

Michał Bernardelli

ALGORYTMY Z UŻYCIEM FORMUŁ WSTECZNEGO RÓŻNICZKOWANIA

[**Słowa kluczowe:** równania różniczkowe zwyczajne, formuły wstecznego różniczkowania, EBDF, EB^rDF , stabilność, stabilność absolutna, $A(\alpha)$ -stabilność, sztywność, rząd, zgodność, predyktor, korektor, schemat wielokrokowy]

Streszczenie

Praca ta jest poświęcona numerycznym algorytmom rozwiązywania sztywnych zagadnień początkowych równań różniczkowych zwyczajnych. Opisana klasa schematów EB^rDF jest uogólnieniem już istniejących formuł wstecznego różniczkowania BDF i rozszerzonych formuł EBDF, charakteryzuje się jednak lepszymi własnościami numerycznymi. Zastosowanie tak rozbudowanych schematów do rozwiązywania dużych układów równań różniczkowych zwyczajnych jest w dzisiejszych czasach możliwe dzięki niesamowitemu postępowi w dziedzinie komputeryzacji.

Zaimplementowane algorytmy do badania kąta α z definicji $A(\alpha)$ -stabilności ujawniły błędy w analizie zaprojektowanych w czasach osiemdziesiątych ubiegłego stulecia schematów. Umożliwiły także szybkie i dokładne badanie obszarów stabilności absolutnej bardziej skomplikowanych schematów typu predyktor-korektor, w tym formuł EB^rDF . Przedstawione w pracy porównania i wyliczenia stanowią motywację do dalszych badań w tym kierunku zarówno na polu teoretycznym jak i praktycznym.

1. Wprowadzenie

Na równania różniczkowe napotyka się przy badaniach w naukach przyrodniczych, takich jak fizyka, chemia czy biologia, lecz również

w ekonomii, medycynie, elektronice czy teorii chaosu deterministycznego. Zagadnień tych na ogół nie potrafimy rozwiązywać analitycznie. Jest to powodem, dla którego numeryczne rozwiązywanie równań różniczkowych jest tak potrzebnym narzędziem w rękach naukowców. W miarę rozwoju nauki rozpatrywane zjawiska stały się na tyle skomplikowane, że mimo wzrostu mocy obliczeniowej komputerów, niezbędne okazują się prace nad bardziej efektywnymi metodami znajdowania rozwiązań równań różniczkowych. Szczególne trudną klasę takich problemów stanowią zagadnienia sztywne.

Będziemy chcieli rozwiązywać zagadnienie początkowe Cauchy'ego

$$\begin{cases} \frac{dx}{dt} = f(t, x(t)), & x \in (C^1[t_0, T])^m \\ x(t_0) = x_0 \end{cases} \quad (1)$$

Z twierdzenia Peano i Picarda-Lindelöfa [14, str. 73] wiemy, że zadanie to ma jednoznaczne rozwiązanie dla funkcji

$$f(t, x) : \mathbb{R}^{m+1} \longrightarrow \mathbb{R}^m,$$

która jest ciągła i ograniczona w zbiorze

$$Q = \{(t, x) : |t - t_0| \leq a, |x - x_0| \leq b\},$$

oraz spełnia w nim warunek Lipschitza względem zmiennej x . W celu rozwiązania zagadnienia na przedziale $[t_0, T]$ dzielimy go na N równych części za pomocą punktów

$$t_k = t_0 + kh \quad (2)$$

dla $k = 0, 1, \dots, N$. Wielkość

$$h = \frac{T - t_0}{N}, \quad (3)$$

czyli długość jednej części, na jakie podzieliliśmy przedział, nazywamy *krokiem całkowania*. Szukamy ciągu $x_0, x_1, \dots, x_N$ takiego, który możliwie dobrze przybliży ciąg $x(t_0), x(t_1), \dots, x(t_N)$, gdzie $x(t)$ jest rozwiązaniem zagadnienia (1). Przyjmijmy też wygodne w użyciu oznaczenie $f_k = f(t_k, x_k)$.

Sposób na wyznaczanie kolejnych wyrazów ciągu $x_0, x_1, \dots, x_N$ nazywa się *schematem*. *Liniowy schemat wielokrokowy* zdefiniowany jest zależnościami

$$\sum_{j=0}^q \alpha_j x_{k+j} = h \sum_{j=0}^q \beta_j f_{k+j}, \quad (4)$$

które są spełnione dla $k = 0, 1, \dots$. Jeśli $\alpha_q \neq 0$ i $|\alpha_0| + |\beta_0| \neq 0$, to q nazywa się *liczbą kroków* schematu. Schemat q -krokowy nazywamy *zamkniętym*, jeśli $\beta_q \neq 0$. W przeciwnym razie mówimy o schemacie *otwartym*.

Zauważmy, że aby wyliczyć x_{k+q} musimy znać $x_k, x_{k+1}, \dots, x_{k+q-1}$, czyli q poprzednich wartości. Stąd nazwa: *schemat q -krokowy*. w przypadku schematów zamkniętych znalezienie wartości x_{k+q} wymaga rozwiązania równania (na ogół nieliniowego), ponieważ x_{k+q} występuje zarówno po lewej jak i po prawej stronie wzoru. Twierdzenie dotyczące jednoznaczności rozwiązania schematów zamkniętych można znaleźć w [9, str. 188].

Podstawowymi pojęciami przy porównywaniu własności schematów są *rzęd*, *zgodność* (konsystentność) oraz *stała błędu* schematu. z dokładnymi definicjami tych pojęć można zapoznać się między innymi w [12], [14] i [15]. Warto zauważyć, że powyższe pojęcia są związane z lokalnym błędem schematu i mają znaczący wpływ na szybkość jego *zbieżności*. Trzeba przy tym wyraźnie zaznaczyć, iż w definicji zbieżność q -krokowych schematów linowych chodzi o zbieżność $x_k \rightarrow x(t)$ przy $h \rightarrow 0$, $k \rightarrow \infty$ i dowolnym, ustalonym t , to jest

$$x_k \rightarrow x(\underbrace{t_0 + hk}_{const}). \quad (5)$$

Do badania numerycznych własności schematów niezbędne jest pojęcie *stabilności*. Wprowadźmy najpierw dwa wielomiany pomocnicze związane ze schematem (4), zwane *pierwszym* i *drugim wielomianem charakterystycznym*:

$$\rho(z) = \alpha_0 + \alpha_1 z + \dots + \alpha_1 z^q = \sum_{j=0}^q \alpha_j z^j, \quad (6)$$

$$\sigma(z) = \beta_0 + \beta_1 z + \dots + \beta_1 z^q = \sum_{j=0}^q \beta_j z^j. \quad (7)$$

Schemat (4) jest *zero-stabilny* (lub po prostu *stabilny*) jeśli wszystkie pierwiastki wielomianu ρ znajdują się w kole jednostkowym $|z| \leq 1$ na płaszczyźnie zespolonej, zaś te pierwiastki, które leżą na okręgu $|z| = 1$, są jednokrotne. Stabilność jest warunkiem koniecznym do tego, aby rozwiązania schematu (4) nie były rozbieżne do nieskończoności, gdy $h \rightarrow 0$. Więcej szczegółów na ten temat można znaleźć na przykład

w [14, str. 159]. Okazuje się, że zgodność i stabilność są wystarczające do uzyskania zbieżności schematu. Mówi o tym (patrz [6]):

Twierdzenie 1 (Dahlquist, 1956)

Niech rozwiązanie $x(t)$ zagadnienia Cauchy'ego (1) będzie klasy $C^{p+1}([t_0, T])$. Jeśli schemat (4) rzędu $p \geq 1$ jest stabilny oraz punkty startowe $x_0, x_1, \dots, x_{q-1}$ są dobrane tak, że

$$|x_j - x(t_j)| \leq R_1 h^p \quad j = 0, 1, \dots, q-1$$

dla $R_1 > 0$, to jest on zbieżny oraz zachodzi oszacowanie

$$|x_q - x(t_q)| \leq R h^p,$$

gdzie R jest pewną dodatnią stałą.

Dowód tego twierdzenia można znaleźć też między innymi w [14, str. 165].

