\begin{equation*}
\binom{k}{l}=0 \text { for } l>k . \tag{6.18}
\end{equation*}
$$

Then by (6.13) and (6.14), for any function $z(x)$ which is $m+1$ times continuously differentiable on $\left[x_{0}, x_{0}+m h\right]$, it is valid that

$$
\begin{equation*}
z\left(x_{k}\right)-z\left(x_{0}\right)=h \sum_{i=0}^{m-1} \gamma_{k i} \Delta^{i} z^{\prime}\left(x_{0}\right)+O\left(h^{m+1}\right), \tag{6.19}
\end{equation*}
$$

$$
\begin{equation*}
\Delta^{l} z\left(x_{0}\right)=h \sum_{i=0}^{m-1} a_{i} \Delta^{i+l-1} z^{\prime}\left(x_{0}\right)+O\left(h^{m+1}\right) \quad(l=1,2, \ldots, m) \tag{6.20}
\end{equation*}
$$

Suppose that $n \geqq r+1$ and let $A$ be the matrix

$$
A=\left(\begin{array}{c}
a_{2}, a_{3}, \cdots \cdots \cdots, a_{l+1}  \tag{6.21}\\
a_{3}, a_{4}, \cdots \cdots \cdots, a_{l+2} \\
\bullet \\
\bullet \\
a_{l+1}, a_{l+2}, \cdots \cdots, \\
\cdot
\end{array}\right)
$$

where $l=n-r$. Then, from (6.20), we have a system of linear equations

$$
\left.A \left\lvert\, \begin{array}{c}
h \Delta^{n+1} z^{\prime}\left(x_{0}\right)  \tag{6.22}\\
\cdot \\
\cdot \\
h \Delta^{n+l} z^{\prime}\left(x_{0}\right)
\end{array}\right.\right)=\left(\begin{array}{c}
\Delta^{n} z\left(x_{0}\right)-h \sum_{i=0}^{1} a_{i} \Delta^{n-1+i} z^{\prime}\left(x_{0}\right)+O\left(h^{m+1}\right) \\
\cdot \\
\Delta^{r+1} z\left(x_{0}\right)-h \sum_{i=0}^{l} a_{i} \Delta^{r+i} z^{\prime}\left(x_{0}\right)+O\left(h^{m+1}\right)
\end{array}\right)
$$

To show that $A$ is non-singular, making use of (6.15) and (6.16), we rewrite $|A|$ as follows:

$$
|A|=\left|\begin{array}{l}
a_{0}, a_{1}, \cdots, a_{l}  \tag{6.23}\\
0, \\
\cdot \\
\cdot \\
\cdot \\
0,
\end{array}\right|=(-1)^{l}\left|\begin{array}{ccc}
\frac{1}{1+0}, \frac{1}{1+1}, \cdots \cdots \cdots \cdots, \frac{1}{1+l} \\
\frac{1}{2+0}, \frac{1}{2+1}, \cdots \cdots \cdots \cdots, \frac{1}{2+l} \\
\frac{1}{l+1+0}, \frac{1}{l+1+1}, \cdots \cdots, \frac{1}{l+1+l}
\end{array}\right| .
$$

Then, from the well-known identity due to Cauchy, it follows that

$$
\begin{equation*}
|A|=(-1)^{l} \frac{[1!2!\cdots l!]^{3}}{(l+1)!(l+2)!\cdots(2 l+1)!} \neq 0 \tag{6.24}
\end{equation*}
$$

Hence (6.22) can be solved and the solution can be written in the following form:

$$
\begin{gather*}
h \Delta^{n+j} z^{\prime}\left(x_{0}\right)=\sum_{k=1}^{n-r} \alpha_{j k} \Delta^{r+k} z\left(x_{0}\right)+h \sum_{k=r}^{n} \beta_{j k} \Delta^{k} z^{\prime}\left(x_{0}\right)+O\left(h^{m+1}\right)  \tag{6.25}\\
(j=1,2, \cdots, n-r) .
\end{gather*}
$$

