A CLASS OF DBDF METHODS WITH THE DERIVATIVE MODIFYING TERM*1)

XIANG JIA-XIANG (项家祥) KUANG JIAO-XUN (匡蛟勋)
(Shanghai Teachers University, Shanghai, China)

Abstract

In this paper, a class of DBDF methods with the derivative modifying term is presented. The form of the methods is

$$\sum_{j=0}^{k}\alpha_j y_{n+j}=hf_{n+k}+\alpha h^2 f'_{n+k}$$

which is of k-step and order $k+1$. The numerical stability of the new methods is much better than both Gear's methods and Enright's methods.

§ 1. Introduction

Since the famous thesis of G. Dahlquist——the order of any A-stable linear multistep method cannot exceed 2, and the smallest error constant is obtained for the trapezoidal rule was presented in 1963, the research of the numerical methods for stiff systems has been divided into two classes: 1) non-linear methods, such as one-leg methods and implicit Runge-Kutta methods; 2) stiff-stable linear multistep methods. The former is comparatively stable but more complex in computation while the latter is simple in the construction but weak in numerical stability. C. W. Gear introduced a class of backward differential methods (BDF) with a perfect program for automatic computation. Confined by the numerical stability, however, only the lower order methods can be used for highly oscillating systems, so it is inadequate for such problems. In this paper, we shall introduce a class of improved BDF methods which has a modifying term using the second derivative. Being assured of the zero-stability, the new methods have excellent absolute stability as compared with Gear's methods and Enright's methods which, as our new methods, contain the second derivative.

§ 2. The Construction of DBDF Methods

Following the notation in [5], we let D and E be the differential and displacement operator respectively, that is

$$Dy(x)=y'(x),\quad Ey(x)=y(x+h) \tag{1}$$

then the backward difference operator ∇ satisfies

$$\nabla y(x)=y(x)-y(x-h)=(I-E^{-1})y(x)$$

so that

* Received May 13, 1986.
1) The Project Supported by National Natural Science Foundation of China.

$$\nabla = I - E^{-1}. \tag{2}$$

By the operator formula (see [5]),

$$hD = \nabla + \frac{1}{2}\nabla^2 + \frac{1}{3}\nabla^3 + \cdots = \sum_{j=1}^{\infty}\frac{1}{j}\nabla^j \tag{3}$$

we have

$$h^2D^2 = \sum_{j=2}^{\infty}\sum_{i=1}^{j-1}\frac{1}{i(j-i)}\nabla^j. \tag{4}$$

Using (3) and (4) we obtain immediately

$$hD + \alpha h^2D^2 = \nabla + \sum_{j=2}^{\infty}\left(\frac{1}{j} + \sum_{i=1}^{j-1}\frac{\alpha}{i(j-i)}\right)\nabla^j, \tag{5}$$

where α is a real parameter.

Truncating the first k terms of the right hand side of (5) and together with (2), we have

$$hDE^k + \alpha h^2D^2E^k = E^k\left\{(I-E^{-1}) + \sum_{j=2}^{k}\left(\frac{1}{j} + \sum_{i=1}^{j-1}\frac{\alpha}{i(j-i)}\right)(I-E^{-1})^j\right\} = \rho_k(E),$$

and thus we have constructed a class of modified BDF methods with derivative modifying term, i.e. DBDF methods:

$$\sum_{j=0}^{k}\alpha_j y_{n+j} = hf_{n+k} + \alpha h^2 f'_{n+k}, \tag{6}$$

with its first characteristic polynomial

$$\rho_k(\zeta) = \zeta^k\left\{(1-\zeta^{-1}) + \sum_{j=2}^{k}\left[\frac{1}{j} + \sum_{i=1}^{j-1}\frac{\alpha}{i(j-i)}\right](1-\zeta^{-1})^j\right\}. \tag{7}$$

