Journal of Science of the Hiroshima University, Ser. A, Vol. 19, No. 2, November, 1955

A Method of Numerical Integration of Analytic Differential Equations

By

Minoru URABE and Shigetoshi MISE

(Received June 6, 1955)

1. Introduction

For the truncation error committed in the use of approximate formulas, P. Davis [1] proposed a new estimate by introducing a Hilbert space of analytic functions and using essentially the Riesz representation of bounded linear functionals. In the present note, we seek for the formulas of numerical integration of analytic differential equations such that their truncation errors may be least in the error estimate of Davis, and compare them with the traditional formulas.

2. Derivation of new formulas

We consider the formula of numerical integration of the form

$$\int_{x_0}^{x_1} f(x)\,dx = h\sum_{j=-1}^{N} a_j f(x_{-j}), \tag{1}$$

where $x_{-j} \doteqdot x_0 - jh$ $(h>0)$. We assume that the complex extension $f(z)$ of $f(x)$ is regular and single-valued in $|z-x_0|<\rho$ and belongs to the class H^2 there. That is to say, $f(z)$ has the Taylor expansion

$$f(z) = \sum_{n=0}^{\infty} c'_n (z-x_0)^n \qquad (|z-x_0|<\rho) \tag{2}$$

and it is valid that

$$\|f\|^2 = \int_0^{2\pi} f\bar{f}\,d\theta = 2\pi\sum_{n=0}^{\infty} |c'_n\rho^n|^2 < \infty, \tag{3}$$

θ being such that $z-x_0=\rho\exp(i\theta)$. We take h so that $|x_{-j}-x_0|<\rho$ for $j=$

1) P. Davis, *Errors of numerical approximation for analytic functions*, J. Rational Mech. Anal. **2**, 303-313 (1953).

$-1, 0, 1, 2, \cdots, N$. Put $x = x_0 + \rho u$, $x_{-j} = x_0 + \rho u_{-j}$, $h = \rho h_0$ and $f(x_0 + \rho u) = \phi(u)$. Let the error committed by the formula (1) be

$$(4) \qquad E(f) = \int_{x_0}^{x_1} f(x)\, dx - h \sum_{j=-1}^{N} a_j f(x_{-j}),$$

then

$$E(f) = \rho\, E_0(\phi),$$

where

$$(5) \qquad E_0(\phi) = \int_{u_0}^{u_1} \phi(u)\, du - h_0 \sum_{j=-1}^{N} a_j \phi(u_{-j}).$$

From (2) and (3), $\phi(u)$ has the Taylor expansion

$$(6) \qquad \phi(u) = \sum_{n=0}^{\infty} c_n u^n \qquad (c_n = c'_n \rho^n, \quad |u| < 1)$$

and it is valid that

$$(7) \qquad \|\phi\|^2 = \|f\|^2 = 2\pi \sum_{n=0}^{\infty} |c_n|^2 < \infty.$$

From $|x_{-j} - x_0| < \rho$, $|u_{-j}| = |-jh_0| < 1$ for $j = -1, 0, 1, 2, \cdots N$. From (6) and (7), for $|u| < 1$, it is valid that

$$|\phi(u)|^2 \leqq \sum_{n=0}^{\infty} |c_n|^2 \cdot \sum_{n=0}^{\infty} |u^n|^2 = \frac{1}{2\pi} \frac{1}{1-|u|^2} \|\phi\|^2.$$

From this follows the boundedness of the linear functionals E and E_0. Then, from (6), it follows that

$$(8) \qquad E_0(\phi) = \sum_{n=0}^{\infty} c_n E_0(u^n),$$

consequently, if $\sum_{n=0}^{\infty} |E_0(u^n)|^2$ converges, it is valid that

$$(9) \qquad |E_0(\phi)|^2 \leqq \sum_{n=0}^{\infty} |c_n|^2 \cdot \sum_{n=0}^{\infty} |E_0(u^n)|^2 = \sigma_{E_0}^2 \|\phi\|^2,$$

where

$$(10) \qquad 2\pi\, \sigma_{E_0}^2 = \sum_{n=0}^{\infty} |E_0(u^n)|^2.$$

Now, from (5),

(11) $$E_0(u^n)=\left(\frac{1}{n+1}-(-1)^n\sum_{j=-1}^{N}a_j\,j^n\right)h_0^{n+1}.$$