Substituting this into (6.19), we can rewrite (6.19) as follows:

$$
\begin{equation*}
z\left(x_{k}\right)-z\left(x_{0}\right)=h \sum_{j=0}^{n} \mu_{k j} \Delta^{j} z^{\prime}\left(x_{0}\right)+\sum_{j=r+1}^{n} \nu_{k j} \Delta^{j} z\left(x_{0}\right)+O\left(h^{m+1}\right) . \tag{6.26}
\end{equation*}
$$

Since

$$
\begin{equation*}
\Delta^{j}=(E-1)^{j}=\sum_{l=0}^{j}(-1)^{j-l}\binom{j}{l} E^{l} \tag{6.27}
\end{equation*}
$$

if we put

$$
\begin{equation*}
\sum_{l=0}^{n}(-1)^{l-j} \mu_{k j}\binom{l}{j}=a_{k j} \tag{6.28}
\end{equation*}
$$

and

$$
\begin{equation*}
\sum_{l=r+1}^{n}(-1)^{l-j} \nu_{k j}\binom{l}{j}=b_{k j} \quad(j=0,1, \cdots, n) \tag{6.29}
\end{equation*}
$$

then it follows that

$$
\begin{gather*}
z\left(x_{k}\right)-z\left(x_{0}\right)=h \sum_{j=0}^{n} a_{k j} z^{\prime}\left(x_{j}\right)+\sum_{j=0}^{n} b_{k j} z\left(x_{j}\right)+O\left(h^{m+1}\right)  \tag{6.30}\\
\sum_{j=r+1}^{n} \nu_{k j} \Delta^{j} z\left(x_{0}\right)=\sum_{j=0}^{n} b_{k j} z\left(x_{j}\right) \tag{6.31}
\end{gather*}
$$

and

$$
\begin{equation*}
a_{0 j}=b_{0 j}=0 \quad(j=0,1, \cdots, n) \tag{6.32}
\end{equation*}
$$

because $\gamma_{0 i}=0(i=0,1,2, \ldots, m-1)$.
It is readily seen from (6.19) that (6.30), (6.31) and (6.32) are valid also for the case $n=r$ with $\nu_{k j}=b_{k j}=0$ because $m-1=n$.

Since

$$
\begin{equation*}
z\left(x_{j}\right)=\sum_{l=0}^{m} j^{l} \frac{1}{l!} h^{l} z^{(l)}\left(x_{0}\right)+O\left(h^{m+1}\right) \tag{6.33}
\end{equation*}
$$

$$
\begin{equation*}
z^{\prime}\left(x_{j}\right)=\sum_{l=1}^{m} l_{j}^{l-1} \frac{1}{l!} h^{l-1} z^{(l)}\left(x_{0}\right)+O\left(h^{m+1}\right) \tag{6.34}
\end{equation*}
$$

$$
\begin{equation*}
\Delta^{j} z\left(x_{0}\right)=O\left(h^{j}\right) \tag{6.35}
\end{equation*}
$$

and $z(x)$ is any function smooth enough, from (6.30) and (6.31) follow (2.33) and (2.24). This proves lemma 2.

Lemma 3 follows readily from (6.30), (6.31) and (6.35).

### 6.3 Proof of lemma 4

From (2.7), (2.6), (2.18) and (2.19) it follows that

$$
\begin{equation*}
y\left(x_{k}\right)-y\left(x_{0}\right)=y_{k}-y_{0}-d_{k} \tag{6.36}
\end{equation*}
$$

$$
\begin{equation*}
\sum_{j=0}^{n} b_{k j} y\left(x_{j}\right)=\sum_{j=0}^{n} b_{k j} y_{j}-\sum_{j=0}^{n} b_{k j} d_{j} \tag{6.37}
\end{equation*}
$$

$$
\begin{equation*}
\sum_{j=0}^{n} b_{k j} y_{j}=y_{0} \sum_{j=0}^{n} b_{k j}+h \sum_{j=1}^{n} b_{k j} \sum_{q=1}^{j} p_{q}, \tag{6.38}
\end{equation*}
$$

and

$$
\begin{equation*}
\sum_{j=0}^{n} b_{k j} d_{j}=h^{p}\left[\sum_{i=1}^{m} h^{i} \sum_{j=0}^{n} b_{k j} P_{i}(j)+O\left(h^{m+1}\right)\right] . \tag{6.39}
\end{equation*}
$$