It is easy to calculate the coefficients α_j in (6):

$$\alpha_k = \sum_{j=1}^{k}\frac{1}{j} + \sum_{j=2}^{k}\sum_{i=1}^{j-1}\frac{\alpha}{i(j-i)},$$

$$\alpha_{k-p} = (-1)^p\sum_{j=p}^{k}\binom{j}{p}\left[\frac{1}{j} + \sum_{i=1}^{j-1}\frac{\alpha}{i(j-i)}\right], \quad p = 1, 2, \cdots, k.$$

Let $\mathscr{L}[E; h]$ denote the difference operator of DBDF methods (6), that is

$$\begin{aligned}\mathscr{L}[E; h] &\equiv \rho_k(E) - hDE^k - \alpha h^2D^2E^k \\ &= E^k\left\{hD + \alpha h^2D^2 - \sum_{j=k+1}^{\infty}\left[\frac{1}{j} + \sum_{i=1}^{j-1}\frac{\alpha}{i(j-i)}\right]\nabla^j\right\} - hDE^k - \alpha h^2D^2E^k \\ &= -E^k\sum_{j=k+1}^{\infty}\left\{\frac{1}{j} + \sum_{i=1}^{j-1}\frac{\alpha}{i(j-i)}\right\}\nabla^j \\ &\sim -\left[\frac{1}{k+1} + \sum_{i=1}^{k}\frac{\alpha}{i(k+1-i)}\right](hD)^{k+1}, \quad h \to 0.\end{aligned}$$

Hence the error constant

$$C_{k+1} = -\frac{1}{k+1} - \sum_{i=1}^{k}\frac{\alpha}{i(k+1-i)}. \tag{8}$$

Theorem 1. *DBDF methods (6) is of order $k+1$ if and only if* $\alpha = -\left(2\sum_{j=1}^{k}\frac{1}{j}\right)^{-1}$.

Proof. By (8), DBDF methods (6) is of order $k+1$ if and only if $C_{k+1} = 0$, i.e.

$$\alpha = -\left(\sum_{i=1}^{k}\frac{k+1}{i(k+1-i)}\right)^{-1}.$$

Notice

$$\sum_{j=1}^{k}\frac{1}{j}-\frac{1}{2}\sum_{j=1}^{k}\frac{k+1}{j(k+1-j)}=\sum_{j=1}^{k}\frac{1}{j}\Big(1-\frac{k+1}{2(k+1-j)}\Big)=\frac{1}{2}\Big(\sum_{j=1}^{k}\frac{1}{j}-\sum_{j=1}^{k}\frac{1}{k+1-j}\Big)=0$$

so we have

$$\sum_{j=1}^{k}\frac{1}{j}=\frac{1}{2}\sum_{j=1}^{k}\frac{k+1}{j(k+1-j)}$$

and this completes our proof.

By Theorem 1, we see that the method (6) is of order $k+1$ when the parameter α equals to half of the negative reciprocal of the first part of the leading coefficient α_k in $\rho_k(\zeta)$. The coefficients α_j, parameter α and the truncation error constant C_{k+2} of DBDF methods of order 3 to 9 are given in Table 1.