Since $|jh_0|<1$, from this readily follows the convergence of $\sum_{n=0}^{\infty}|E_0(u^n)|^2$. The inequality (9) is nothing but the estimate proposed by P. Davis.[1)]

If, in the formula (1), we consider the coefficients a_j's as the parameters, $\sum_{n=0}^{\infty}|E_0(u^n)|^2$ becomes a function of a_j's. Let this function be $\sum(a)$, then, from (11), $\sum(a)$ can be written as follows :

(12) $$\sum(a)=h_0^2\left[\sum_{j,k=-1}^{N}g_{jk}\,a_j\,a_k-2\sum_{k=-1}^{N}g_k\,a_k+g\right],$$

where

(13) $$\begin{cases} g_{jk}=g_{kj}=\displaystyle\sum_{n=0}^{\infty}(j\,k\,h_0^2)^n=\frac{1}{1-j\,k\,h_0^2}, \\ g_k=\displaystyle\sum_{n=0}^{\infty}(-1)^n\frac{(k\,h_0^2)^n}{n+1}=\frac{1}{k\,h_0^2}\log(1+k\,h_0^2). \end{cases}$$

Since $\sum_{j,k}g_{jk}a_ja_k=\sum_{n=0}^{\infty}(\sum_j a_j j^n h_0^n)^2\geqq 0$, $\|g_{jk}\|$ is non-negative semi-definite. If $\sum_{j,k}g_{jk}a_ja_k=0$, it must be that $\sum_j a_j j^n=0$ for $n=0, 1, 2, \cdots$, from which follows that $a_j=0$. Thus we see that $\|g_{jk}\|$ is positive definite. Then $\sum(a)$ becomes least when and only when a_j satisfy the conditions

(14) $$\frac{\partial\sum}{\partial a_k}=2h_0^2\left[\sum_{j=-1}^{N}g_{kj}\,a_j-g_k\right]=0, \qquad (k=-1, 0, 1, \cdots, N)$$

and a_j's satisfying these conditions are uniquely determined. Then, from (9), it is seen that this unique solution a_j of (14) provides a most accurate formula of the form (1) in the error estimate of P. Davis. In the sequel, we shall call this newly obtained formulas simply the new formulas.

3. Comparison with traditional formulas

The traditional formulas of the form (1) are such that (1) holds exactly for a polynomial of the degree as high as possible. Consequently, if we write their coefficients as $a_j^{(0)}$, then they satisfy

(15) $$h_0^{-(n+1)}E_0(u^n)=0, \qquad (n=0, 1, 2, \cdots, M)$$

1) P. Davis, ibid.

where $M=N+1$ or N accoding as the formula (1) represents a formula of interpolation or extrapolation, namely $a_{-1}\neq 0$ or $=0$. For such $a_j=a_j^{(0)}$, from (8),

$$E_0(\phi)=\sum_{n=K}^{\infty} c_n E_0(u^n), \qquad (K=M+1)$$

consequently

$$|E_0(\phi)|^2 \leqq \sum_{n=K}^{\infty}|c_n|^2 \cdot \sum_{n=K}^{\infty}|E_0(u^n)|^2.^{1)}$$

Hence the traditional formulas are in general inferior in accuracy to the new ones when

$$\frac{\sum_{n=K}^{\infty}|E_0(u^n)|^2}{\sum} > \frac{\sum_{n=0}^{\infty}|c_n|^2}{\sum_{n=K}^{\infty}|c_n|^2} = 1+\frac{\sum_{n=0}^{M}|c_n|^2}{\sum_{n=K}^{\infty}|c_n|^2},$$

where $\sum=\sum(a)$ for a_j satisfying (14). Let us write the above conditions in the form

$$(16) \qquad \sum_{n=0}^{M}|c_n|^2 < \lambda \sum_{n=K}^{\infty}|c_n|^2,$$

where

$$(17) \qquad \lambda=\frac{\sum_{n=K}^{\infty}|E_0(u^n)|^2}{\sum}-1=\frac{\sum(a^{(0)})}{\sum(a)}-1.$$

Since $\lambda>0$ from the definition of $\sum$, the condition (16) is valid when

$$\sum_{n=0}^{M}|c_n|^2 \Big/ \sum_{n=K}^{\infty}|c_n|^2 \ll 1.$$

This condition means that the function $f(x)=\phi(u)$ behaves governed mostly by the terms of the orders higher than M, in other words, that $\nabla^M f(x)$ [2] loses constancy and reveals remarkable changes. Thus we see that, *when $\nabla^M f(x)$ reveals remarkable changes, the new formulas are more accurate than the traditional ones.*

4. Deviation from traditional formulas

Assuming that $h_0=h/\rho\ll 1$, let us seek for the deviations of a_j's of the new formulas from those of the traditional ones.

