J. Math. Tokushima Univ.
Vol. I. (1967), 29-35.

On the Remainder Terms of the Optimal Formulas

By

Yoshitane SHINOHARA

(Received September 30, 1967)

Abstract

The remainder term of the optimal multi-step formula, that is the stable formula of maximum order, is considered from the analytical point of view.

§1. Introduction

For the numerical integration of ordinary differential equation

(1.1) $$\frac{dy}{dx}=f(x,\ y),$$

we consider the multi-step formula

(1.2) $$\sum_{j=0}^{n}\alpha_j y(x_{k+j})=h\sum_{j=0}^{n}\beta_j f\big(x_{k+j},\ y(x_{k+j})\big)+Ch^{p+1}y^{(p+1)}(\xi)\quad (x_k\leqq\xi\leqq x_{k+n}),$$

where $y=y(x)$ is a solution of (1.1) on the interval $[a, b]$ satisfying the initial condition $y(a)=y_0$ and we shall always assume that $\alpha_n=1$, $|\alpha_0|+|\beta_0|>0$. If we associate with the formula (1. 2) the polynomials

(1.3) $$\begin{cases}\rho(\zeta)=\alpha_n\zeta^n+\alpha_{n-1}\zeta^{n-1}+\cdots\cdots+\alpha_1\zeta+\alpha_0,\\ \sigma(\zeta)=\beta_n\zeta^n+\beta_{n-1}\zeta^{n-1}+\cdots\cdots+\beta_1\zeta+\beta_0,\end{cases}$$

where the polynomials $\rho(\zeta)$ and $\sigma(\zeta)$ have no common factor, then, the stability conditions of (1.2) can be written as follows:

(1.4) $$\text{if } \rho(\zeta)=0,\text{ then}\begin{cases}\text{either} & |\zeta|<1\\ \text{or} & |\zeta|=1 \text{ and } \rho'(\zeta)\neq 0.\end{cases}$$

Moreover, if we postulate the conditions

(1.5) $$\alpha_\nu=-\alpha_{n-\nu},\quad \beta_\nu=\beta_{n-\nu},$$

then, the order p of the local truncation error of the stable multi-step method (1.2) becomes $p=n+2$ for even n, by the theorem of Dahlquist [1]. Hence, from (1.2), (1.3), (1.4) and (1.5), we have

$$\rho(E)y_k-h\sigma(E)f_k=Ch^{n+3}y^{(n+3)}(\xi) \quad (x_k\leqq\xi\leqq x_{k+n}), \tag{1.6}$$

where $y_n=y(x_n)$, $f_n=f(x_n, y(x_n))$ and E is the operator such that $Ey(x)=y(x+h)$. The stable multi-step formula (1.6) of maximum order will be called an optimal formula [3].

In the present paper, first, we express the constant C explicitly in terms of the coefficients α_i. Next, we have obtained the optimal formulas for $n=4$ and $n=6$, and then we show that the optimal formula for which the quantity $|C|$ attains a minimum does not exist for $n=4$ and $n=6$. Lastly, we compare them with Henrici's optimal formula [3] and Gorbunov's optimal formula [2] in a numerical example.

§2. Determination of the constant C

As is well-known, we have $E=e^{hD}$ formally where $D=d/dx$. Hence, from (1.6), we have

$$\rho(e^{hD}) \quad -h\sigma(e^{hD})D \quad \sim Ch^{n+3}D^{n+3} \quad (h\to 0)$$

formally. Replacing the operator e^{hD} by the scalar ζ, we see that the above relation is equivalent to the following one:

$$\rho(\zeta) \quad -(\log\zeta)\sigma(\zeta) \quad \sim C(\zeta-1)^{n+3} \quad (\zeta\to 1), \tag{2.1}$$

where $\log\zeta$ is a branch such that $\log\zeta\to 0$ as $\zeta\to 1$.

If we put $\eta=\zeta-1$, then (2.1) can be rewritten as follows:

$$\frac{\rho(1+\eta)}{\log(1+\eta)} \quad -\sigma(1+\eta) \quad \sim C\eta^{n+2} \quad (\eta\to 0). \tag{2.2}$$

According to the conditions (1.5) of maximum order, we may write:

$$\begin{aligned}\rho(\zeta)&=(\zeta^n-1)\alpha_0+(\zeta^{n-2}-1)\zeta\alpha_1+\cdots+(\zeta^2-1)\zeta^{\frac{n}{2}-1}\alpha_{\frac{n}{2}-1}\\&=\sum_{i=0}^{\frac{n}{2}-1}\zeta^i(\zeta^{n-2i}-1)\alpha_i\\&=\sum_{i=0}^{\frac{n}{2}-1}\zeta^i(\zeta^{2(\frac{n}{2}-i)}-1)\alpha_i\end{aligned}$$

$$=\sum_{i=0}^{\frac{n}{2}-1}\zeta^i(\zeta^2-1)(\zeta^{n-2i-2}+\zeta^{n-2i-4}+\cdots+\zeta^2+1)\alpha_i.$$