By (2.24), further (6.38) and (6.39) can be written as follows:

$$
\begin{equation*}
\sum_{j=0}^{n} b_{k j} y_{j}=h \sum_{j=1}^{n} b_{k j} \sum_{q=1}^{j} p_{q} \tag{6.40}
\end{equation*}
$$

and

$$
\begin{equation*}
\sum_{j=0}^{n} b_{k j} d_{j}=h^{p} \sum_{i=r+1}^{m} h^{i} \sum_{j=0}^{n} b_{k j} P_{i}(j)+O\left(h^{p+m+1}\right) \tag{6.41}
\end{equation*}
$$

From (6.10) and (2.20) it follows that

$$
\begin{equation*}
y^{\prime}\left(x_{j}\right)=f\left(x_{j}, y\left(x_{j}\right)\right)=f_{j}-h^{p} \sum_{i=1}^{m} h^{i} Q_{i}(j)+O\left(h^{p+m+1}\right) . \tag{6.42}
\end{equation*}
$$

Substituting (6.36), (6.42), (6.37), (6.40) and (6.41) into (2.25), we have

$$
\begin{align*}
y_{k}-y_{0}-d_{k}= & h \sum_{j=0}^{n} a_{k j} f_{j}+h \sum_{j=1}^{n} b_{k j} \sum_{q=1}^{j} p_{q}-  \tag{6.43}\\
& -h^{p}\left[\sum_{i=1}^{m} h^{i+1} \sum_{j=0}^{n} a_{k j} Q_{i}(j)+\sum_{i=r+1}^{m} b_{k j} P_{i}(j)\right]+O\left(h^{m+1}\right) .
\end{align*}
$$

Since by (2.23)

$$
\begin{equation*}
\sum_{j=0}^{n} j^{l-1} a_{k j}=\frac{1}{l} k^{l}-\frac{1}{l} \sum_{j=0}^{n} b_{k j} j^{l} \quad(l=1,2, \ldots, m) \tag{6.44}
\end{equation*}
$$

from (2.31) it follows that

$$
\begin{equation*}
\sum_{j=0}^{n} a_{k j} Q_{i}(j)=R_{i+1}(k)-\sum_{j=0}^{n} b_{k j} R_{i+1}(j) . \tag{6.45}
\end{equation*}
$$

Further by (2.24) it holds that

$$
\begin{equation*}
\sum_{j=0}^{n} b_{k j} R_{i+1}(j)=0 \quad(i=0,1, \cdots, r-1) \tag{6.46}
\end{equation*}
$$

Substituting (6.45) and (6.46) into (6.43) and making use of (2.27) and (2.28), we have (2.30), where

$$
\begin{equation*}
c_{k i}=\sum_{j=0}^{n} b_{k j}\left[P_{i}(j)-R_{i}(j)\right] . \tag{6.47}
\end{equation*}
$$

Clearly (2.33) holds and (2.32) follows from (6.32). Thus lemma 4 has been proved.