Do tej pory byliśmy zainteresowani własnościami teoretycznymi schematu (4) przy $h \rightarrow 0$ i $k \rightarrow \infty$, ale stałej wartości $t_k = t_0 + kh$. W praktyce jednak mamy do czynienia z inną sytuacją, a mianowicie długość pojedynczego kroku $h > 0$ jest stała (nie dąży do zera), zaś liczba wykonywanych kroków jest duża ($k \rightarrow \infty$). Zachowanie się numerycznego rozwiązania równania różniczkowego zwyczajnego i jego stabilności względem zaburzeń wartości początkowych jest w ogólności niezmiernie trudno weryfikowalna. Rozważa się w związku z tym *testowe zagadnienie liniowe*:

$$\begin{cases} x'(t) = \lambda x(t), & x \in (C^1[t_0, T])^m, \lambda \in \mathbb{C} \\ x(t_0) = x_0 \end{cases} \quad (8)$$

i sprawdza się działanie schematów na tym szczególnym zagadnieniu. Mówimy, że schemat (4) jest *absolutnie stabilny* w punkcie $\bar{h} \in \mathbb{C}$, jeśli pierwiastki $z_j(\bar{h})$ dla $j = 1, 2, \dots, q$ jego wielomianu charakterystycznego

$$\pi(z, \bar{h}) = \rho(z) - \bar{h}\sigma(z) \quad (9)$$

są mniejsze lub równe od jedynki. Zbiór $\Omega \subset \mathbb{C}$ nazywa się *obszarem stabilności absolutnej* schematu, jeśli schemat ten jest absolutnie stabilny dla każdej liczby zespolonej $\bar{h} \in \Omega$. Wielkość $\bar{h}$ jest związana z krokiem całkowania zależnością:

$$\bar{h} = \lambda h. \quad (10)$$

Można udowodnić [14, str. 229], że jeśli $\bar{h}$ leży w obszarze stabilności absolutnej, to globalny błąd schematu jest ograniczony przez stałą niezależną od k i h . Dla kroku całkowania h , jeżeli tylko $\bar{h}$ jest w obszarze stabilności absolutnej, można prowadzić obliczenia bez obawy drastycznego wzrostu błędu globalnego. Warto dodać, iż obszary stabilności absolutnej schematów zamkniętych są większe niż schematów otwartych, a zwiększenie rzędu – czyli dokładności schematu – jest związane na ogół ze zmniejszeniem się obszaru stabilności absolutnej.

Rozszerzenie teorii na układy równań liniowych prowadzi do pojęcia *szywności*. Niech A będzie macierzą wymiaru $m \times m$ i niech λ_j dla $j = 1, 2, \dots, m$ będą wartościami własnymi macierzy A . Dla układów równań liniowych

$$\begin{cases} x'(t) = Ax(t) \\ x(0) = x_0 \end{cases} \quad (11)$$

gdzie $x(t)$, x_0 są wektorami m -wymiarowymi, obszar stabilności absolutnej Ω powinien zawierać wszystkie liczby $h\lambda_j$, $j = 1, 2, \dots, m$ [13, str. 220]. Układ m równań różniczkowych liniowych ze stałą macierzą A

$$x'(t) = Ax(t) \quad (12)$$

nazywa się *układem szywnym*, jeśli dla każdego $j = 1, 2, \dots, m$ zachodzi $\operatorname{Re}(\lambda_j) < 0$, zaś *współczynnik szywności*

$$s = \frac{\max_{j=1,2,\dots,m} |\lambda_j|}{\min_{j=1,2,\dots,m} |\lambda_j|} \quad (13)$$

jest znacznie większy od 1 (nawet rzędu miliona).

Ponieważ dla problemów szywnych wartości własne odpowiedniej macierzy Jacobiego mogą mieć bardzo wielkie moduły, więc do całkowania takich równań nadają się tylko takie schematy, które mają duży obszar stabilności absolutnej. Skoro bowiem wielkości $h\lambda$ muszą się znaleźć w obszarze stabilności absolutnej, a λ jest bardzo duża, to albo będziemy zmuszeni wybrać bardzo mały krok całkowania h , albo – co z praktycznego punktu widzenia wydaje się znacznie bardziej rozsądnym rozwiązaniem – musimy zadbać, aby obszar stabilności absolutnej był dostatecznie duży. Najlepiej byłoby, gdyby obszar ten obejmował całą półpłaszczyznę $\operatorname{Re}(z) < 0$.

Schemat (4) nazywamy *A-stabilnym* jeśli jego obszar stabilności absolutnej zawiera półpłaszczyznę $\operatorname{Re}(z) < 0$. Niestety definicja ta

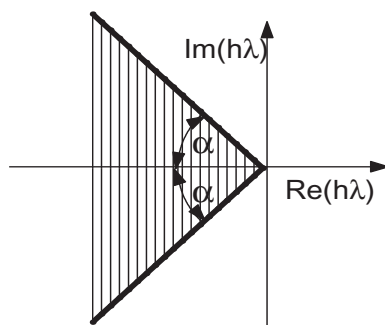
okazuje się zbyt restryktywna, co pokazuje następujące twierdzenie, którego dowód można przeczytać w [7]:

Twierdzenie 2 (Dahlquist, 1963)

- (i) Żaden wielokrokowy schemat otwarty nie może być A -stabilny.
- (ii) Rząd wielokrokowego A -stabilnego schematu zamkniętego nie może przekraczać 2.
- (iii) A -stabilny schemat wielokrokowy rzędu 2 z najmniejszą stałą błędu to schemat trapezów:

$$x_{k+1} = x_k + \frac{1}{2}h(f_k + f_{k+1}).$$

Nieco mniej ograniczającym warunkiem stabilności, spełnianym przez znacznie szerszą klasę schematów, jest $A(\alpha)$ -stabilność. Schemat na-



Rysunek 1: Schemat $A(\alpha)$ -stabilności.

zywa się $A(\alpha)$ -stabilny, jeśli jego obszar stabilności absolutnej zawiera obszar kątowy

$$\pi - \alpha < \arg z < \pi + \alpha,$$

gdzie $0 \leq \alpha \leq \frac{\pi}{2}$ rys. 1. Schemat jest $A(0)$ -stabilny jeżeli jest $A(\alpha)$ -stabilny dla pewnego $\alpha \in (0, \pi/2)$. Analogicznie jak w przypadku A -stabilności (twierdzenie 2) można udowodnić (patrz [16])

Twierdzenie 3 (Widlund, 1967)

- (i) Żaden wielokrokowy schemat otwarty nie może być $A(\alpha)$ -stabilny.

- (ii) *Rząd q -krokowego $A(\alpha)$ -stabilnego schematu nie może być większy niż q .*

Tworząc schemat wielokrokowy musimy zatem zadbać, aby jego rząd był odpowiednio wysoki, bo gwarantuje to odpowiednią dokładność obliczeń. z drugiej strony obszar stabilności absolutnej schematu powinien być jak największy, gdyż implikuje to ograniczoność błędu globalnego. Zachowanie odpowiednich proporcji pomiędzy tymi dwoma warunkami zdaje się być receptą na zbudowanie dobrego schematu wielokrokowego.

Definicje dotyczące liniowych układów równań różniczkowych zwyczajnych uogólniają się na układy nieliniowe. Do tego celu wykorzystuje się *metodę lokalnej linearyzacji*. Założenie, że funkcja f z równania (1) jest dwukrotnie różniczkowalna względem t i x w pewnym otoczeniu punktu (t_0, x_0) , pozwala na wyodrębnienie części liniowej względem x . Otrzymamy wówczas układ równań liniowych postaci

$$\frac{dx}{dt} = f_x(t_0, x_0)x(t), \quad (14)$$

gdzie $f_x(t_0, x_0)$ jest stałą macierzą Jacobiego.

W przypadku rozważania obszaru stabilności absolutnej wystarczy żądać, aby należały do niego wszystkie liczby $h\lambda_j$, $j = 1, 2, \dots, m$, gdzie λ_j są wartościami własnymi macierzy Jacobiego $f_x(t_0, x_0)$. Natomiast nieliniowy system (1) jest uważany za sztywny w punkcie t_0 , jeżeli układ równań różniczkowych z macierzą Jacobiego $\frac{\partial f(t_0, x(t_0))}{\partial x}$ jest sztywny.

Praca ta jest poświęcona numerycznym metodom rozwiązywania sztywnych zagadnień początkowych równań różniczkowych zwyczajnych. Ograniczono się przy tym tylko do klasy schematów znanych pod nazwą *formuł wstecznego różniczkowania*. Obok znanych już schematów przedstawiono ich uogólnienia i zbadam je pod kątem własności numerycznych.