Hence, we have

(2.3) $$\rho(1+\eta)=\sum_{i=0}^{\frac{n}{2}-1}\eta(2+\eta)(1+\eta)^i[1+(1+\eta)^2+\cdots+(1+\eta)^{n-2(i+1)}]\alpha_i$$

$$=\sum_{i=0}^{\frac{n}{2}-1}\sum_{j=0}^{\frac{n}{2}-1-i}\eta(2+\eta)(1+\eta)^i(1+\eta)^{2j}\alpha_i$$

$$=\sum_{i=0}^{\frac{n}{2}-1}\sum_{j=0}^{\frac{n}{2}-1-i}\eta(2+\eta)(1+\eta)^{i+2j}\alpha_i.$$

Then, since

$$\frac{\eta(1+\eta)^{r-1}}{\log(1+\eta)}=\int_0^1(1+\eta)^{x+r-1}dx,$$

we have, from (2.3),

(2.4) $$\frac{\rho(1+\eta)}{\log(1+\eta)}=(2+\eta)\sum_{i=0}^{\frac{n}{2}-1}\alpha_i\sum_{j=0}^{\frac{n}{2}-1-i}\int_0^1(1+\eta)^{x+i+2j}dx.$$

Since $\sigma(1+\eta)$ is a polynomial of degree n at most, the constant C must be the coefficient of η^{n+2} in the power series $\rho(1+\eta)/\log(1+\eta)$ at $\eta=0$. Therefore, we obtain from (2.2) and (2.4),

(2.5) $$C=\sum_{i=0}^{\frac{n}{2}-1}\alpha_i\sum_{j=0}^{\frac{n}{2}-1-i}\left\{\frac{2}{(n+2)!}\int_0^1(x+i+2j)^{[n+2]}dx+\frac{1}{(n+1)!}\int_0^1(x+i+2j)^{[n+1]}dx\right\}.$$

A simple calculation gives for instance the following values of the constant C for $n=2, 4, 6$:

(2.6) $$\begin{cases} n=2, & C=\dfrac{1}{90}\alpha_0 \\ n=4, & C=\dfrac{1}{3780}(5\alpha_1+32\alpha_0) \\ n=6, & C=\dfrac{1}{907200}(832\alpha_1-184\alpha_2+5832\alpha_0) \end{cases}$$

where $\alpha_0=-1$.

§3. Table of the optimal formulas and a numerical example

Taking into account the properties $\alpha_n=1$, $\alpha_\nu=-\alpha_{n-\nu}$, $\beta_\nu=\beta_{n-\nu}$, (2.6) and stability condition (1.4), we have

$n=2$:

$$y_{k+2}=y_k+\frac{1}{3}h(f_{k+2}+4f_{k+1}+f_k)-\frac{1}{90}h^5y^{(5)}(\xi)$$

(Simpson's rule),

$n=4$:

(3.1)
$$\begin{aligned}&y_{k+4}+\mu(y_{k+3}-y_{k+1})-y_k\\&=h(\gamma_1f_{k+4}+\gamma_2f_{k+3}+\gamma_3f_{k+2}+\gamma_2f_{k+1}+\gamma_1f_k)\\&+\frac{5\mu-32}{3780}h^7y^{(7)}(\xi),\end{aligned}$$

where

$$\mu=\alpha_1,$$

$$\gamma_1=\frac{1}{1+\lambda^2}\left(\frac{1}{3}\lambda^2+\frac{13}{45}\right),$$

$$\gamma^2=\frac{1}{1+\lambda^2}\left(\frac{2}{3}\lambda^2+\frac{98}{45}\right),$$

$$\gamma_3=\frac{1}{1+\lambda^2}\left(-2\lambda^2+\frac{138}{45}\right),$$

$$\lambda^2=\frac{2-\mu}{2+\mu}\qquad(|\mu|<2).$$

For $n=6$, we have

(3.2)
$$\begin{aligned}&y_{k+6}+\mu(y_{k+5}-y_{k+1})+\lambda(y_{k+4}-y_{k+2})-y_k\\&=h(\delta_1f_{k+6}+\delta_2f_{k+5}+\delta_3f_{k+4}+\delta_4f_{k+3}+\delta_3f_{k+2}+\delta_2f_{k+1}+\delta_1f_k)\\&-\frac{1}{907200}(184\lambda-832\mu+5832)h^9y^{(9)}(\xi),\end{aligned}$$

where

$$\mu=\alpha_1, \quad \lambda=\alpha_2,$$

$$\delta_1=\frac{1}{1+\alpha^2+\beta^2}\left\{\frac{1}{2}+\frac{1}{6}(3\alpha^2-1)+\frac{1}{90}(45\beta^2-15\alpha^2-4)\right.$$

$$\left.-\frac{1}{1890}(84\alpha^2+315\beta^2+44)\right\},$$

$$\delta_2=\frac{1}{1+\alpha^2+\beta^2}\left\{3+\frac{1}{3}(3\alpha^2-1)-\frac{1}{45}(45\beta^2-15\alpha^2-4)\right.$$