### 6.4 Proof of lemma 5

Since by (2.14) and (3.11)

$$
\begin{align*}
F(x, v(x), S(x)+g(x)) & =F(x, v(x), d(x))  \tag{6.48}\\
& =f(x, y(x)+d(x))-f(x, y(x))
\end{align*}
$$

by the same reasoning as for (6.10), it is easily seen that

$$
\begin{equation*}
F(x, v(x), S(x)+g(x))=h^{p}\left[\sum_{i=1}^{m} h^{i} Q_{i}(u)+O\left(h^{m+1}\right)\right] . \tag{6.49}
\end{equation*}
$$

Differentiating (3.8), we have

$$
\begin{equation*}
g^{\prime}(x)=h^{p}\left[\sum_{i=1}^{m} h^{i} Q_{i}(u)+\sum_{i=r+1}^{m} h^{i-1} c_{i}^{\prime}(u)\right]+h^{m} b^{\prime}(u) . \tag{6.50}
\end{equation*}
$$

Since $m \geqq p+r$ by (3.4), by our assumptions there exists a constant $K$ such that

$$
\begin{equation*}
\left|g^{\prime}(x)-F(x, v(x), S(x)+g(x))\right| \leqq K h^{p+r} \quad(0 \leqq u \leqq n) \tag{6.51}
\end{equation*}
$$

Let $L$ be a constant such that the inequality

$$
\begin{equation*}
\left|f_{y}(x, y)\right| \leqq L \tag{6.52}
\end{equation*}
$$

is valid for any point $(x, y)$ belonging to $S$, then by (2.14) the inequality

$$
\begin{equation*}
\left|F\left(x, v(x), S(x)+w_{1}\right)-F\left(x, v(x), S(x)+w_{2}\right)\right| \leqq L\left|w_{1}-w_{2}\right| \tag{6.53}
\end{equation*}
$$

holds for any numbers $w_{1}$ and $w_{2}$.
Since $g\left(x_{0}\right)=0$, if we put

$$
\begin{equation*}
w(x)=w(x ; 0) \tag{6.54}
\end{equation*}
$$

from the well-known theorem in the theory of ordinary differential equations, it follows that

$$
\begin{equation*}
|g(x)-w(x)| \leq h^{p+r} \frac{K}{L}\left(e^{L u h}-1\right) \quad(0 \leq u \leq n) \tag{6.55}
\end{equation*}
$$

so that

$$
\begin{equation*}
g(x)=w(x)+O\left(h^{p+r+1}\right) \tag{6.56}
\end{equation*}
$$

Therefore we put

$$
\begin{equation*}
w(x)=h^{p} \sum_{i=1}^{m} h^{i+1} R_{i+1}(u)+h^{p+r+1} z(u)+h^{p+r+2} z(u, h), \tag{6.57}
\end{equation*}
$$

where

$$
\begin{equation*}
z(0)=z(0, h)=0 . \tag{6.58}
\end{equation*}
$$

Then we have

$$
\begin{equation*}
w^{\prime}(x)=h^{p} \sum_{i=1}^{m} h^{i} Q_{i}(u)+h^{p+r} z^{\prime}(u)+h^{p+r+1} z_{u}(u, h) . \tag{6.59}
\end{equation*}
$$

On the other hand, since by (3.12), (3.8) and (6.57)

$$
\begin{equation*}
S(x)+w(x)=d(x)+h^{p+r+1}\left[z(u)-c_{r+1}(u)\right]+O\left(h^{\rho+r+2}\right), \tag{6.60}
\end{equation*}
$$

it follows that

$$
\begin{align*}
& F(x, v(x), S(x)+w(x))= f(x, y(x)+d(x))-f(x, y(x))+  \tag{6.61}\\
&+f(x, y(x))-f(x, y(x)+d(x)-S(x)-w(x)) \\
&=h^{p} \sum_{i=1}^{m} h^{i} Q_{i}(u)+h^{p+r+1} f_{y}\left(x_{0}, y_{0}\right)\left[z(u)-c_{r+1}(u)\right]+O\left(h^{p+r+2}\right) .
\end{align*}
$$

Comparing (6.61) with (6.59), we find that

$$
\begin{equation*}
z(u)=0 \tag{662}
\end{equation*}
$$

so that

$$
\begin{equation*}
w(x)=h^{p} \sum_{i=1}^{m} h^{i+1} R_{i+1}(u)+O\left(h^{p+r+2}\right), \tag{6.63}
\end{equation*}
$$

and

$$
\begin{equation*}
d(x)=S(x)+w(x)+h^{p+r+1} c_{r+1}(u)+O\left(h^{p+r+2}\right) . \tag{6.64}
\end{equation*}
$$