1) This is not an estimate of Davis but is a sharper one of Davis' form, because the right-hand side is less than $\sigma_{E_0^2}\|\phi\|^2$.

2) ∇ denotes a backward difference operator with respect to the intervals of breadth h.

Introducing the parameters r and $\mathcal{E}$, let us consider the functions

(18)
$$\begin{cases} g_{kj}^{(0)}(r) = \sum_{n=0}^{M} (kjr)^n, \\ g_k^{(m)}(r) = (-1)^{mK} k^{mK} \sum_{n=0}^{M} (-1)^n \dfrac{(kr)^n}{mK+n+1}, \end{cases}$$

and the series

(19)
$$\begin{cases} g_{kj}(r, \mathcal{E}) = g_{kj}^{(0)}(r) \sum_{m=0}^{\infty} (kj\mathcal{E})^{mK} = \dfrac{g_{kj}^{(0)}(r)}{1-(kj\mathcal{E})^K} \\ g_k(r, \mathcal{E}) = \sum_{m=0}^{\infty} g_k^{(m)}(r)\, \mathcal{E}^{mK}. \end{cases}$$

We assume that $|r|$, $|\mathcal{E}| \ll 1$. It is evident that

(20)
$$g_{kj}(h_0^2, h_0^2) = g_{kj}, \quad g_k(h_0^2, h_0^2) = g_k.$$

Since $|kr| < 1$, $|g_k^{(m)}(r)| < K \cdot \dfrac{k^{mK}}{mK+1}$, consequently the series $g_k(r, \mathcal{E})$'s are convergent, because $|k\mathcal{E}| < 1$. From $|kj\mathcal{E}| < 1$ readily follows the convergence of $g_{kj}(r, \mathcal{E})$.

In $G = \|g_{ij}^{(0)}(r)\|$, let the minor determinant complementary to $g_{ij}^{(0)}(r)$ be $G_{ij}(r)$. Then it holds that

(21)
$$\begin{cases} |G| = \text{det.}\, |G| = r^{M(M+1)/2} |\Delta|^2, \\ G_{ij} = r^{M(M+1)/2} \sum_{p=0}^{M} r^{-p} \Delta_{pi} \Delta_{pj}, \end{cases}$$

where $\Delta = \|j^n\| = \sum_j \oplus \begin{pmatrix} j^0 = 1 \\ j^1 \\ j^2 \\ \cdots \\ j^M \end{pmatrix}$ [1] and Δ_{pi} is a minor determinant complementary i^p in Δ. For, by the definition, $g_{ij}^{(0)}(r) = \sum_{n=0}^{M} (ijr)^n = \sum_{n=0}^{M} i^n \cdot (jr)^n$, consequently

$$G = \|i^n\| \cdot \|(jr)^n\|.$$

Since det. $\|(jr)^n\| = r^{M(M+1)/2}$ det. $|j^n|$, the first of (21) is valid. In the same way, the matrix composed of the elements of G_{ij} can be decomposed like G in a product of two rectangular matries, of which the former is a matrix lacking one row of $\|i^n\|$ and the latter is one lacking one column of $\|(jr)^n\|$. From this decomposition, the second of (21) follows immediately by the generalized multipli-

1) $\sum_j \oplus$ denotes a matrix whose columns are of the forms written behind the symbol. Here j runs over $-1, 0, 1, 2, \cdots, N$ or $0, 1, \cdots, N$ according as $M = N+1$ or N.

cation rule of determinants.

The values of $|\Delta|$ and Δ_{pi} are easily calculated by means of the well-known formula

$$\begin{vmatrix} 1 & 1 & \cdots\cdots\cdots & 1 \\ x_1 & x_2 & \cdots\cdots\cdots & x_L \\ x_1{}^2 & x_2{}^2 & \cdots\cdots\cdots & x_L{}^2 \\ \vdots & \vdots & & \vdots \\ x_1^{p-1} & x_2^{p-1} & \cdots\cdots\cdots & x_L^{p-1} \\ x_1^{p+1} & x_2^{p+1} & \cdots\cdots\cdots & x_L^{p+1} \\ \vdots & \vdots & & \vdots \\ x_1^M & x_2^M & \cdots\cdots\cdots & x_L{}^M \end{vmatrix} = A\, S_{L-p}, \tag{22}$$