Praca zorganizowana jest w następujący sposób. W następnej części podana jest ogólna teoria związana z metodami typu predyktor-korektor. Przedstawione są również formuły wstecznego różniczkowania BDF oraz ich rozszerzone odpowiedniki EBDF, ze szczególnym uwzględnieniem obszarów stabilności absolutnej tych schematów. w trzeciej części przedstawiono uogólnione formuły wstecznego różniczkowania EB^rDF charakteryzujące się lepszymi własnościami

numerycznymi. Udowodniono twierdzenie dotyczące rzędu tych metod oraz pokazano obliczone obszary stabilności absolutnej tych formuł. W czwartej, ostatniej części, dokonano krótkiego podsumowania przedstawionych w tej pracy algorytmów i oceniono ich przydatność do rozwiązywania sztywnych nieliniowych układów równań różniczkowych zwyczajnych.

2. Formuły wstecznego różniczkowania

Konieczność używania schematów zamkniętych jest oczywista. Mają one szereg lepszych cech w porównaniu ze schematami otwartymi, na przykład większy obszar stabilności absolutnej. Niestety, stosując metodę niejawną, na każdym kroku obliczeń zmuszeni jesteśmy do rozwiązania równania lub układu równań algebraicznych

$$\mathbf{x}_{\mathbf{k}+\mathbf{q}} = \Phi(x_k, x_{k+1}, \dots, x_{k+q-1}, \mathbf{x}_{\mathbf{k}+\mathbf{q}}), \quad (15)$$

gdzie funkcja Φ jest na ogół nieliniowa. Równania typu (15) rozwiązujemy metodami iteracyjnymi, np. metodą Newtona [9].

Metody iteracyjne wymagają podania przybliżenia początkowego, przy czym metody te są tym szybciej (jeżeli w ogóle) zbieżne, im lepsze jest podane początkowe przybliżenie. w praktyce stosuje się metody typu *predyktor-korektor*. Najpierw obliczamy przybliżoną wartość x_{k+q} schematem otwartym (*predyktor*), a następnie poprawiamy tę wartość schematem zamkniętym (*korektor*). Wyznaczając dokładniejszą wartość x_{k+q} metodą niejawną wykorzystujemy wspomniane powyżej metody iteracyjne. Dzięki doborowi odpowiedniego predyktora wystarczy wykonać zwykle tylko kilka iteracji, ponieważ startujemy z dobrego przybliżenia. Warto dodać, iż oba schematy, to jest predyktor i korektor, powinny być podobnych rzędów, gdyż o dokładności obliczeń decyduje składnik o niższym rzędzie, a o koszcie metody składnik o rzędzie wyższym. Mówi o tym:

Twierdzenie 4

Niech x_k oznacza końcowy wynik działania predyktora i korektora na k -tym kroku, zaś C_{p+1} będzie stałą błędu korektora. Jeżeli rząd p korektora jest mniejszy lub równy od rzędu p^ predyktora, to dla dowolnej liczby iteracji predyktora i korektora*

$$x_k - x(t_k) = C_{p+1} h^{p+1} x^{(p+1)}(t_k) + O(h^{p+2}),$$

Zatem cały schemat typu predyktor-korektor jest rzędu p ze stałą błędu C_{p+1} .

Jeżeli rząd p^* predyktora jest mniejszy od rzędu p korektora ($p^* = p - r$ dla $r = 1, 2, \dots$), to

$$\begin{aligned} x_k^{(m)} - x(t_k) &= O(h^{p-r+m+1}) & \text{dla } m = 1, 2, \dots, r, \\ x_k^{(m)} - x(t_k) &= C_{p+1} h^{p+1} x^{(p+1)}(t_k) + O(h^{p+2}) & \text{dla } m > r, \end{aligned}$$

gdzie m jest liczbą wykonanych iteracji przy wyznaczaniu przybliżenia rozwiązania równania korektora.

Dowód tego twierdzenia można znaleźć w [12, str. 103]. Można także pokazać, że jeżeli schemat używany w roli korektora jest zbieżny, to zbieżna jest także cała metoda typu predyktor-korektor (bez względu na zbieżność predyktora). Z powyższego twierdzenia widać, że najkorzystniej jest przyjąć $p^* = p$, ewentualnie $p^* = p - 1$.

Przejdźmy teraz do *schematów ze wstecznym różniczkowaniem* lub po prostu *formuł wstecznego różniczkowania*. Najczęściej używa się na ich określenie skrót BDF od angielskiego *Backward Differentiation Formulae*. Dane są one wzorem:

$$\sum_{j=0}^q \alpha_j x_{k+j} = h \beta_q f_{k+q}, \quad (16)$$

przy czym współczynniki α_j i β_q podane są w tabeli 1. Uwzględniono tam także kąt α z definicji $A(\alpha)$ -stabilności, rząd i stałą błędu metody. Można wykazać [5], że dla $q \geq 7$ formuły wstecznego różniczkowania rzędu q nie są zero-stabilne, dlatego są one bezużyteczne z praktycznego punktu widzenia.

Duże kąty α z definicji $A(\alpha)$ -stabilności dla początkowych formuł BDF wskazują na przydatność tych schematów w numerycznym rozwiązywaniu układów równań różniczkowych zwyczajnych o dużym współczynniku sztywności. Nazwa schematów BDF wzięła się od pewnej formy ich zapisu. Okazuje się bowiem [8, str. 350], iż każdą formułę wstecznego różniczkowania można przedstawić przy pomocy *operatora wstecznego różniczkowania*

$$\nabla x_{n+1} = x_{n+1} - x_n. \quad (17)$$

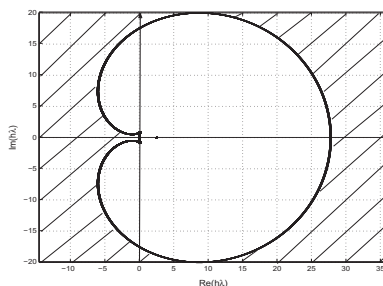
Tablica 1: Dane dotyczące formuł wstecznego różniczkowania (BDF) dla $q = 1 \dots 6$.

q	α_6	α_5	α_4	α_3	α_2	α_1	α_0	β_q	rzęd p	C_{p+1}	α_{max}
1						1	-1	1	1	$-\frac{1}{2}$	90°
2					1	$-\frac{4}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	2	$-\frac{2}{9}$	90°
3				1	$-\frac{18}{11}$	$\frac{9}{11}$	$-\frac{2}{11}$	$\frac{6}{11}$	3	$-\frac{3}{22}$	86.03°
4			1	$-\frac{48}{25}$	$\frac{36}{25}$	$-\frac{16}{25}$	$\frac{3}{25}$	$\frac{12}{25}$	4	$-\frac{12}{125}$	73.35°
5		1	$-\frac{300}{137}$	$\frac{300}{137}$	$-\frac{200}{137}$	$\frac{75}{137}$	$-\frac{12}{137}$	$\frac{60}{137}$	5	$-\frac{10}{137}$	51.84°
6	1	$-\frac{360}{147}$	$\frac{450}{147}$	$-\frac{400}{147}$	$\frac{225}{147}$	$-\frac{72}{147}$	$\frac{10}{147}$	$\frac{60}{147}$	6	$-\frac{20}{147}$	17.84°

Granica obszaru stabilności absolutnej metod wstecznego różniczkowania dana jest wzorami:

$$\begin{cases} Re(\psi(\theta)) &= \frac{1}{\beta_q} \sum_{j=0}^q \alpha_j \cos(q-j)\theta, \\ Im(\psi(\theta)) &= -\frac{1}{\beta_q} \sum_{j=0}^q \alpha_j \sin(q-j)\theta. \end{cases} \quad (18)$$

Przykładowy obszar stabilności absolutnej jest podany na rysunku 2.



Rysunek 2: Obszar stabilności absolutnej (zakreskowany) schematu BDF rzędu 6.

W 1980 roku J. R. Cash [2] zaproponował rozszerzenie schematów wstecznego różniczkowania (ang. *Extended Backward Differentiation*

Formulae). Są one oparte na formule

$$x_{k+q} + \sum_{j=0}^{q-1} \alpha_j x_{k+j} = h\beta_q f_{k+q} + h\beta_{q+1} f_{k+q+1}. \quad (19)$$

Zauważmy, że w przeciwieństwie do typowych liniowych schematów wielokrokowych, w których do wyznaczenia wartości x_{k+q} , przybliżającej dokładne rozwiązanie $x(t_{k+q})$ zagadnienia (1), wykorzystuje się punkty $x_k, x_{k+1}, \dots, x_{k+q-1}$, tutaj dochodzi dodatkowy, jeszcze nie wyznaczony, punkt x_{k+q+1} . Wartości wszystkich współczynników dobrane są tak, żeby formuła miała rząd równy $q+1$. Wymaganie to jednoznacznie definiuje współczynniki (19).