$$\left.+\frac{1}{315}(84\alpha^2+315\beta^2+44)\right\},$$

$$\delta_3=\frac{1}{1+\alpha^2+\beta^2}\left\{\frac{15}{2}-\frac{1}{6}(3\alpha^2-1)-\frac{1}{90}(45\beta^2-15\alpha^2-4)\right.$$

$$\left.-\frac{1}{126}(84\alpha^2+315\beta^2+44)\right\},$$

$$\delta_4=\frac{1}{1+\alpha^2+\beta^2}\left\{10-\frac{2}{3}(3\alpha^2-1)+\frac{2}{45}(45\beta^2-15\alpha^2-4)\right.$$

$$\left.+\frac{2}{189}(84\alpha^2+315\beta^2+44)\right\},$$

$$\alpha^2=\frac{40-8\lambda}{8\mu+4\lambda+12},$$

$$\beta^2=\frac{-8\mu+4\lambda+12}{8\mu+4\lambda+12}.$$

In this case, the stability domain D is as follows:

$$D:\quad\begin{cases}\lambda<\dfrac{\mu^2}{4}+1\\ \lambda>-2\mu-3\\ \lambda\geqq 2\mu-3\\ |\mu|<4.\end{cases}$$

The above analysis shows that the minimum value of the constant C in absolute value can not be obtained under the stability conditions in both cases for $n=4$ and $n=6$, because $\mu=2$ (for $n=4$) and $\mu=4$, $\lambda=5$ (for $n=6$) are not contained in stability domain, respectively. The formula (3.1) for $\mu=1.41477653$ is Gorbunov's formula [2]. The formulas (3.2) for $\mu=-1$, $\lambda=1$ and $\mu=1.416369190$, $\lambda=1.002253240$ are Henrici's formula [3] and Gorbunov's formula [2], respectively.

We illustrate the optimal formula for which the error is smaller than that of the formulas obtained by Gorbunov and Henrici.

The Cauchy problem:

$$\frac{dy}{dx}=y, \quad y(0)=1$$

is integrated numerically with a fixed step-size $h=0.2$.
The starting values are computed from the exact solution e^x and the following predictor formulas are used for (3.1) and (3.2), respectively.

$$y_{k+4}-y_{k+3}=\frac{1}{24}h(55f_{k+3}-59f_{k+2}+37f_{k+1}-9f_k),$$

$$y_{k+6}-y_{k+5}=\frac{1}{1440}h(4277f_{k+5}-7923f_{k+4}+9982f_{k+3}$$
$$-7298f_{k+2}+2877f_{k+1}-475f_k).$$

$n=4$

μ	$y_k-e^{x_k}(x_k=10.6)$	C
0.9999991252	3.10×10^{-2}	$-7.142858300\times10^{-3}$
1.4147765300	2.52×10^{-2}	$-6.594210939\times10^{-3}$
1.4999986880	2.42×10^{-2}	$-6.481483217\times10^{-3}$
1.7499984690	2.14×10^{-2}	$-6.150795676\times10^{-3}$
1.9999982500	1.99×10^{-2}	$-5.820108134\times10^{-3}$

$n=6$

μ	λ	$y_k-e^{x_k}(x_k=11)$	C
-1.000000000	1.000000000	2.92×10^{-3}	7.548500882×10^{-3}
1.416369190	1.002253240	6.98×10^{-4}	5.332887379×10^{-3}
2.394427191	1.788854382	4.72×10^{-4}	4.595442883×10^{-3}

The computation is carried out in the floating-point arithmetic with 37 bits mantissa and rounding is done by chopping.

As is well-known, the optimal multi-step method is weakly unstable. From the practical point of view, therefore, the corrector formula such that the quantity $|C|$ of the remainder term is minimum under the strong stability should be considered. For a study of these we refer to Urabe's formulas [4, 5].

Department of Applied Mathematics
Faculty of Engineering
Tokushima University

References

[1] Dahlquist, G., Convergence and stability in the numerical integration of ordinary differential equations, Math. Scand., **4** (1956), 33-53.

[2] Gorbunov, A. D. and Šebalina, O. P., Predicting-correcting methods with optimum correction formula, Compt. Methods and Programming (Compt. Center Moscow Univ. Collect Works 2) (Russian), Izdat. Moscow Univ., Moscow (1965), 275-280.

[3] Henrici, P., Discrete variable methods in ordinary differential equations, Wiley, 1962.

[4] Shinohara, Y., On Urabe's predictor-corrector method for the numerical solution of ordinary differential equations, Bulletin of Faculty of Engineering, Tokushima Univ., **3-1** (1966), 75-85.

[5] Urabe, M., Yanagiwara, H. and Shinohara, Y., Periodic solution of van der Pol's equation with damping coefficient $\lambda=2$-10, J. Sci. Hiroshima Univ., Ser. A, **23** (1960), 325-366.