Table 1

k	2	3	4	5	6	7	8
α_k	$\frac{3}{2}+\alpha$	$\frac{11}{6}+2\alpha$	$\frac{25}{12}+\frac{35}{12}\alpha$	$\frac{137}{60}+\frac{15}{4}\alpha$	$\frac{147}{60}+\frac{203}{45}\alpha$	$\frac{363}{140}+\frac{469}{90}\alpha$	$\frac{761}{280}+\frac{29531}{5040}\alpha$
α_{k-1}	$-2-2\alpha$	$-3-5\alpha$	$-4-\frac{26}{3}\alpha$	$-5-\frac{77}{6}\alpha$	$-6-\frac{87}{5}\alpha$	$-7-\frac{223}{10}\alpha$	$-8-\frac{1924}{70}\alpha$
α_{k-2}	$\frac{1}{2}+\alpha$	$\frac{3}{2}+4\alpha$	$3+\frac{19}{2}\alpha$	$5+\frac{107}{6}\alpha$	$\frac{15}{2}+\frac{117}{4}\alpha$	$\frac{21}{2}+\frac{879}{20}\alpha$	$14+\frac{621}{10}\alpha$
α_{k-3}		$-\frac{1}{3}-\alpha$	$-\frac{4}{3}-\frac{14}{3}\alpha$	$-\frac{10}{3}-13\alpha$	$-\frac{20}{3}-\frac{254}{9}\alpha$	$-\frac{35}{3}-\frac{949}{18}\alpha$	$-\frac{56}{3}-\frac{8012}{90}\alpha$
α_{k-4}			$\frac{1}{4}+\frac{11}{12}\alpha$	$\frac{5}{4}+\frac{61}{12}\alpha$	$\frac{15}{4}+\frac{33}{2}\alpha$	$\frac{35}{4}+41\alpha$	$\frac{35}{2}+\frac{691}{8}\alpha$
α_{k-5}				$-\frac{1}{5}-\frac{5}{6}\alpha$	$-\frac{6}{5}-\frac{27}{5}\alpha$	$-\frac{21}{5}-\frac{201}{10}\alpha$	$-\frac{56}{5}-\frac{282}{5}\alpha$
α_{k-6}					$\frac{1}{6}+\frac{137}{180}\alpha$	$\frac{7}{6}+\frac{1019}{80}\alpha$	$\frac{14}{3}+\frac{2143}{90}\alpha$
α_{k-7}						$-\frac{1}{7}-\frac{7}{10}\alpha$	$-\frac{8}{7}-\frac{412}{70}\alpha$
α_{k-8}							$\frac{1}{8}+\frac{363}{560}\alpha$
C_{k+2}	$-\frac{1}{4}-\frac{11}{12}\alpha$	$-\frac{1}{5}-\frac{5}{6}\alpha$	$-\frac{1}{6}-\frac{137}{180}\alpha$	$-\frac{1}{7}-\frac{7}{10}\alpha$	$-\frac{1}{8}-\frac{363}{560}\alpha$	$-\frac{1}{9}-\frac{761}{1260}\alpha$	$-\frac{1}{10}-\frac{21387}{37800}\alpha$
α	$-\frac{1}{3}$	$-\frac{3}{11}$	$-\frac{6}{25}$	$-\frac{30}{137}$	$-\frac{30}{147}$	$-\frac{70}{363}$	$-\frac{140}{761}$

One can see at once that the absolute value of the error constant of DBDF methods is much smaller than that of Gear's methods, for example, in the case of $k=6$, $C_8\approx 0.0072$, and the error constant of Gear's method is $1/7\approx 0.14$, about 20 times larger than the former.

§ 3. Zero-Stability

Similar to the results in [5], considering the real roots of $\rho_k(\zeta)$, we have

Theorem 2. *Let α be given such that DBDF method* (6) *is of order $k+1$, and let ξ be any real root of $\rho_k(\zeta)$ such that $\xi\neq 1$, then $0<\xi<1$.*

Proof. We rewrite (7) as

$$\rho_k(\zeta)=\zeta^k(1-\zeta^{-1})\left\{1+\sum_{j=2}^{k}\left[\frac{1}{j}+\sum_{i=1}^{j-1}\frac{\alpha}{i(j-i)}\right](1-\zeta^{-1})^{j-1}\right\}=\zeta^k(1-\zeta^{-1})f(\zeta).$$

Obviously, for this given α, $\rho_k(0)=(-1)^k\left[\frac{1}{k}+\sum_{i=1}^{k-1}\frac{\alpha}{i(k-i)}\right]\neq 0$. So except for $\xi=1$, all the real roots of $\rho_k(\zeta)$ are in $f(\zeta)$. Let

$$z=1-\zeta^{-1} \tag{9}$$

and still denote by $f(z)$, that is

$$f(z)=1+\sum_{j=2}^{k}\left[\frac{1}{j}+\sum_{i=1}^{j-1}\frac{\alpha}{i(j-i)}\right]z^j. \tag{10}$$