Trzeba jeszcze sprecyzować w jaki sposób wykonywać obliczenia takim schematem, mając dane startowe punkty $x_k, x_{k+1}, \dots, x_{k+q-1}$. Otóż przy pomocy jakiegoś predyktora wyznaczamy punkty u_j będące przybliżeniami wartościami punktów x_j . Najpierw liczymy punkt u_{k+q} – przybliżenie punktu x_{k+q} . Następnie tym samym schematem predyktora wyznaczamy punkt u_{k+q+1} , czyli aproksymację punktu x_{k+q+1} . Dopiero na końcu obliczamy szukaną wartość x_{k+q} na podstawie $x_k, x_{k+1}, \dots, x_{k+q-1}, u_{k+q+1}$. Oczywiście podstawowe własności schematu, typu stabilność czy rząd, są zależne w dużej mierze od tego, jaki predyktor został użyty do obliczeń. Z twierdzenia 4 wiemy, że rząd takiego predyktora musi być równy co najmniej q , żeby cały schemat predyktor-korektor był rzędu $q+1$. Cash użył jako predyktora q -krokowej metody wstecznego różniczkowania

$$x_{k+q} = a_0 x_k + a_1 x_{k+1} + \dots + a_{q-1} x_{k+q-1} + h b_q f_{k+q}. \quad (20)$$

Predyktor jest schematem zamkniętym, więc na każdym kroku jego obliczeń jesteśmy zmuszeni rozwiązywać równanie algebraiczne. Dla każdej metody iteracyjnej, używanej do rozwiązania tego równania, potrzebny jest jednak jakiś punkt startowy – przybliżenie początkowe, znacznie przyspieszające, a w wielu przypadkach w ogóle umożliwiające poprawne działanie metody iteracyjnej. Taki swego rodzaju pre-predyktor powinien być oczywiście schematem otwartym. Cash ([3], porównaj też [8, str. 311]) wykorzystał w tym celu własności operatorów wstecznego różniczkowania

$$x_{n+1} = \sum_{j=0}^q \nabla^j x_n + O(h^{q+1}), \quad (21)$$

dla ustalonego kroku całkowania h i punktów $x_n, x_{n+1}, \dots, x_{n-q}$. Schemat (19) z q -krokową metodą BDF jako predyktorem oraz

$$x_{n+1} = \sum_{j=0}^q \nabla^j x_n \quad (22)$$

jako pre-predyktorem nosi miano metody EBDF. Algorytm EBDF wygląda zatem następująco:

Krok 1. Użyj schematu BDF do obliczenia u_{k+q} :

$$u_{k+q} = a_0 x_k + a_1 x_{k+1} + \dots + a_{q-1} x_{k+q-1} + h b_q f(t_{k+q}, u_{k+q}), \quad (23)$$

przyjmując za przybliżenie początkowe punkt u_{k+q+1} z poprzedniej iteracji.

Krok 2. Użyj schematu BDF do obliczenia u_{k+q+1} :

$$u_{k+q+1} = a_0 x_{k+1} + a_1 x_{k+2} + \dots + a_{q-2} x_{k+q-1} + a_{q-1} u_{k+q} + h b_q f(t_{k+q+1}, u_{k+q+1}), \quad (24)$$

przyjmując za przybliżenie początkowe $u_{k+q+1}^{(0)}$ punkt

$$u_{k+q+1}^{(0)} = \sum_{j=0}^q \nabla^j u_{k+q}, \quad (25)$$

obliczony na podstawie punktów $x_k, x_{k+1}, \dots, x_{k+q-1}, u_{k+q}$.

Krok 3. Użyj korektora do uzyskania rozwiązania rzędu $q + 1$ w punkcie t_{k+q} :

$$x_{k+q} = \alpha_0 x_k + \alpha_1 x_{k+1} + \dots + \alpha_{q-1} x_{k+q-1} + h \beta_q f_{k+q} + h \beta_{q+1} f(t_{k+q+1}, u_{k+q+1}), \quad (26)$$

przyjmując za przybliżenie początkowe punkt u_{k+q} z danej iteracji.

Wektory u_{k+q} i u_{k+q+1} są wyznaczane z dokładnością rzędu q , zaś schemat (26) jest rzędu $q + 1$. Całkowity rząd schematu EBDF jest zatem równy $q + 1$. Okazuje się jednak, iż schematy tego typu mają większy obszar stabilności absolutnej niż standardowe schematy BDF tego samego rzędu. Współczynniki schematów EBDF można znaleźć w oryginalnym opracowaniu [2, str. 238]. Wartości kątów α_{max} z tego

opracowania okazały się jednak niepoprawne. Obliczenia własne, znajdujące potwierdzenie w [11, str. 289], dały w wyniku kąty, które można znaleźć w tabeli 2.

Zauważmy, że każdy z trzech kroków algorytmu wymaga rozwiązania nieliniowego układu równań w celu uzyskania żądanego rozwiązania. Efektywnym sposobem (porównaj [10]) rozwiązywania takich układów jest metoda iteracyjna Newtona. Niestety wymaga ona dokonania rozkładu LU odpowiedniej macierzy Jacobiego. W algorytmie EBDF wykorzystuje się następujące iteracje:

$$\begin{aligned}
 (I - hb_q J_k)(u_{k+q}^{(p)} - u_{k+q}^{(p-1)}) &= -u_{k+q}^{(p-1)} + \sum_{j=0}^{q-1} a_j x_{k+j} \\
 &\quad + hb_q f(t_{k+q}, u_{k+q}^{(p-1)}), \\
 (I - hb_q J_{k+1})(u_{k+q+1}^{(p)} - u_{k+q+1}^{(p-1)}) &= -u_{k+q+1}^{(p-1)} + a_{q-1} u_{k+q} \\
 &\quad + \sum_{j=0}^{q-2} a_j x_{k+j+1} + hb_q f(t_{k+q+1}, u_{k+q+1}^{(p-1)}), \\
 (I - h\beta_q \bar{J}_k)(x_{k+q}^{(p)} - x_{k+q}^{(p-1)}) &= -x_{k+q}^{(p-1)} + \sum_{j=0}^{q-1} \alpha_j x_{k+j} \\
 &\quad + h\beta_q f(t_{k+q}, x_{k+q}^{(p-1)}) + h\beta_{q+1} f(t_{k+q+1}, u_{k+q+1}),
 \end{aligned}$$

gdzie J_k jest macierzą Jacobiego równania predyktora w punkcie t_{k+q} , J_{k+1} macierzą Jacobiego równania predyktora w punkcie t_{k+q+1} , $\bar{J}_k$ macierzą Jacobiego równania korektora w punkcie t_{k+q} , zaś $x_j^{(l)}$ i $u_j^{(l)}$ są przybliżeniami odpowiednio x_j i u_j w l -tej iteracji. Na ogół przyjmuje się $J_k = J_{k+1}$ i nie zmienia się tych macierzy tak długo jak długo procesy iteracyjne pozostają zbieżne. Zatem wystarczy dokonać rozkładu LU dwóch różnych macierzy Jacobiego – jednego dla równań predyktora (23) – (24), jednego dla równania korektora (26).

Warto dodać, że w celu uniknięcia dużych nakładów obliczeniowych związanych z dwukrotnym rozkładem LU w 1983 roku J. R. Cash [4] oraz [3] zaproponował modyfikację metody EBDF, tak zwaną MEBDF, czyli *Modified Extended Backward Differentiation Formulae*.

