



WSTĘP DO METOD NUMERYCZNYCH (WNUM)

RÓWNANIA RÓŻNICZKOWE

ZAŁOŻENIA PODSTAWOWE

W ramach wykładu będziemy zajmować się metodami rozwiązywania *zagadnienia początkowego* (zagadnienia Cauchy'ego), które zdefiniowane jest w następujący sposób:

Poszukiwany jest zbiór M funkcji $y_1(x), y_2(x), \dots, y_M(x)$ które spełniają układ równań różniczkowych:

$$\begin{array}{l}
 y_m(x) \\
 \text{-- szukane funkcje}
 \end{array}
 \left\{
 \begin{array}{l}
 \frac{dy_1}{dx} = f_1(x, y_1, \dots, y_M) \\
 \vdots \\
 \frac{dy_M}{dx} = f_M(x, y_1, \dots, y_M)
 \end{array}
 \right.
 \begin{array}{l}
 \text{Jakieś} \\
 \text{funkcje} \\
 \text{potencjalnie} \\
 \text{„składające} \\
 \text{się” z } y_m(x) \\
 \text{oraz } x
 \end{array}$$

Funkcje te będziemy wyznaczać w oparciu o założone (dane) warunki początkowe jakie funkcje $y_m(x_0) = y_{m,0}$ dla każdej z funkcji $y_m(x)$ $m = 1, 2, \dots, M$

Przykład:

Dynamika układu
Nieliniowego:

$$\frac{d q(t)}{dt} = \frac{1}{R} (e(t) - u(t))$$

$y_m(x)$

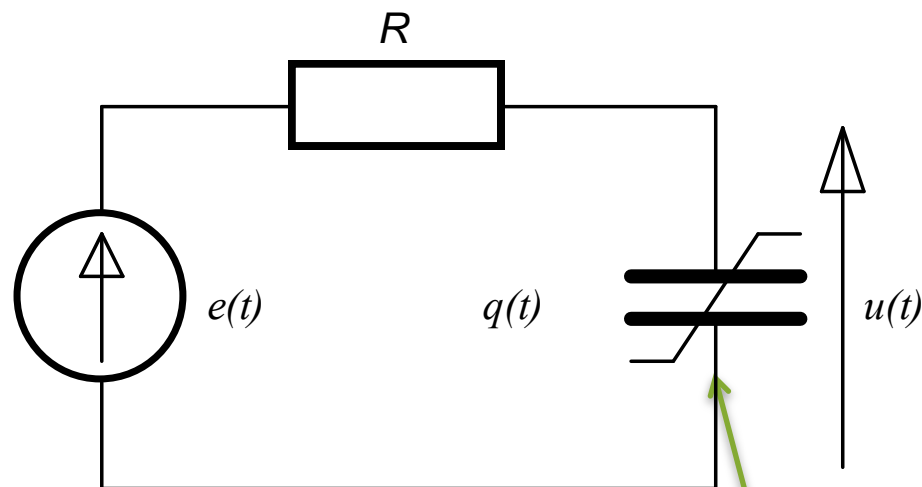
– szukane funkcje

$M=1$

Warunki początkowe:

Warunkiem początkowym dla takiego układu jest jakaś wartość ładunku początkowego w chwili $q(t_0)$, którą akurat w analizie obwodów wygodnie wyznacza się z napięcia początkowego $u(t_0)$, które jest wartością mierzalną:

$$q(t_0) = Q(u(t_0))$$



Funkcja $f(x, y_1, \dots, y_M)$

Nieliniowa
pojemność

Przykład c.d.:

Gdyby pojemność była liniowa, to moglibyśmy napisać:

$$Q(u(t)) = Cu(t)$$

W zależności od przyjętych warunków początkowych będziemy uzyskiwać różne rozwiązania.

Np. Jeżeli przyjmiemy, że:

$q(t_0) = Q(u(t_0)) = 0$ (kondensator całkowicie rozładowany)
oraz

$e(t) = E_0 \cdot 1(t - t_0)$ (załączenie napięcia w chwili t_0)

To umiemy układ rozwiązać analitycznie:

$$u(t) = E_0(1 - e^{-t/\tau})\mathbf{1}(t - t_0)$$

W większości innych przypadków nie umiemy!