Now, from (2.11) and (3.14), it follows that

$$
\begin{align*}
w^{\prime}(x)+c^{\prime}(x)= & F(x, v(x), S(x)+w(x))+F(x, y(x), c(x))  \tag{6.65}\\
= & F(x, v(x), S(x)+w(x)+c(x))+F(x, y(x), S(x)+w(x)-d(x)) \\
& -F(x, y(x)-c(x), S(x)+w(x)-d(x)) \\
= & F(x, v(x), S(x)+w(x)+c(x))+O\left(h^{p+r+1}\right)
\end{align*}
$$

Since by (2.12), (6.54) and (3.15)

$$
\begin{equation*}
w\left(x_{0}\right)+c\left(x_{0}\right)=e, \tag{6.66}
\end{equation*}
$$

by the same reasoning as for $g(x)$ and $w(x)$, it is seen that

$$
\begin{equation*}
w(x)+c(x)=w(x ; e)+O\left(h^{p+r+2}\right) . \tag{6.67}
\end{equation*}
$$

From (6.67), (6.64), (2.10) and (3.11), it follows that

$$
\begin{equation*}
v(x)-y(x ; e)=S(x)+w(x ; e)+h^{p+r+1} c_{r+1}(u)+O\left(h^{p+r+2}\right) . \tag{6.68}
\end{equation*}
$$

Then, by (3.13), (3.16) holds. This completes the proof of lemma 5.

### 6.5 Proof of theorem 2

From (3.7) and (3.8) we have

$$
\begin{equation*}
h^{k} \frac{d^{k}}{d x^{k}} e(x)=\sum_{i=1}^{m} h^{i} P_{i}^{(k)}(u)+h^{m+1} a^{(k)}(u) \tag{6.69}
\end{equation*}
$$

and

$$
\begin{gather*}
h^{k} \frac{d^{k}}{d x^{k}} g(x)=h^{p}\left[\sum_{i=1}^{m} h^{i+1} R_{i+1}^{(k)}(u)+\sum_{i=r+1}^{m} h^{i} c_{i}^{(k)}(u)\right]+h^{m+1} b^{(k)}(u)  \tag{6.70}\\
(k=1,2, \cdots) .
\end{gather*}
$$

Since $a(u), c_{i}(u)$ and $b(u)$ are polynomials of degree at most $n, P_{i}(u)$ and $R_{i}(u)$ are polynomials of degree at most $i$ and $m \geqq n+1$ by (3.5) and (2.15), from
(6.69) and (6.70) it follows that

$$
\begin{equation*}
h^{k} \frac{d^{k}}{d x^{k}} e(x)=O\left(h^{k}\right) \tag{6.71}
\end{equation*}
$$

and

$$
h^{k} \frac{d^{k}}{d x^{k}} g(x)= \begin{cases}O\left(h^{p+k}\right), & 0 \leqq k \leqq r+1 \quad \text { or } \quad k \geqq n+1,  \tag{6.72}\\ O\left(h^{p+r+1}\right), & r+1<k \leqq n,\end{cases}
$$

so that by (3.10) and (3.12) we have

$$
\begin{equation*}
h^{k} \frac{d^{k}}{d x^{k}} d(x)=O\left(h^{p+k}\right) \tag{6.73}
\end{equation*}
$$

and

$$
\begin{equation*}
h^{k} \frac{d^{k}}{d x^{k}} S(x)=O\left(h^{p+\alpha_{k}}\right) \tag{6.74}
\end{equation*}
$$

where

$$
\begin{equation*}
\alpha_{k}=\min (k, r+1) \tag{6.75}
\end{equation*}
$$

Further by (3.11) we have

$$
\begin{equation*}
\frac{d^{k}}{d x^{k}} v(x)=O(1) \tag{6.76}
\end{equation*}
$$

and

$$
\begin{equation*}
\frac{d^{k}}{d x^{k}} S(x)=O(1) \quad(k=0,1, \ldots) \tag{6.77}
\end{equation*}
$$

because $p \geqq 1$ by (3.5) and $p+r+1-n \geqq 0$ by (3.22). Also by (2.13), (6.60) and (6.67) it is seen that

$$
\begin{equation*}
S(x)+w(x ; e)=e \cdot O(1)+O\left(h^{p+1}\right) . \tag{6.78}
\end{equation*}
$$