Tablica 2: Dane dotyczące rozszerzonych formuł wstecznego różniczkowania (EBDF) dla $q = 1 \dots 8$.

q	1	2	3	4	5	6	7	8
rzęd p	2	3	4	5	6	7	8	9
α_{max}	90°	90°	90°	87.61°	80.21°	67.73°	48.82°	19.98°
C_{p+1}	$\frac{5}{12}$ 0.4167	$\frac{17}{138}$ 0.1232	$\frac{111}{1970}$ 0.0563	$\frac{394}{12505}$ 0.0315	$\frac{690}{34811}$ 0.0198	$\frac{2515}{186578}$ 0.0135	$\frac{12145}{1253418}$ 0.0097	$\frac{270172}{37211841}$ 0.0073

3. Uogólnienie formuł wstecznego różniczkowania

W formule BDF do wyznaczenia punktu x_{k+q} wykorzystywało się punkty $x_k, x_{k+1}, \dots, x_{k+q-1}$. W schemacie EBDF trzeba było dodatkowo wyliczyć przy pomocy predyktora punkt u_{k+q+1} będący przybliżeniem punktu x_{k+q+1} . Następowало w tym przypadku wybiegnięcie w przyszłość, ale tylko o jeden krok. Zostaną przedstawione własności schematów, w których korzysta się z punktów wybiegających w przeszłość o więcej niż jeden krok. Oznaczmy zatem przez q_1 liczbę kroków predyktora, przez q_2 liczbę kroków korektora, zaś przez r liczbę kroków w przyszłość. Niech $q = \max \{q_1, q_2\}$. Nowe schematy składać się będą z formuł:

$$\text{Pre-predyktor: } x_{k+q} = \sum_{j=0}^q \nabla^j x_{k+q-1}, \quad (27)$$

$$\text{Predyktor: } x_{k+q_1} = h b_{q_1} f_{k+q_1} - \sum_{j=0}^{q_1-1} a_j x_{k+j}, \quad (28)$$

$$\text{Korektor: } x_{k+q_2} = h \sum_{j=q_2}^{q_2+r} \beta_j f_{k+j} - \sum_{j=0}^{q_2-1} a_j x_{k+j}. \quad (29)$$

Na obliczenia będzie składać się $r+1$ wywołań predyktora oraz jedna iteracja schematu korektora. Wprowadźmy następujące oznaczenia na nowe schematy:

- $B^r DF$ - q -krokowa formuła wstecznego różniczkowania z r krokami w przyszłość. Spośród współczynników β tylko

$\beta_q, \beta_{q+1}, \dots, \beta_{q+r}$ są niezerowe. W szczególności B^0DF to zwykła BDF.

- EB^rDF - schemat typu predyktor-korektor, w którym używamy formuły B^0DF do obliczenia punktów $u_{k+q}, u_{k+q+1}, \dots, u_{k+q+r}$ (predyktor), a następnie formułę B^rDF do wyznaczenia punktu x_{k+q} (korektor). Predyktora używamy zatem $r+1$ razy, a korektora tylko raz. Liczba kroków predyktora i korektora mogą być różne. W szczególności EB^1DF to zwykła EBDf.

Algorytm oparty na uogólnionych schematach wstecznego różniczkowania EB^rDF można przedstawić w następujących krokach:

Krok 0. Użyj schematu predyktora BDF do obliczenia u_{k+q} :

$$u_{k+q} = hb_{q_1} f(t_{k+q}, u_{k+q}) - \sum_{j=0}^{q_1-1} a_j x_{k+q-q_1+j},$$

przyjmując za przybliżenie początkowe punkt u_{k+q+1} z poprzedniej iteracji.

Krok 1. Użyj schematu predyktora BDF do obliczenia u_{k+q+1} :

$$u_{k+q+1} = hb_{q_1} f(t_{k+q+1}, u_{k+q+1}) - \sum_{j=0}^{q_1-2} a_j x_{k+q-q_1+j+1} - a_{q_1-1} u_{k+q},$$

przyjmując za przybliżenie początkowe punkt u_{k+q+2} z poprzedniej iteracji.

...

Krok $(r-1)$. Użyj schematu predyktora BDF do obliczenia $u_{k+q+r-1}$:

$$u_{k+q+r-1} = hb_{q_1} f(t_{k+q+r-1}, u_{k+q+r-1}) - \sum_{j=0}^{q_1-r} a_j x_{k+q-q_1+j+r-1} - \sum_{j=q_1-r+1}^{q_1-1} a_j u_{k+q-q_1+j+r-1},$$

przyjmując za przybliżenie początkowe punkt u_{k+q+r} z poprzedniej iteracji.

Krok r . Użyj schematu predyktora BDF do obliczenia u_{k+q+r} :

$$u_{k+q+r} = hb_{q_1} f(t_{k+q+r}, u_{k+q+r}) - \sum_{j=0}^{q_1-r-1} a_j x_{k+q-q_1+j+r} - \sum_{j=q_1-r}^{q_1-1} a_j u_{k+q-q_1+j+r},$$

przyjmując za przybliżenie początkowe punkt

$$u_{k+q+r}^{(0)} = \sum_{j=0}^q \nabla^j u_{k+q+r-1},$$

obliczony na podstawie punktów

$$x_{k+r-1}, x_{k+r}, \dots, x_{k+q-1}, u_{k+q}, u_{k+q+1}, \dots, u_{k+q+r-2}, u_{k+q+r-1}.$$

Krok $(r+1)$. Użyj schematu korektora $B^r DF$ do uzyskania przybliżonego rozwiązania w punkcie t_{k+q} :

$$x_{k+q} = h\beta_{q_2} f(t_{k+q}, x_{k+q}) + h \sum_{j=q_2+1}^{q_2+r} \beta_j f(u_{k+q-q_2+j}) - \sum_{j=0}^{q_2-1} \alpha_j x_{k+q-q_2+j},$$

przyjmując za przybliżenie początkowe punkt u_{k+q+r} z danej iteracji.

W schemacie $EB^r DF$ predyktorem jest zwykła metoda wstecznego różniczkowania (BDF). Współczynniki dla takich formuł dla $q = 1, 2, 3, 4, 5, 6$ zostały podane w tabeli 1. Używanie samych formuł BDF dla $q > 6$ do numerycznego rozwiązywania równań różniczkowych zwyczajnych nie ma sensu ze względu na ich niestabilność. Mają one jednak duże znaczenie jako formuły predyktora między innymi w schematach EBDF, MEbDF i podanych w tej pracy. Dla q większych od 6 współczynniki można znaleźć w tabeli 3.

Korektor w nowych schematach jest q -krokową formułą BDF z r krokami w przyszłość. Współczynniki takiej metody wyznaczone są w taki sposób, by jej rząd był jak najwyższy. Dla dwóch i trzech kroków w przyszłość współczynniki zamieszczone są w tabelach 4 i 5.

Ważnym wyznacznikiem jakości schematu, poza jego stabilnością, jest jego rząd. Dla pojedynczych formuł badanie rzędu jest algorytmicznie proste. Trudniejsze jest badanie rzędu schematów bardziej złożonych. O rzędzie nowych schematów $EB^r DF$ wychodzących r kroków w przyszłość mówi:

Twierdzenie 5

Niech predyktor będzie q_1 -krokową formułą BDF, korektor zaś q_2 -krokową formułą BDF z r krokami w przyszłość. Predyktor jest rzędu q_1 ze stałą błędu C_0 , czyli dla wartości $x_k, x_{k+1}, \dots, x_{k+q_1-1}$ równym rozwiązaniu rzeczywistemu zachodzi:

$$x(t_{k+q_1}) - x_{k+q_1} = C_0 h^{q_1+1} x^{(q_1+1)}(t_{k+q_1}) + O(h^{q_1+2}).$$

Korektor jest rzędu q_2 ze stałą błędu D_0 , czyli dla wartości $x_k, x_{k+1}, \dots, x_{k+q_2-1}$ oraz $x_{k+q_2+1}, x_{k+q_2+2}, \dots, x_{k+q_2+r}$ równym rozwiązaniu rzeczywistemu zachodzi:

$$x(t_{k+q_2}) - x_{k+q_2} = D_0 h^{q_2+r+1} x^{(q_2+r+1)}(t_{k+q_2}) + O(h^{q_2+r+2}).$$

Wówczas rząd całego schematu typu predyktor-korektor $EB^r DF$ jest równy $\min\{q_1 + 1, q_2 + r\}$.

Dowód:

Niech $q = \max\{q_1, q_2\}$. Załóżmy, że wartości $x_k, x_{k+1}, \dots, x_{k+q-1}$ są równe rzeczywistemu rozwiązaniu $x(t_k), x(t_{k+1}), \dots, x(t_{k+q-1})$. Wykorzystując fakt, że predyktor jest rzędu q_1 i postępując zgodnie z algorytmem dostajemy:

Krok 0.

$$x(t_{k+q}) - u_{k+q} = C_0 h^{q_1+1} x^{(q_1+1)}(t_{k+q}) + O(h^{q_1+2}).$$

Krok 1.

$$\begin{aligned} x(t_{k+q+1}) - u_{k+q+1} &= x(t_{k+q+1}) - hb_{q_1} f(t_{k+q+1}, u_{k+q+1}) \\ &\quad + \sum_{j=0}^{q_1-2} a_j x_{k+q-q_1+j+1} + a_{q_1-1} u_{k+q} \\ &= x(t_{k+q+1}) - hb_{q_1} f(t_{k+q+1}, u_{k+q+1}) \\ &\quad + \sum_{j=0}^{q_1-2} a_j x_{k+q-q_1+j+1} + a_{q_1-1} x(t_{k+q}) \\ &\quad - a_{q_1-1} (x(t_{k+q}) - u_{k+q}) \\ &= C_0 h^{q_1+1} x^{(q_1+1)}(t_{k+q+1}) + O(h^{q_1+2}) \\ &\quad - a_{q_1-1} C_0 h^{q_1+1} x^{(q_1+1)}(t_{k+q}) + O(h^{q_1+2}) \\ &= (C_0 - a_{q_1-1} C_0) h^{q_1+1} x^{(q_1+1)}(t_{k+q}) + O(h^{q_1+2}) \\ &= C_1 h^{q_1+1} x^{(q_1+1)}(t_{k+q}) + O(h^{q_1+2}). \end{aligned}$$