where $L=K$ or M according as $p>M$ or $p\leqq M$ and

$$\begin{aligned} A = (x_L - x_{L-1})\,(x_L - x_{L-2}) \cdots\cdots (x_L - x_1) \\ \times\, (x_{L-1}-x_{L-2}) \cdots\cdots (x_{L-1}-x_1) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \times\, (x_2 - x_1) \end{aligned}$$

and S_{L-p} is an elementary symmetric function of the form

$$S_{L-p} = \sum x_L\, x_{L-1} \cdots x_{p+1}.$$

From (22), it follows that

$$\left\{ \begin{aligned} &|\Delta| = 1!\ 2! \cdots M!\,, \\ &\Delta_{pi} = \frac{T_{M-p,\,i}}{(N-i)!\,(M-N+i)!}\,|\Delta| = \frac{1}{M!}\binom{M}{N-i} T_{M-p,\,i}\,|\Delta|, \end{aligned} \right. \tag{23}$$

where $T_{M-p,i}$ is a quantity obtained from S_{M-p} by substituting

$$x_j = \left\{ \begin{array}{ll} j - 1 - (M - N) & \text{for}\quad j \leqq i + (M-N), \\ j - (M - N) & \text{for}\quad j > i + (M-N). \end{array} \right.$$

Consequently, if $r\neq 0$, from (21), $|G|\neq 0$ and

$$G^{-1} = \|G^{ij}(r)\| = \left\| \frac{(-1)^{i+j}\sum_{p=0}^{M} r^{-p}\, T_{M-p,\,i}\, T_{M-p,\,j}}{(N-i)!\,(M-N+i)!\,(N-j)!\,(M-N+j)!} \right\|. \tag{24}$$

From (19), det. $|g_{kj}(r, \varepsilon)|$ can be expanded as follows:

$$\text{det.}\ |g_{kj}(r,\ \varepsilon)| = |G| + \sum_{m=1}^{\infty} H_m(r)\, \varepsilon^{mK},$$

where $H_m(r)$ is a linear combination of the minor determinants in G of the order

$(K-1), (K-2), \cdots\cdots, (K-m)$ [1]. As in G_{ij}, it is readily seen that the minor determinants of $(K-l)$-th order have the order at least

$$[M(M+1)/2] - [M + (M-1) + \cdots + (M-l+1)]$$

with respect to r, consequently $H_m(r)r^{mK}$ has the order at least

$$[M(M+1)/2] + R$$

with respect to r, where

$$R = \begin{cases} m(m+1)/2 & \text{when } m \leqq K, \\ \{K(K+1)/2\} + \{(m-K)K\} & \text{when } m > K. \end{cases}$$

Then, when $0 < r \ll 1$, from (21), for $|\varepsilon| \leqq r$,

$$|\det. |g_{kj}(r, \varepsilon)|| > r^{M(M+1)/2}(|\varDelta|^2 - [r]_1) > 0 \text{ }^{2)},$$

since $|\varDelta| > 0$ from (23).

Now, corresponding to (14), let us consider the equations

$$\sum_j g_{kj}(r, \varepsilon)\, a_j = g_k(r, \varepsilon). \tag{25}$$

From (19), these equations are of the forms that follows:

$$\sum_j \frac{g_{kj}^{(0)}(r)}{1-(kj\varepsilon)^K}\, a_j = \sum_{m=0}^{\infty} g_k^{(m)}(r)\, \varepsilon^{mK}.$$

When $r \neq 0$, by the above results, $|\det. |g_{kj}(r, \varepsilon)|| > 0$ for $|\varepsilon| \leqq r$, consequently a_j's satisfying the above equations are unique and expanded as

$$a_j = \sum_{m=0}^{\infty} a_j^{(m)}(r)\, \varepsilon^{mK}, \tag{26}$$

of which the series of right-hand side converge for $|\varepsilon| \leqq r$. Then, substituting (26) into (25), we have:

$$\begin{cases} \sum_j g_{kj}^{(0)}(r)\, a_j^{(0)}(r) = g_k^{(0)}(r), \\ \sum_j g_{kj}^{(0)}(r)\, a_j^{(1)}(r) = g_k^{(1)}(r) - k^K \sum_j g_{kj}^{(0)}(r)\, j^K a_j^{(0)}(r), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases} \tag{27}$$

By (18), the first of these equations are written as follows:

1) Here we agree that the minor determinants of the order zero are 1 and those of the negative order are zero.