Now, to determine the order of the truncation errors of $w_{i}(e)(i=l, 2 l$, $\cdots, t l)$, we shall show by induction that

$$
\begin{equation*}
h^{k+1} \frac{d^{k+1}}{d x^{k+1}} w(x ; e)=e \cdot O\left(h^{k+1}\right)+O\left(h^{p+\alpha_{k}+1}\right)^{5} \quad(k=0,1, \cdots, p+r+1) . \tag{6.79}
\end{equation*}
$$

By definition $w(x ; e)$ satisfies the equation

[^2]\[

$$
\begin{align*}
w^{\prime}(x ; e) & =F(x, v(x), S(x)+w(x ; e))  \tag{6.80}\\
& =f(x, v(x))-f(x, v(x)-S(x)-w(x ; e))
\end{align*}
$$
\]

From this we have by (6.78)

$$
\begin{align*}
h \frac{d}{d x} w(x ; e) & =[S(x)+w(x ; e)] \cdot O(1)  \tag{6.81}\\
& =e \cdot O(h)+O\left(h^{p+2}\right)
\end{align*}
$$

because $f_{y}(x, y)$ is bounded in $S$. Thus (6.79) is valid for $k=0$. Hence suppose that (6.79) holds for $k=0,1, \cdots, j-1(j \leqq p+r+1)$. Then we have

$$
\begin{equation*}
\frac{d^{k+1}}{d x^{k+1}} w(x ; e)=e \cdot O(1)+O\left(h^{p+\alpha_{k}-k}\right)=O(1) \quad(k=0,1, \ldots, j-1), \tag{6.82}
\end{equation*}
$$

and so

$$
\begin{equation*}
S^{(i)}(x)+w^{(i)}(x ; e)=O(1) \quad(i=0,1, \cdots, j) \tag{6.83}
\end{equation*}
$$

because $p+\alpha_{k}-k \geqq 0$ provided $k \leqq p+r+1$, where $f^{(i)}(x)$ denotes $\frac{d^{i}}{d x^{i}} f(x)$. Further, from the assumption, it follows that

$$
\begin{align*}
& h^{j+1}\left[S^{(i)}(x)+w^{(i)}(x ; e)\right]  \tag{6.84}\\
& =h^{j+1-i}\left[h^{i} S^{(i)}(x)+h^{i} w^{(i)}(x ; e)\right] \\
& =h^{j+1-i}\left[O\left(h^{p+\alpha_{i}}\right)+e \cdot O\left(h^{i}\right)+O\left(h^{p+\alpha_{i-1}+1}\right)\right] \\
& =e \cdot O\left(h^{j+1}\right)+O\left(h^{p+\alpha_{j}+1}\right) \quad(i=0,1, \cdots, j),
\end{align*}
$$

because (6.84) is evidently valid by (6.78) for $i=0$, and

$$
j+1-i+p+\alpha_{i}= \begin{cases}p+1+r+1+(j-i) & (i \geqq r+1)  \tag{6.85}\\ p+1+j & (i<r+1)\end{cases}
$$

and
(6.86) $j+1-i+p+\alpha_{i-1}+1= \begin{cases}p+1+r+1+(j+1-i) & (i-1 \geqq r+1) \\ p+1+j & (i-1<r+1)\end{cases}$
for $i \geqq 1$.
Differentiating (6.80) $j$ times and multiplying it by $h^{j+1}$, we see that $h^{j+1} w^{(j+1)}(x ; e)$ can be expressed as a linear combination of the terms of the
form as follows:

$$
\begin{align*}
& U\left(\mu_{1}, \mu_{2}, \ldots, \mu_{j} ; \nu_{1}, \nu_{2}, \ldots, \nu_{j} ; \alpha, \beta\right)=h^{j+1}\left[\prod_{i=1}^{i}\left(v^{\left(\mu_{i}\right)}(x)\right)^{\nu} \frac{\partial^{\alpha+\beta}}{\partial x^{\alpha} \partial y^{\beta}} f(x, v(x))-\right.  \tag{6.87}\\
& \left.\quad-\prod_{i=1}^{j}\left(v^{\left(\mu_{i}\right)}(x)-S^{\left(\mu_{i}\right)}(x)-w^{\left(\mu_{i}\right)}(x ; e)\right)^{\nu_{i}} \frac{\partial^{\alpha+\beta}}{\alpha x^{\alpha} \partial y^{\beta}} f(x, v(x)-S(x)-w(x ; e))\right]
\end{align*}
$$

where $\mu_{i}, \nu_{i}(i=1,2, \cdots, j), \alpha$, and $\beta$ are non-negative integers not greater than $j$ such that

$$
\begin{equation*}
\alpha+\beta \leqq j, \quad \sum_{i=1}^{j} \mu_{i} \cdot \nu_{i} \leqq j \tag{6.88}
\end{equation*}
$$

Making use of (6.76), 6.83) and (6.84), we can rewrite $U$ as follows:

$$
\begin{gather*}
\text { 9) } U\left(\mu_{1}, \mu_{2}, \ldots, \mu_{j} ; \nu_{1}, \nu_{2}, \ldots, \nu_{j} ; \alpha, \beta\right)=h^{j+1} \prod_{i=1}^{j}\left(v^{\left(\mu_{i}\right)}(x)\right)^{\nu_{i}}\left[\frac{\partial^{\alpha+\beta}}{\partial x^{\alpha} \partial y^{\beta}} f(x, v(x))-\right.  \tag{6.89}\\
\left.-\frac{\partial^{\alpha+\beta}}{\partial x^{\alpha} \partial y^{\beta}} f(x, v(x)-S(x)-w(x ; e))\right]+h^{j+1}\left[\prod_{i=1}^{j}\left(v^{\left(\mu_{i}\right)}(x)\right)^{\nu_{i}}-\right. \\
\left.-\prod_{i=1}^{j}\left(v^{\left(\mu_{i}\right)}(x)-S^{\left(\mu_{i}\right)}(x)-w^{\left(\mu_{i}\right)}(x ; e)\right)^{\nu_{i}}\right] \frac{\partial^{\alpha+\beta}}{\partial x^{\alpha} \partial y^{\beta}} f(x, v(x)-S(x)-w(x ; e)) \\
=e \cdot O\left(h^{j+1}\right)+O\left(h^{p+\alpha_{j}+1}\right)
\end{gather*}
$$

because the partial derivatives of $f(x, y)$ are bounded by the the assumption. Hence (6.79) is valid also for $k=j$.

From (6.79) it follows that

$$
\begin{equation*}
w\left(x_{i} ; e\right)=w_{i}(e)+e \cdot O\left(h^{q+1}\right)+O\left(h^{p+\alpha_{q}+1}\right) \quad(i=l, 2 l, \cdots, t l), \tag{6.90}
\end{equation*}
$$

because a one-step method of order $q(q \leqq p+r+1)$ is applied for approximating $w(x ; \mathrm{e})$. Substitution of (6.90) into (3.16) yields (3.23). Thus theorem 2 has been proved.