The transformation (9) maps the interior of the unit circle $|\zeta|=1$ in ζ-plane onto the outside of the circle $|z-1|=1$ in z-plane conformally and maps $0<\zeta<1$ onto $z<0$. Denote

$$g(x,\alpha)=\frac{1}{x}+\sum_{i=1}^{[x]-1}\frac{\alpha}{i(x-i)}$$

then by Theorem 1, $g(k+1,\alpha)=0$ for $\alpha=-(2\sum 1/j)^{-1}$. It is easy to verify that

$$g(2,\alpha)>g(3,\alpha)>\cdots>g(k,\alpha)>g(k+1,\alpha)=0.$$

This means that (10) is a polynomial with positive real coefficients and hence has no positive real roots. By the transformation (9), we see that the real roots of $\rho_k(\zeta)$ except 1 are all within the interval (0, 1).

Theorem 2 locates the interval of the real spurious roots of $\rho_k(\zeta)$. To locate all the spurious roots of $\rho_k(\zeta)$, we give (see [6])

Theorem 3. *Let*

$$f(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$$

be a polynomial with positive real coefficients, and ξ be any root of $f(x)$. Then

$$\min_{1<i<n}\left\{\frac{a_{i-1}}{a_i}\right\}<|\zeta|<\max_{1<i<n}\left\{\frac{a_{i-1}}{a_i}\right\}.$$

Corollary. Let the polynomial in Theorem 3 be a polynomial with the coefficients a_i:

$$a_i=\sigma(-1)^i|a_i|,(\sigma=\pm 1),\ |a_n|>|a_{n-1}|>\cdots>|a_1|>|a_0|. \tag{11}$$

Let ξ be any root of $f(x)$. Then $|\xi|<1$.

Proof. Let

$$\begin{aligned}h(x)=f(-x)&=a_n(-x)^n+a_{n+1}(-x)^{n+1}+\cdots+a_1(-x)+a_0\\&=\sigma(|a_n|x^n+|a_{n-1}|x^{n-1}+\cdots+|a_1|x+|a_0|).\end{aligned}$$

Then except for a constant factor σ, $h(x)$ satisfies the condition in Theorem 3. Moreover, by (11)

$$\max_{1<i<n}\left\{\frac{|a_{i-1}|}{|a_i|}\right\}<1.$$

By Theorem 3, for any root of $f(x)$, we have

$$|\xi| = |-\xi| < \max_{1<i<n}\left\{\frac{|a_{i-1}|}{|a_i|}\right\} < 1.$$

Denote

$$\rho_k(\zeta)/(\zeta-1) = b_n\zeta^{n-1} + b_{n-1}\zeta^{n-2} + \cdots + b_2\zeta + b_1. \tag{12}$$

Table 2 gives the coefficients of the polynomial (12) of DBDF method (6). One can see at once that these polynomials satisfy the condition of the corollary. Hence all the spurious roots of $\rho_k(\zeta)$ are within the unit circle. That means DBDF methods of order $\leqslant 9$ are all zero-stable. Unfortunately, it is not true when $k=9$.

Table 2

k	b_k	b_{k-1}	b_{k-2}	b_{k-3}	b_{k-4}	b_{k-5}	b_{k-6}	b_{k-7}	$\max\left\{\frac{\lvert b_{i-1}\rvert}{\lvert b_i\rvert}\right\}$
2	1.17	−0.17							0.14
3	1.29	−0 35	0.06						0.27
4	1.38	−0 54	0.18	−0.03					0.39
5	1.46	−0.72	0.36	−0.12	0.017				0.49
6	1.53	−0.92	0.61	−0.30	0 087	−0.01			0.66
7	1.59	−1.11	0.91	−0.59	0.26	−0.07	0.01		0.82
8	1.61	−1 30	1.26	−1.02	0.59	−0.23	0.05	−0.01	0.97

Notice that only an upper bound has been given, and the module of the spurious roots may certainly be smaller than the bound. For example, in the case of $k=4$, $\max_{i\neq 1}|\xi_i| = 0.31$, while the estimate bound in Table 2 is 0.39.