2) $[r]_1$ denotes a sum of the terms of the first and higher orders with respect to r.

$$\sum_{n=0}^{M} (kr)^n \left(\sum_j j^n a_j^{(0)}(r)\right) = \sum_{n=0}^{M} (kr)^n (-1)^n \frac{1}{n+1}\,.$$

These equations are evidently satisfied by $a_j^{(0)}(r) = a_j^{(0)}$ determined by (15). Thus we see that, for $r \neq 0$, $a_j^{(0)}(r)$'s coincide with the coefficients $a_j^{(0)}$'s of the traditional formulas and are independent of r. The equations (15) are easily solved by means of (23) as follows:

$$a_j^{(0)} = \frac{(-1)^{j+(M-N)}}{(N-j)!\,(M-N+j)!} \sum_{p=0}^{M} \frac{T_{M-p,\,j}}{p+1}\,. \tag{28}$$

Put

$$\delta a_j = \sum_{m=1}^{\infty} a_j^{(m)}(r)\, r^{mK}, \tag{29}$$

then, for $r = h_0^2$, by (20), from (25) and (26), δa_j's denote the deviations of a_j's of the new formulas from those of the traditional ones. In the sequel, we agree that $r = h_0^2$. Now, from (24), by induction, it is readily seen that the solutions $a_j^{(m)}(r)$ of (27) have the order at least $(-mM)$ with respect to r, consequently $a_j^{(m)}(r) r^{mK}$ has the order at least m. Then, by means of (18) and (24), from (27), δa_j's are calculated as follows:

$$\begin{aligned} \delta a_j &= a_j^{(1)}(r)\, r^K + [r]_2 \\ &= \sum_k G^{jk} \Big[g_k^{(1)} - k^K \sum_i g_{ki}^{(0)}\, i^K a_i^{(0)}\Big] r^K + [r]_2 \\ &= \sum_k \Big[\frac{(-1)^{j+k}}{(N-j)!\,(M-N+j)!\,(N-k)!\,(M-N+k)!} r^{-M} \\ &\qquad \times \Big\{(-1)^K \frac{k^K}{K+1} - k^K \sum_i i^K a_i^{(0)}\Big\}\Big] \cdot r^K + [r]_2, \end{aligned}$$

where $[r]_2$ is a sum of the terms of the second and higher orders with respect to r. Thus, within the first order with respect to r, we have:

$$\delta a_j \doteqdot \frac{(-1)^j}{(N-j)!\,(M-N+j)!} \left[\frac{(-1)^K}{K+1} - \sum_i i^K a_i^{(0)}\right] \left[\sum_k \frac{(-1)^k k^K}{(N-k)!\,(M-N+k)!}\right] r. \tag{30}$$

From this, we see that, *while the traditional formulas are independent of the adopted breadth of the intervals, the new formulas varies with the adopted breadth of the intervals and the deviations of the coefficients of the new formulas from those of the traditional ones increase in absolute values with the breadth of the intervals provided that the breadth of the intervals is sufficiently small.*

Let the deviation of $\sum(a)$ from $\sum(a^{(0)})$ be $\delta\sum$. Then, from (12), it follows that

$$\delta \sum = \sum_k \frac{\partial \sum(a^{(0)})}{\partial a_k^{(0)}} \delta a_k + \frac{1}{2} \sum_{k,j} \frac{\partial^2 \sum(a^{(0)})}{\partial a_k^{(0)} \partial a_j^{(0)}} \delta a_k \, \delta a_j$$

$$= 2r \sum_k [\sum_j g_{kj} \, a_j^{(0)} - g_k] \, \delta a_k + r \sum_{k,j} g_{kj} \, \delta a_k \, \delta a_j.$$

Since $a_j = a_j^{(0)} + \delta a_j$ satisfy (14), the above exression becomes

$$\delta \sum = - r \sum_{k,j} g_{kj} \, \delta a_k \, \delta a_j.$$

By means of (27) and (18), this can be calculated as follows:

$$\delta \sum \doteqdot - r^{K+1} \sum_k \delta a_k \, [g_k^{(1)} - k^K \sum_j g_{kj}^{(0)} \, j^K \, a_j^{(0)}]$$

$$\doteqdot - r^{K+1} \sum_k \delta a_k \left[\frac{(-1)^K k^K}{K+1} - k^K \sum_j j^K \, a_j^{(0)} \right]$$

$$= - r^{K+1} \left[\frac{(-1)^K}{K+1} - \sum_j j^K \, a_j^{(0)} \right] \cdot \sum_k k^K \, \delta a_k.$$