Remark 3. When $m=p+r$, replacing $c_{r+1}(u)$ by $c_{r+1}(u)+b(u)$ in (6.60), (6.61), (6.64) and (6.68), and modifying (3.16) as

$$
\begin{equation*}
y_{k}-y\left(x_{k} ; e\right)=S_{k}+w\left(x_{k} ; e\right)+O\left(h^{p+r+1}\right) \quad(k=0,1, \ldots, n), \tag{3.16}
\end{equation*}
$$

we have the following
Theorem $2^{\prime}$. Under the conditions (3.22) and

$$
\begin{equation*}
2(n-r)+1=p \geqq 1 \tag{3.5}
\end{equation*}
$$

it is valid that

$$
\begin{equation*}
y_{j}-y\left(x_{j} ; e\right)=T_{j}(e)+e \cdot O\left(h^{q+1}\right)+O\left(h^{p+\alpha+1}\right) \quad(j=l, 2 l, \cdots, t l) \tag{3.23}
\end{equation*}
$$

where $T_{j}(e)$ is defined by (3.24) and

$$
\begin{equation*}
\alpha=\min (q, r) . \tag{3.25}
\end{equation*}
$$

## References

[1] Bieberbach, L.: On the remainder of the Runge-Kutta formula in the theory of ordinary differential equations, Z. Angew. Math. Phys., 2 (1951), 233-248.
[2] Carr, J. W., III: Error bounds for the Runge-Kutta single-step integration process, J. Assoc. Comput. Mach., 5 (1958), 39-44.
[3] Ceschino, F.: Critère d'utilisation du procédé de Runge-Kutta, C. R. Acad. Sci. Paris, 238 (1954), 986-988; 1553-1555.
[4] Collatz, L.: The numerical treatment of differential equations, Berlin (1960).
[5] Galler, B. A., and D.P. Rozenberg: A generalization of a theorem of Carr on error bounds for RungeKutta procedures, J. Assoc. Comput. Mach., 7 (1960), 57-60.
[6] Henrici, P.: Discrete variable methods in ordinary differential equations, New York (1962).
[7] Hildebrand, F. B.: Introduction to numerical analysis, New York (1956).
[8] Kurmit, A.: Adaptation of an error estimate for difference methods of numerical integration to RungeKutta's method, Latvijas PSR Zinātnu Akad. Vēstis, 191 (1963), 49-58; 192 (1963), 75-84.
[9] Kurmit, A.: Error theory of Runge-Kutta method, Vestnik Leningrad. Univ., 19 (1963), 35-41.
[10] Lotkin, M: On the accuracy of Runge-Kutta's method, Math. Tables Aids Comput., 5 (1951), 128133.
[11] Schechter, E.: Sur l'erreur du procédé d'intégration numérique de Runge-Kutta, Acad. R. P. Romîne Fil. Cluj. Stud. Cerc. Mat. Fiz., 8 (1957), 115-124.
[12] Schechter, E.: De la délimitation des erreurs dans certains procédés d'intégration numérique des équations différentielles, Acad. R. P. Romîne Fil. Cluj. Stud. Cerc. Mat. Fiz., 9 (1958), 343-350.
[13] Schröder, J.: Fehlerabschätzungen mit Rechenanlagen bei gewöhnlichen Differentialgleichungen erster Ordnung, Numer. Math., 3 (1961), 39-61.
[14] Schröder, J.: Verbesserung einer Fehlerabschätzung für gewöhnliche Differentialgleichungen erster Ordnung, Numer. Math., 3 (1961), 125-130.
[15] Tihonov, A. N. and A. D. Gorbunov: Asymptotic error bounds for the Runge-Kutta method, Ž. Vyčisl. Mat. i. Mat. Fiz., 3 (1963), 195-197.
[16] Urabe, M.: Theory of errors in numerical integration of ordinary differential equations, J. Sci. Hiroshima Univ., Ser. A-I, 25 (1961), 3-62.

Department of Mathematics<br>Faculty of Science<br>Hiroshima University


[^0]:    1) Numbers in square brackets refer to the references listed at the end of this paper.
[^1]:    2) The assumptions stated here are too severe for practical purposes but they are relaxed at the end of paragraph 3.
    3) Here it is assumed that $s \geqq p$.
[^2]:    5) Here it is assumed that $s \geqq k+1$.