Przedostatnia z równości wynika z faktu:

$$x^{(q_1+1)}(t_{n+j}) = x^{(q_1+1)}(t_n + jh) = x^{(q_1+1)}(t_n) + O(h^{q_1+2}).$$

W kolejnych krokach algorytmu postępujemy analogicznie. W kroku drugim przykładowo dostajemy:

$$x(t_{k+q+2}) - u_{k+q+2} = x(t_{k+q+2}) - hb_{q_1} f(t_{k+q+2}, u_{k+q+2})$$

$$\begin{aligned}
& + \sum_{j=0}^{q_1-3} a_j x_{k+q-q_1+j+2} + \sum_{j=q_1-2}^{q_1-1} a_j u_{k+q-q_1+j+2} \\
& = x(t_{k+q+2}) - hb_{q_1} f(t_{k+q+2}, u_{k+q+2}) + \sum_{j=0}^{q_1-3} a_j x_{k+q-q_1+j+2} \\
& \quad + \sum_{j=q_1-2}^{q_1-1} a_j x(t_{k+q-q_1+j+2}) - \sum_{j=q_1-2}^{q_1-1} a_j (x(t_{k+q-q_1+j+2}) - u_{k+j+2}) \\
& = C_0 h^{q_1+1} x^{(q_1+1)}(t_{k+q+2}) + O(h^{q_1+2}) \\
& \quad - \sum_{j=q_1-2}^{q_1-1} a_j \left(C_j h^{q_1+1} x^{(q_1+1)}(t_{k+q+j}) + O(h^{q_1+2}) \right) \\
& = (C_0 - a_{q_1-2} C_0 - a_{q_1-1} C_1) h^{q_1+1} x^{(q_1+1)}(t_{k+q}) + O(h^{q_1+2}) \\
& = C_2 h^{q_1+1} x^{(q_1+1)}(t_{k+q}) + O(h^{q_1+2}).
\end{aligned}$$

W kroku r -tym natomiast

$$\begin{aligned}
x(t_{k+q+r}) - u_{k+q+r} & = x(t_{k+q+r}) - hb_{q_1} f(t_{k+q+r}, u_{k+q+r}) \\
& \quad + \sum_{j=0}^{q_1-r-1} a_j x_{k+q-q_1+j+r} + \sum_{j=q_1-r}^{q_1-1} a_j u_{k+q-q_1+j+r} \\
& = x(t_{k+q+r}) - hb_{q_1} f(t_{k+q+r}, u_{k+q+r}) + \sum_{j=0}^{q_1-r-1} a_j x_{k+q-q_1+j+r} \\
& \quad + \sum_{j=q_1-r}^{q_1-1} a_j x(t_{k+q-q_1+j+r}) - \sum_{j=q_1-r}^{q_1-1} a_j (x(t_{k+q-q_1+j+r}) - u_{k+q-q_1+j+r}) \\
& = C_0 h^{q_1+1} x^{(q_1+1)}(t_{k+q+r}) + O(h^{q_1+2}) \\
& \quad - \sum_{j=q_1-r}^{q_1-1} a_j \left(C_j h^{q_1+1} x^{(q_1+1)}(t_{k+q+j}) + O(h^{q_1+2}) \right) \\
& = \left(C_0 - \sum_{j=q_1-r}^{q_1-1} a_j \right) h^{q_1+1} x^{(q_1+1)}(t_{k+q+r}) + O(h^{q_1+2}) \\
& \quad + O(h^{q_1+2}) = C_r h^{q_1+1} x^{(q_1+1)}(t_{k+q+r}) + O(h^{q_1+2}).
\end{aligned}$$

W ostatnim $(r+1)$ -szym kroku

$$x(t_{k+q}) - x_{k+q} = x(t_{k+q}) - h\beta_{q_2} f_{k+q}$$

$$\begin{aligned}
& -h \sum_{j=q_2+1}^{q_2+r} \beta_j f(t_{k+q-q_2+j}, u_{k+q-q_2+j}) + \sum_{j=0}^{q_2-1} \alpha_j x_{k+q-q_2+j} \\
& = x(t_{k+q}) - h\beta_{q_2} f_{k+q} - h \sum_{j=q_2+1}^{q_2+r} \beta_j f(t_{k+q-q_2+j}, x(t_{k+q-q_2+j})) \\
& \quad + \sum_{j=0}^{q_2-1} \alpha_j x_{k+q-q_2+j} + h \sum_{j=q_2+1}^{q_2+r} \beta_j (f(t_{k+q-q_2+j}, x(t_{k+q-q_2+j})) \\
& \quad - f(t_{k+q-q_2+j}, u_{k+q-q_2+j})) \\
& = D_0 h^{q_2+r+1} x^{(q_2+r+1)}(t_{k+q}) + O(h^{q_2+r+2}) \\
& \quad + h \sum_{j=q_2+1}^{q_2+r} \beta_j (f(t_{k+q-q_2+j}, x(t_{k+q-q_2+j})) - f(t_{k+q-q_2+j}, u_{k+q-q_2+j})).
\end{aligned}$$

Wiadomo jednak, że

$$\begin{aligned}
f(t_{k+q+j}, x(t_{k+q+j})) - f(t_{k+q+j}, u_{k+q+j}) & = E_{q+j} (x(t_{k+q+j}) - u_{k+q+j}) \\
& = E_{q+j} C_j h^{q_1+1} x^{(q_1+1)}(t_{k+q}) + O(h^{q_1+2}) \\
& = \overline{C}_j h^{q_1+1} x^{(q_1+1)}(t_{k+q}) + O(h^{q_1+2}),
\end{aligned}$$

dla pewnych stałych E_j i $j = 0, 1, \dots, r$. Stąd

$$\begin{aligned}
x(t_{k+q}) - x_{k+q} & = D_0 h^{q_2+r+1} x^{(q_2+r+1)}(t_{k+q}) + O(h^{q_2+r+2}) \\
& \quad + h \sum_{j=q_2+1}^{q_2+r} \left(\beta_j \overline{C}_{j-q_2} h^{q_1+1} x^{(q_1+1)}(t_{k+q}) + O(h^{q_1+2}) \right) \\
& = D_0 h^{q_2+r+1} x^{(q_2+r+1)}(t_{k+q}) \\
& \quad + h^{q_1+2} x^{(q_1+1)}(t_{k+q}) \sum_{j=q_2+1}^{q_2+r} \beta_j \overline{C}_{j-q_2} + O(h^{q_1+3}) + O(h^{q_2+r+2}) \\
& = \text{const}_1 \cdot h^{q_2+r+1} + \text{const}_2 * h^{q_1+2} + O(h^{q_1+3}) + O(h^{q_2+r+2}) \\
& = \text{const} \cdot h^{\min\{q_1+2, q_2+r+1\}} + O(h^{\min\{q_1+3, q_2+r+2\}}),
\end{aligned}$$

co dowodzi, że schemat jest rzędu $\min\{q_1 + 1, q_2 + r\}$.

□

W szczególności dla $q_1 = q_2 + r - 1$ schemat jest rzędu $q_2 + r$, zaś dla $q_1 = q_2$ rzędu $q_1 + 1$. Dla $r = 1$ oba powyższe przypadki sprowadzają się do jednego $q = q_1 = q_2$ rozważanego przez Casha.

Poza rzędem schematu pozostałe kryteria służące do porównywania jakości schematów numerycznego rozwiązywania równań różniczkowych zwyczajnych są ściśle związane z ich obszarem stabilności absolutnej. Wyznaczenie takiego obszaru dla schematów typu predyktorkorektor jest znacznie trudniejszym zadaniem niż w przypadku wielokrokowych schematów liniowych [1]. Obliczone kąty z $A(\alpha)$ -stabilności podane są w tabelach 6 i 7. Przykład konturu obszaru stabilności absolutnej metody E^2BDF przedstawiony jest na rysunku 3.