Finally, substituting (30) for δa_k, we have:

(31) $$\delta \sum \doteqdot - r^{K+1} \left[\frac{(-1)^K}{K+1} - \sum_i i^K \, a_i^{(0)} \right]^2 \cdot \left[\sum_k \frac{(-1)^k k^K}{(N-k)!\,(M-N+k)!} \right]^2 r.$$

Now, since $r = h_0^2 \ll 1$, from (15),

$$\sum (a^{(0)}) \doteqdot \left[\frac{(-1)^K}{K+1} - \sum_i i^K \, a_i^{(0)} \right]^2 r^{K+1}.$$

Consequently we have

(32) $$\delta \sum = - r \kappa \sum (a^{(0)}),$$

where

(33) $$\kappa = \left[\sum_k \frac{(-1)^k k^K}{(N-k)!\,(M-N+k)!} \right]^2.$$

Since $2\pi \sigma_{E_0}^2 = \sum(a)$ by (10), from (32), for the deviation $\delta \sigma_{E_0}$ of σ_{E_0}, it holds that

(34) $$\delta \sigma_{E_0} = - \frac{1}{2} r \kappa \sigma_{E_0}.$$

If we substitute (32) into (17), then $\lambda = r\kappa$, consequently the condition (16) becomes

$$\sum_{n=0}^{M} |c_n|^2 < r\,\kappa \sum_{n=K}^{\infty} |c_n|^2,$$

or

$$\sum_{n=0}^{M} |c_n' \rho^n|^2 < r\,\kappa \sum_{n=K}^{\infty} |c_n' \rho^n|^2. \tag{35}$$

Thus, by the reasonings of the preceding paragraph, we see that, *when* (35) *is fulfilled for the function* $f(x)$, *the new formulas are more accurate than the traditional ones provided that* $h_0 \ll 1$, *namely that* h *is sufficiently small compared with* ρ.

5. Approximate solution of (14)

In the preceding paragraph, assuming that $h_0 \ll 1$, we have calculated $\delta a_j = a_j - a_j^{(0)}$ within the first order with respect to $r = h_0^2$. In order to obtain the more minute solution of (14), it is necessary to calculate the terms of higher orders of δa_j's given by (29). In this paragraph, we calculate the terms of the second order. The terms of the third and higher orders can be calculated analogously, though the calculation itself becomes considerably complicated.

By the preceding paragraph, the terms of the second order appears only in $a_j^{(1)}(r) r^K$ and $a_j^{(2)}(r) r^{2K}$. Now, by (27),

$$a_j^{(1)}(r)\, r^K = \sum_k G^{jk} [g_k^{(1)} - k^K \sum_i g_{ki}^{(0)}\, i^K\, a_i^{(0)}]\, r^K,$$

consequently, substituting (24) and (18) into the right-hand side, we see that the coefficients of the terms of the second order appearing in $a_j^{(1)}(r) r^K$ are

$$\frac{(-1)^j}{(N-j)!\,(M-N+j)!}\left[\frac{(-1)^{K+1}}{K+2} - \sum_i i^{K+1} a_i^{(0)}\right] \cdot \left[\sum_k \frac{(-1)^k k^{K+1}}{(N-k)!\,(M-N+k)!}\right] \tag{36}$$

$$+ \frac{(-1)^j T_{1,j}}{(N-j)!\,(M-N+j)!}\left[\frac{(-1)^K}{K+1} - \sum_i i^K a_i^{(0)}\right] \cdot \left[\sum_k \frac{(-1)^k k^K T_{1,k}}{(N-k)!\,(M-N+k)!}\right],$$

where, by the definition,

$$T_{1,j} = \frac{N(N+1)}{2} - (M-N) - j.$$

Also, by (27),

$$a_j^{(2)}(r)\, r^{2K} = \sum_k G^{jk} [g_k^{(2)} - k^K \sum_i g_{ik}^{(0)}\, i^K a_i^{(1)} - k^{2K} \sum_i g_{ki}^{(0)}\, i^{2K}\, a_i^{(0)}]\, r^{2K},$$

consequently, substituting (24) and (18) into the right-hand side, we see that the terms of the second order appearing in $a_j^{(2)}(r) r^{2K}$ appear only in the expression

$$-\frac{(-1)^j}{(N-j)!\,(M-N+j)!}\left[\sum_k \frac{(-1)^k k^K}{(N-k)!\,(M-N+k)!}\right]\cdot\left[\sum_i i^K a_i^{(1)}(r)\,r^K\right]r.$$