§ 4. Numerical Stability

Let A_s be the set of the absolute stability of DBDF method (6) when it is applied to the test equation: $y'=\lambda y$, $y(0)=1$, where Re $\lambda<0$,

$$A_s = \{\bar{h}\in\mathbb{C} \mid |\xi_i(\bar{h})| < 0, \quad i=1, 2, \cdots, k\},$$

where $\bar{h}=\lambda h$, $\xi_i(\bar{h})$ are the roots of the characteristic equation

$$\pi(\xi;\bar{h}) = \rho_k(\xi) - \bar{h}\xi^k - \alpha\bar{h}^2\xi^k.$$

Denote the infinite wedge

$$W_\theta = \left\{\bar{h}\in\mathbb{C} \mid -\theta < \pi - \arg\bar{h} < \theta, \quad 0 \leqslant \theta \leqslant \frac{\pi}{2}\right\}.$$

The positive real parameter $a_{\min}$ and $\theta_{\max}$ defined in the following are very important for absolute stability

$$a_{\min} = \min\{a \mid \text{Re}\,\bar{h} < -a \rightarrow \bar{h}\in A_s\},$$

$$\theta_{\max} = \max\{\theta \mid W_\theta \subseteq A_s\}.$$

When $a_{\min}=0$ or $\theta_{\max}=90°$, the method is said to be A-stable. We need a numerical method with an $a_{\min}$ as small as possible, or a $\theta_{\max}$ as large as possible in order to deal with the highly oscillating systems.

The comparasion of stability parameter a_{min} and θ_{max} is given in Table 3.

Table 3

	k	2	3	4	5	6	7	8
a_{min}	DBDF	0	0	0.015	0.13	0.40	0.88	1.60
	GEAR	0	0.10	0.70	2.37	6.08		
θ_{max}	DBDF	90°	90°	89°	86°	79°	71°	69°
	GEAR	90°	86°	73°	51°	18°		

The following figure sketches the bound of absolute stability of DBDF methods, the dotted line sketches the bound of Enright's methods.