4. Podsumowanie

Jako parametry jakości schematu przyjąłem w swojej pracy rząd schematu oraz kąt α_{max} z $A(\alpha)$ -stabilności. Wszystkie dane dotyczące opracowanych przeze mnie uogólnionych formuł wstecznego różniczkowania są zebrane w tabelach 6 i 7. Są one potwierdzeniem tezy mówiącej, iż zwiększenie rzędu schematu, czyli jego dokładności, jest związane ze zmniejszeniem obszaru jego stabilności absolutnej. Na szczęście parametry istniejących formuł EBDF i MEBDF autorstwa Casha można poprawić. Porównania tych formuł z ich uogólnionymi wersjami są na to dowodem.

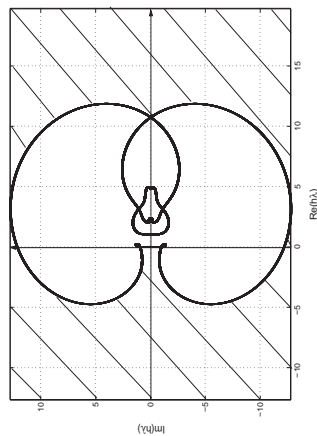
Schematy EB^rDF dla $q_1 = q_2$ (tabela 6)

Rząd tych schematów jest równy $q_2 + 1$, czyli nie zależy od liczby kroków w przyszłość. Dla $r = 1$ dostajemy podane przez Casha formuły EBDF. Ich uogólnienie poprzez „wyjście w przyszłość” nie poprawia rzędu, ale znacznie zwiększa ich obszar stabilności absolutnej. Warto zwrócić szczególną uwagę na przypadek $r = 2$, gdyż kąty α_{max} są tutaj największe. Wraz ze wzrostem liczby kroków w przyszłość obszary stabilności absolutnej zaczynają maleć, dla $r \leq 4$ okazują się być nawet mniejsze od podstawowej EBDF. Ze względu na nakłady obliczeniowe z nimi związane, to jest $r + 1$ wywołań predyktora, nie opłaca z nich korzystać w praktyce. Najbardziej wydajny wydaje się schemat EB^2DF . Jest on $A(0)$ -stabilny dla $q_2 = 1, 2, \dots, 9$ podczas gdy zwykła formuła EBDF nie jest dla $q_2 = 9$ $A(\alpha)$ -stabilna dla żadnego α . Zwiększone obszary stabilności absolutnej w większości przypadków będą rekompensować zwiększenie nakładów obliczeniowych, które w tym przypadku sprowadzają się do skorzystania tylko jeden raz więcej z formuły predyktora. Najdokładniejszy z $A(0)$ -stabilnych schematów ($q_2 = 9$) ma rząd równy 10.

Schematy EB^rDF dla $q_1 = q_2 + r - 1$ (tabela 7)

Tablica 3: Współczynniki formuł wstecznego różniczkowania bez kroków w przyszłość ($r = 0$) dla $q = 7 \dots 10$.

q	α_{10}	α_9	α_8	α_7	α_6	α_5	α_4	α_3	α_2	α_1	α_0	β_q
7				1	$-\frac{2940}{1089}$	$\frac{4410}{1089}$	$-\frac{4900}{1089}$	$\frac{3675}{1089}$	$-\frac{1764}{1089}$	$\frac{490}{1089}$	$-\frac{60}{1089}$	$\frac{420}{1089}$
8			1	$-\frac{6720}{2283}$	$\frac{11760}{2283}$	$-\frac{15680}{2283}$	$\frac{14700}{2283}$	$-\frac{9408}{2283}$	$\frac{3920}{2283}$	$-\frac{960}{2283}$	$\frac{105}{2283}$	$\frac{840}{2283}$
9		1	$-\frac{22680}{7129}$	$\frac{45360}{7129}$	$-\frac{70560}{7129}$	$\frac{79380}{7129}$	$-\frac{63504}{7129}$	$\frac{35280}{7129}$	$-\frac{12960}{7129}$	$\frac{2835}{7129}$	$-\frac{280}{7129}$	$\frac{2520}{7129}$
10	1	$-\frac{25200}{7381}$	$\frac{56700}{7381}$	$-\frac{100800}{7381}$	$\frac{132300}{7381}$	$-\frac{127008}{7381}$	$\frac{88200}{7381}$	$-\frac{43200}{7381}$	$\frac{14175}{7381}$	$-\frac{2800}{7381}$	$\frac{252}{7381}$	$\frac{2520}{7381}$



Rysunek 3: Obszar stabilności absolutnej schematu E^2BDF ($r = 2$) dla $q_1 = q_2 = 9$ (zakreskowany).

Tablica 4: Współczynniki formuł wstecznego różniczkowania z dwoma krokami w przyszłość ($r = 2$) dla $q = 1 \dots 9$.

q	1	2	3	4	5	6	7	8	9
α_9									1
α_8								1	$-\frac{4615595089560}{2213203583279}$
α_7							1	$-\frac{116781210560}{59361109467}$	$\frac{5010293088360}{2213203583279}$
α_6						1	$-\frac{5285658420}{2863446891}$	$\frac{111193894840}{59361109467}$	$-\frac{4876603044912}{2213203583279}$
α_5					1	$-\frac{33101560}{19236471}$	$\frac{4326228865}{2863446891}$	$-\frac{92876972384}{59361109467}$	$\frac{3752443035684}{2213203583279}$
α_4				1	$-\frac{5360100}{3368607}$	$\frac{22655475}{19236471}$	$-\frac{3015526150}{2863446891}$	$\frac{59598234500}{59361109467}$	$-\frac{2181757246704}{2213203583279}$
α_3			1	$-\frac{326384}{224133}$	$\frac{2947300}{3368607}$	$-\frac{12654400}{19236471}$	$\frac{1549317525}{2863446891}$	$-\frac{27734617792}{59361109467}$	$\frac{920521508760}{2213203583279}$
α_2		1	$-\frac{10539}{8018}$	$\frac{135324}{224133}$	$-\frac{1237200}{3368607}$	$\frac{4881025}{19236471}$	$-\frac{541042516}{2863446891}$	$\frac{8779273160}{59361109467}$	$-\frac{265510010160}{2213203583279}$
α_1	1	$-\frac{192}{165}$	$\frac{2934}{8018}$	$-\frac{37968}{224133}$	$\frac{318525}{3368607}$	$-\frac{1137096}{19236471}$	$\frac{114221695}{2863446891}$	$-\frac{1688593440}{59361109467}$	$\frac{46810392237}{2213203583279}$
α_0	$-\frac{12}{12}$	$\frac{27}{165}$	$-\frac{413}{8018}$	$\frac{4895}{224133}$	$-\frac{37132}{3368607}$	$\frac{120085}{19236471}$	$-\frac{10987890}{2863446891}$	$\frac{148882209}{59361109467}$	$-\frac{3805316984}{2213203583279}$
β_q	$\frac{23}{12}$	$\frac{197}{165}$	$\frac{7503}{8018}$	$\frac{179028}{224133}$	$\frac{2398860}{3368607}$	$\frac{12534180}{19236471}$	$\frac{1736552580}{2863446891}$	$\frac{33910320360}{59361109467}$	$\frac{1201469398920}{2213203583279}$
β_{q+1}	$-\frac{16}{12}$	$-\frac{76}{165}$	$-\frac{1926}{8018}$	$-\frac{33696}{224133}$	$-\frac{351600}{3368607}$	$-\frac{1490400}{19236471}$	$-\frac{172548600}{2863446891}$	$-\frac{2879318400}{59361109467}$	$-\frac{88716358080}{2213203583279}$
β_{q+2}	$\frac{5}{12}$	$\frac{17}{165}$	$\frac{333}{8018}$	$\frac{4728}{224133}$	$\frac{41400}{3368607}$	$\frac{150900}{19236471}$	$\frac{15302700}{2863446891}$	$\frac{226944480}{59361109467}$	$\frac{6287531040}{2213203583279}$

Tablica 5: Współczynniki formuł wstecznego różniczkowania z trzema krokami w przyszłość ($r = 3$) dla $q = 1 \dots 9$.