Now, the terms of the first order of $a_i^{(1)}(r)r^K$ are given by (30). Consequently the coefficients of the terms of the second order appearing in $a_j^{(2)}(r)r^{2K}$ become

$$(37)\qquad -\frac{(-1)^j}{(N-j)!\,(M-N+j)!}\left[\sum_k \frac{(-1)^k k^K}{(N-k)!\,(M-N+k)!}\right]^3\cdot\left[\frac{(-1)^K}{K+1}-\sum_i i^K a_i^{(0)}\right].$$

Thus we see that the coefficients of the terms of the second order appearing in δa_j's given by (29) are the sum of the quantities given by (36) and (37).

For brevity, put

$$(38)\qquad \begin{cases} P_1 = \dfrac{(-1)^K}{K+1} - \sum\limits_i i^K a_i^{(0)}, \\ P_2 = \dfrac{(-1)^{K+1}}{K+2} - \sum\limits_i i^{K+1} a_i^{(0)}, \\ Q_1 = \sum\limits_k \dfrac{(-1)^k k^K}{(N-k)!\,(M-N+k)!}, \\ Q_2 = \sum\limits_k \dfrac{(-1)^k k^{K+1}}{(N-k)!\,(M-N+k)!}, \\ T = \dfrac{N(N+1)}{2} - (M-N) = \dfrac{N(N+3)}{2} - M, \\ R = P_1\,(T\,Q_1 - Q_2). \end{cases}$$

Then the coefficients of the terms of the second order appearing in δa_j's are expressed as follows:

$$\frac{(-1)^j}{(N-j)!\,(M-N+j)!}\left[(P_2 Q_2 + TR - P_1 Q_1^3) - jR\right].$$

Combining (30) with the above results, we see that, within the second order with respect to r, δa_j's are given by

$$(39)\qquad \delta a_j = \frac{(-1)^j}{(N-j)!\,(M-N+j)!}\left[P_1 Q_1 r + \{(P_2 Q_2 + TR - P_1 Q_1^3) - jR\}\,r^2\right].$$

6. Examples of new formulas

With a view to having actual utility, for $h_0 = 10^{-1}$, we seek for a_j solving numerically the equations (14) by the method of elimination. Since $|\det.|g_{kj}||$

$\ll 1$, for numerical solution of (14), the iterative method is not adequate, because the speed of convergence of iteration process is very slow. For example, in the case where $M=N+1=4$, as is seen from Table 1, the considerably minute approximate solution is found from correction δa_j's given by (39), but, if, from this approximate solution, we start the iterative procedure — for example, Seidel's procedure, we obtain the solution that follows:

$$\begin{aligned} a_{-1} &= 0.353239, \\ a_0 &= 0.878897, \\ a_1 &= -0.339460, \\ a_2 &= 0.129272, \\ a_3 &= -0.021948. \end{aligned}$$

These values differ only in a_{-1} by 1×10^{-6} from the starting values. Consequently, in this case, the iterative procedure does not serve to seek for a more accurate solution than the starting approximate solution. Thus, in order to solve numerically the equations (14), we must calculate the terms of higher orders of δa_j's given by (29) or else directly solve the equations (14) by the finite procedure — for example, the method of elimination. As is remarked at the beginning of this paragraph, in this note, we have adopted the method of elimination.

In solving (14), since $|\det.|g_{kj}|| \ll 1$, the values of g_k must be known with sufficient accuracy, consequently we have computed them to 20 decimal places directly from the series. The values of a_j's thus obtained are shown in Table 1 as "new". In this table, besides the values of a_j's, those of the other useful quantities are also shown. The values of $\sum(a)$ and $\sum(a^{(0)})$ were computed directly from $\sum(a)=\sum_{n=0}^{\infty}|E_0(u^n)|^2$.

Table 1.