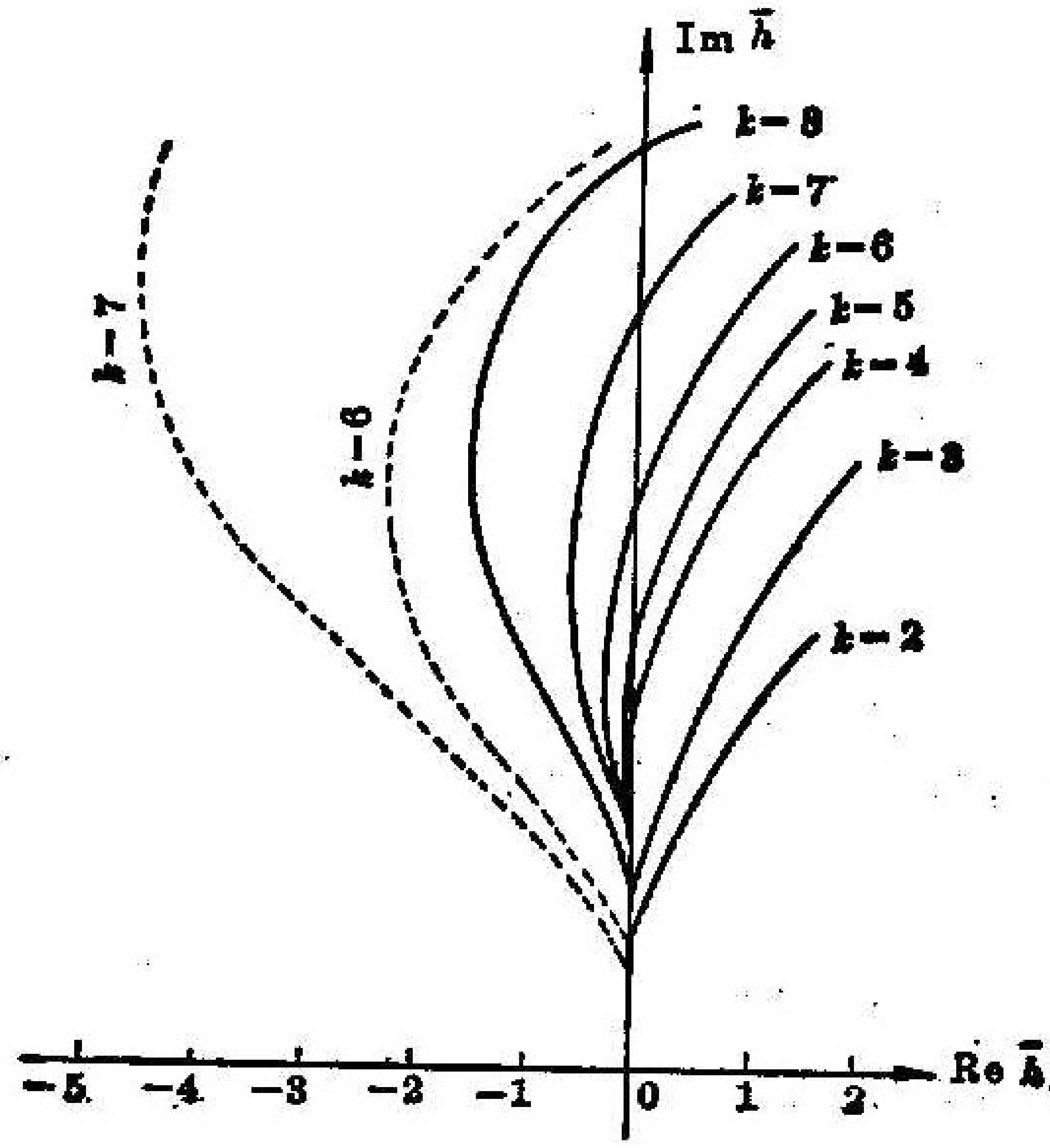

Fig. 1

Table 3 shows the incomparable superior of the numerical stability of DBDF methods, and Fig. 1 shows that our new methods are much more better than Enright's methods, both methods using the second derivative. Since Enright's methods are of k-step and order $k+2$, so k-step Enright's method and $k+1$-step DBDF method have the same precision. The a_{min} of the former is about 4.2 while that of the latter is about 1.6. (see the case of $k=7$)

§ 5. Numerical Test

We compute the following test equations:

$$\begin{cases} y' = -y - 10z \\ z' = 10y - z, \quad y(0) = 1, \ z(0) = 0 \end{cases}$$

using DBDF method and Gear's method of $k=4$ respectively, with the permission error $\varepsilon=0.001$. The computational results are givan in Table 4, in which SN is the step number, and SL is the step length.

Table 4

DBDF				Gear			
SN	SL	T	Error	SN	SL	T	Error
47	0.1	5.00	4.4 E-4	122	0.04	5.00	3.8 E-4
85	0.4	10.0	1.0 E-4	247	0.04	10.0	2.3 E-5
105	6.4	52.0	1.0 E-15	497	0.04	20.0	1.8 E-9
111	12.8	109.6	1.8 E-22				

The numerical test shows that the DBDF methods are very efficient for stiff and highly oscillating systems. They can be used in a large class of problems which Gear's methods may have difficulty in dealing with. However, since DBDF methods need the second derivative, this elicits the limitation of its application.

References

[1] G. Dahlquist, A special stability problem for linear multistep methods, *BIT*, **3** (1963), 27—43.

[2] C. W. Gear, The automatic integration for ordinary differential equations, *Comm. ACM*, **14** (1971), 176—179.

[3] ——, Algorithm 407 DIFSUB for solution of ordinary differential equations, *Comm. ACM*, **14** (1971), 185—190.

[4] W. H. Enright, Second derivative multistep methods for stiff ordinary differential equations, *SIAM J. Num. Anal.*, Vol. 11, No. 2, 321—331.

[5] Kuang Jiao-xun, Xiang Jia-xiang, A class of modified BDF methods, to appear on Mathematica Numerica Sinica.

[6] А. Г. Курош, Курс высшей алгебры, 1952.