q	1	2	3	4	5	6	7	8	9
α_9									1
α_8							1	1	$-\frac{16132872236287590}{8559635851294723}$
α_7							1	$-\frac{5131079849799120}{2869794578568697}$	$\frac{14388035521648824}{8559635851294723}$
α_6						1	$-\frac{20330180452365}{12039969061391}$	$\frac{4017142624977048}{2869794578568697}$	$-\frac{11854134190464552}{8559635851294723}$
α_5				1	$-\frac{1219633993530}{76893354389}$	$\frac{13694547816663}{12039969061391}$	$\frac{12039969061391}{12039969061391}$	$-\frac{2841494053011152}{2869794578568697}$	$\frac{7900657484053572}{8559635851294723}$
α_4				1	$-\frac{2088183575}{1410754417}$	$\frac{68771743275}{76893354389}$	$-\frac{8087571303175}{12039969061391}$	$\frac{1579756079250300}{2869794578568697}$	$-\frac{4049806984063848}{8559635851294723}$
α_3			1	$-\frac{59972224}{43776401}$	$\frac{947132700}{1410754417}$	$-\frac{32564326600}{76893354389}$	$\frac{3601231618425}{12039969061391}$	$-\frac{648250344244272}{2869794578568697}$	$\frac{1527505437696840}{8559635851294723}$
α_2		1	$-\frac{773577}{616669}$	$\frac{20541276}{43776401}$	$-\frac{337270300}{1410754417}$	$\frac{10889936475}{76893354389}$	$-\frac{1109181466491}{12039969061391}$	$\frac{183469305218920}{2869794578568697}$	$-\frac{398300525356824}{8559635851294723}$
α_1	1	$-\frac{6456}{5703}$	$\frac{178227}{616669}$	$-\frac{4892864}{43776401}$	$\frac{75314675}{1410754417}$	$-\frac{2238118014}{76893354389}$	$\frac{209399656865}{12039969061391}$	$-\frac{31905105043248}{2869794578568697}$	$\frac{64067998773177}{8559635851294723}$
α_0	$-\frac{24}{24}$	$\frac{753}{5703}$	$-\frac{21319}{616669}$	$\frac{547411}{43776401}$	$-\frac{7747917}{1410754417}$	$\frac{211404005}{76893354389}$	$-\frac{18214931313}{12039969061391}$	$\frac{2566764082827}{2869794578568697}$	$-\frac{4788357294322}{8559635851294723}$
β_q	$\frac{55}{24}$	$\frac{8018}{5703}$	$\frac{672399}{616669}$	$\frac{40423284}{43776401}$	$\frac{1154188260}{1410754417}$	$\frac{57268937820}{76893354389}$	$\frac{8310555325380}{12039969061391}$	$\frac{1859091009954360}{2869794578568697}$	$\frac{5252768510086440}{8559635851294723}$
β_{q+1}	$-\frac{59}{24}$	$-\frac{4827}{5703}$	$-\frac{271053}{616669}$	$-\frac{11992752}{43776401}$	$-\frac{267287400}{1410754417}$	$-\frac{10776128400}{76893354389}$	$-\frac{1308128994900}{12039969061391}$	$-\frac{250248865800600}{2869794578568697}$	$-\frac{615193133617440}{8559635851294723}$
β_{q+2}	$\frac{37}{24}$	$\frac{2172}{5703}$	$\frac{94419}{616669}$	$\frac{3393672}{43776401}$	$\frac{63514200}{1410754417}$	$\frac{2202855300}{76893354389}$	$\frac{234303344100}{12039969061391}$	$\frac{39844276547040}{2869794578568697}$	$\frac{88087675465440}{8559635851294723}$
β_{q+3}	$-\frac{9}{24}$	$-\frac{413}{5703}$	$-\frac{14685}{616669}$	$-\frac{445584}{43776401}$	$-\frac{7205100}{1410754417}$	$-\frac{219757800}{76893354389}$	$-\frac{20843509260}{12039969061391}$	$-\frac{3196466266560}{2869794578568697}$	$-\frac{6432026665680}{8559635851294723}$

Tablica 6: Wartości kąta α_{max} schematów EB^rDF dla $q_1 = q_2$ (rzęd równy $q_2 + 1$).

q_2	1	2	3	4	5	6	7	8	9
$r = 1$	90°	90°	90°	87.61°	80.22°	67.73°	48.82°	19.98°	—
$r = 2$	90°	90°	90°	88.44°	83.32°	75.06°	63.37°	47.27°	24.31°
$r = 3$	90°	90°	89.97°	86.83°	80.46°	71.30°	59.13°	43.15°	21.08°

Tablica 7: Wartości kąta α_{max} schematów EB^rDF dla $q_1 = q_2 + r - 1$ (rzęd równy $q_2 + r$).

q_2	1	2	3	4	5	6	7	8	9
$r = 1$	90°	90°	90°	87.61°	80.22°	67.73°	48.82°	19.98°	—
$r = 2$	90°	90°	89.33°	85.37°	78.48°	68.77°	55.77°	38.23°	12.77°
$r = 3$	78.72°	77.02°	71.09°	62.04°	49.78°	33.35°	9.01°	—	—

W przeciwieństwie do przypadku opisanego powyżej rząd tych schematów zależy od liczby kroków przyszłość i jest równy $q_2 + r$. Jak poprzednio dla $r = 1$ dostajemy zwykle formuły EBDF. Począwszy od $r = 3$ obszary stabilności absolutnej są w uogólnionych schematach mniejsze niż odpowiadające im obszary dla $r = 1$. Interesujący okazuje się przypadek dla $r = 2$. Dla $q_2 \geq 6$, czyli dla schematów z wyższymi rzędami, kąty α_{max} są wyraźnie większe. Nawet dla $q_2 = 9$ dostajemy schemat A(0)-stabilny i to rzędu 11. Ma on jednak mniejszy obszar stabilności absolutnej od odpowiadającej formuły EB^2DF przy takiej samej liczbie kroków predyktora i korektora.

Spośród schematów typu EB^rDF praktyczne zastosowanie mogą znaleźć schematy EB^2DF dla $q_1 = q_2$ lub $q_1 = q_2 + 1$. Charakteryzują się one bowiem zwiększonym, w porównaniu z podstawowymi schematami EBDF i MEBDF, obszarem stabilności absolutnej, a w przypadku $q_1 = q_2 + 1$ nawet zwiększonym rzędem.

Bibliografia

- [1] Bernardelli M. (2003); *Algorytmy równoległe z użyciem formuł wstecznego różniczkowania*, Praca magisterska na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego, Warszawa.
- [2] Cash J. R. (1980); *On the Integration of Stiff Systems of O.D.E.s Using Extended Backward Differentiation Formulae*, „Numerische Mathematik” 34, pp. 235-246.
- [3] Cash J. R. (1983); *The Integration of Stiff Initial Value Problems in ODEs Using Modified Extended Backward Differentiation Formulae*, „Computers And Mathematics with Applications” Vol. 9 No. 5, pp. 645-657.
- [4] Cash J. R. (2000); *Modified Extended Backward Differentiation Formulae for the Numerical Solution of Stiff Initial Value Problems in ODEs and DAEs*, „Journal of Computational and Applied Mathematics” 125, pp. 117-130.
- [5] Creedon D.M., Miller J.J.H. (1975); *The Stability Properties of q-step Backward Difference Schemes*, „BIT” 15, pp. 244-249
- [6] Dahlquist G. (1956); *Convergence and Stability in the Numerical Integration of Ordinary Differential Equations*, „Math. Scand.” 4, pp. 33-53
- [7] Dahlquist G. (1963); *A Special Stability Problem for Linear Multistep Methods*, „BIT” 3, pp. 27 - 43
- [8] Dahlquist G., Björck Å. (1974); *Numerical Methods*, Prentice-Hall, Inc.
- [9] Dryja M., Jankowscy J. i M. (1988); *Przegląd metod i algorytmów numerycznych. Część 2*, Wydawnictwa Naukowo-Techniczne, Warszawa
- [10] Frank J. E., Houwen P. J. (2001); *Parallel Iteration of the Extended Backward Differentiation Formulas*, „IMA Journal of Numerical Analysis” 21, pp. 367 - 385.

- [11] Hairer E., Wanner G. (1991); *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, Springer-Verlag, Berlin Heidelberg
- [12] Krupowicz A. (1986); *Metody numeryczne zagadnień początkowych równań różniczkowych zwyczajnych*, Państwowe Wydawnictwo Naukowe, Warszawa
- [13] Lambert J. D. (1973); *Computational Methods in Ordinary Differential Equations*, John Wiley & Sons
- [14] Palczewski A. (1999); *Równania Różniczkowe Zwyczajne. Teoria i metody numeryczne z wykorzystaniem komputero-
wego systemu obliczeń symbolicznych*, Wydawnictwa Naukowo-Techniczne, Warszawa
- [15] Süli E. (2001); *Numerical Solution of Ordinary Differential Equations*, <http://web.comlab.ox.ac.uk/oucl/work/andre.suli>
- [16] Widlund O. B. (1967); *A Note on Unconditionally Stable Linear Multistep Methods* „BIT” 7, pp. 65-70.