(1) $N=3$, $M=3$,

	new	traditional	δa_j		
			true	by (30)	by (39)
a_0	2.210478	2.291667	−0.081189	−0.083667	−0.081118
a_1	−2.219167	−2.458333	0.239166	0.251000	0.238753
a_2	1.306852	1.541667	−0.234815	−0.251000	−0.234151
a_3	−0.298163	−0.375000	0.076837	0.083667	0.076517
Σ	6.452×10^{-9}	9.302×10^{-9}	λ	0.442	
σ_{E_0}	3.205×10^{-5}	3.848×10^{-5}	κ	36	

(2) $N=2,\ M=3$

	new	traditional	δa_j		
			true	by (30)	by (39)
a_{-1}	0.377072	0.375000	0.002072	0.002111	0.002071
a_0	0.785420	0.791667	−0.006247	−0.006333	−0.006245
a_1	−0.202055	−0.208333	0.006278	0.006333	0.006277
a_2	0.039564	0.041667	−0.002103	−0.002111	−0.002103
Σ	3.936×10^{-11}	4.103×10^{-11}	λ	0.042	
σ_{E_0}	2.503×10^{-6}	2.555×10^{-6}	κ	4	

(3) $N=4,\ M=4$

	new	traditional	δa_j		
			true	by (30)	by (39)
a_0	2.482958	2.640278	−0.157320	−0.164931	−0.156993
a_1	−3.244882	−3.852778	0.607896	0.659722	0.604880
a_2	2.752659	3.633333	−0.880674	−0.989583	−0.872684
a_3	−1.202509	−1.769444	0.566935	0.659722	0.558699
a_4	0.211775	0.348611	−0.136836	−0.164931	−0.133903
Σ	1.361×10^{-9}	3.631×10^{-9}	λ	1.668	
σ_{E_0}	1.472×10^{-5}	2.404×10^{-5}	κ	100	

(4) $N=3,\ M=4$

	new	traditional	δa_j		
			true	by (30)	by (39)
a_{-1}	0.353240	0.348611	0.004629	0.004688	0.004629
a_0	0.878887	0.897222	−0.018335	−0.018750	−0.018325
a_1	−0.339439	−0.366667	0.027228	0.028125	0.027207
a_2	0.129257	0.147222	−0.017965	−0.018750	−0.017950
a_3	−0.021945	−0.026389	0.004444	0.004688	0.004441
Σ	4.802×10^{-12}	6.234×10^{-12}	λ	0.298	
σ_{E_0}	8.742×10^{-7}	9.961×10^{-7}	κ	25	

As an example, by means of the new formulas for $M=4$, we find the solution of the equation

$$y' = 6y/(x-1)$$

with the initial condition that $y(0)=1$.

The starting values are found by means of Taylor series for $x=-0.2,\ -0.1,\ 0,\ 0.1$ and 0.2.

The solution thus found is tabulated as y_1 in Table 2. The exact solution is easily found to be $y=(x-1)^6$. This is tabulated as "true" in the table. For comparison, the approximate solution found by means of the traditional formulas for $M=4$ is also tabulated in the table as y_2. The table shows that the new formulas are superior in accuracy to the traditional ones in this case. For reference, $\nabla^4 y'$ for y_1 is also shown in the table.

Table 2.

x	y_1		true	y_2		$\nabla^4 y'$
	values	errors		values	errors	
−0.2			2.985984			
−0.1			1.771561			
0.0			1.000000			
0.1			0.531441			
0.2			0.262144			−0.072000
0.3	0.117835	+186	0.117649	0.117412	−237	
	0.117641	−8		0.117659	+10	−0.064731
0.4	0.046857	+201	0.046656	0.046406	−250	
	0.046649	−7		0.046668	+12	−0.057806
0.5	0.015801	+176	0.015625	0.015401	−224	
	0.015619	−6		0.015637	+12	−0.050194
0.6	0.004280	+184	0.004096	0.003847	−249	
	0.004093	−3		0.004106	+10	−0.043299
0.7	0.000877	+148	0.000729	0.000489	−240	
	0.000728	−1		0.000737	+8	−0.035939

Of the values of y, the above show the values obtained by the extrapolation formula and the lower those by the interpolation formula.

7. Remark

As is remarked in § 3, the new formulas are more accurate than the traditional ones when $\nabla^M y'$ reveals remarkable changes, consequently, for minute integration of differential equations, in the process of integration, corresponding to adequate $h_0=h/\rho$, either the traditional formulas or the new formulas should be choosed according to the behavior of $\nabla^M y'$, so long as the number M is kept and the intervals are not subdivided. When $h_0 \ll 1$, (30) or (39) serves to obtain the new formulas corresponding to any h_0.

Department of Mathematics,
Faculty of Science,
Hiroshima University.