Metody rozwiązywania zagadnienia początkowego, o których będziemy mówili opierają się na pewnych ograniczeniach (założeniach) położonych na funkcje $f_m(x, y_1, \dots, y_M)$ $m=1, 2, \dots, M$ tworzące wektor funkcji \mathbf{f} :

- I. Funkcje $f_m(x, y_1, \dots, y_M)$ $m = 1, 2, \dots, M$ są ciągłe w pewnym przedziale $x \in [x_0, b]$ oraz dla każdego zbioru (wektora) funkcji $\mathbf{y} \in \mathbb{R}^m$

$$\mathbf{y} = [y_1, y_2, \dots, y_M]^T$$

- I. Funkcje $f_m(x, y_1, \dots, y_M)$ $m = 1, 2, \dots, M$ spełniają w zdefiniowanym powyżej przedziale x i dla podanych \mathbf{y} warunek Lipschitza względem zmiennej \mathbf{y} , tzn. że istnieje taka liczba L dla $x \in [x_0, b]$ i dowolnych $\mathbf{y} \in \mathbb{R}^m$ że:

$$\|\mathbf{f}(x, y_1) - \mathbf{f}(x, y_2)\| \leq L \|y_1 - y_2\|$$

Spełnienie dwóch podanych warunków powoduje, że:

$$\mathbf{y}(x) = \mathbf{y}(x_0) + \int_{x_0}^x \mathbf{f}(\tau, \mathbf{y}(\tau)) d\tau$$

W dalszej analizie będziemy mówić o *metodach dyskretnych* rozwiązywania układu równań różniczkowych. Polegają one na wyznaczeniu wartości funkcji $\mathbf{y}(x)$ w pewnych punktach x_i $i = 0, 1, 2, \dots, N$.

Dlatego też powyższy warunek należy rozumieć w takim sensie, że obliczenie wartości w dowolnym punkcie x_i można przeprowadzić na podstawie znajomości warunku początkowego i wartości funkcji $f_m(x, y_1, \dots, y_M)$ w kilku punktach $x_0 < x < x_i$ (np. zastępując całkę w powyższym równaniu kwadraturą)

Omawiać będziemy również wyłącznie tzw. *metody różnicowe* i *metody (typu) Rungego-Kutty* rozwiązywania zagadnienia początkowego ze względu na ich uniwersalność.

Metody te charakteryzują się tym, że rozwiązanie problemu w każdym kolejnym $(n+1)$ punkcie x_{n+1} wynoszące \mathbf{y}_{n+1} otrzymuje się za pomocą kilku wcześniej obliczonych wartości w punktach x_{n-j} gdzie $j=0,1,2, \dots$ wynoszących \mathbf{y}_{n-j} oraz kilku wartości wektora funkcji \mathbf{f} .

$$\mathbf{f} = [f_1(x, y_1, \dots, y_M), f_2(x, y_1, \dots, y_M), \dots, f_M(x, y_1, \dots, y_M)]$$

Będziemy zakładać, że punkty, w których wyznaczamy wartości funkcji y są równoodległe od siebie z krokiem h .

Krok ten będziemy nazywać *krokiem całkowania*.

$$\mathbf{y}(x) = \mathbf{y}(x_0) + \int_{x_0}^x \mathbf{f}(\tau, \mathbf{y}(\tau)) d\tau$$

Fakt, że wartość y dla x_{n+1} będziemy wyznaczać na podstawie znajomości poprzednich x_{n-j} będziemy zapisywać w postaci pewnego wzoru (specyficznego dla danej metody) w ogólnej postaci:

$$y_{n+1} = \mathcal{D}(h, y_{n-j}, \dots, y_n)$$

Gdzie $\mathcal{D}(h, y_{n-j}, \dots, y_n)$ jest pewnym równaniem specyficznym dla danej metody będącym równaniem w ogólności nieliniowym.

Będziemy rozróżniali dwa rodzaje błędów metod:

- Błąd lokalny – powstający przez wstawienie do równania metody:

$$y_{n+1} = \mathcal{D}(h, y_{n-j}, \dots, y_n)$$

dokładnych wartości y_{n-j}, \dots, y_n oraz y_{n+1} , **które dla odróżnienia będą oznaczane wielkimi literami** Y_{n-j}, \dots, Y_n oraz Y_{n+1} . Wtedy błąd lokalny to:

$$\mathcal{D}(h, Y_{n-j}, \dots, Y_n) = y_{n+1} + T_{n+1}$$

Błąd
lokalny w
punkcie
n+1

$$T_{n+1} = \mathcal{D}(h, Y_{n-j}, \dots, Y_n) - y_{n+1}$$

Zatem błąd lokalny to błąd jakim obarczone jest rozwiązanie w n+1 punkcie, gdyby poprzednie wartości wykorzystywane do obliczenia funkcji $y_m(x)$ w n+1 punkcie znane byłyby dokładnie.

Będziemy rozróżniali dwa rodzaje błędów metod:

- Błąd globalny – jest to błąd całkowity w punkcie x_{n+1} uwzględniający błąd lokalny oraz fakt, że wartości funkcji $y_m(x)$ w poprzednich punktach, na podstawie których wyznaczyliśmy wartości $y_m(x_{n+1})$ również obciążone są błędami. Błąd ten będziemy oznaczać jako: ε_n (błąd globalny w n-tym punkcie)

Metodę rozwiązywania układów równań różniczkowych będziemy nazywać **stabilną** jeżeli będzie miała pożądane właściwości związane z zachowaniem funkcji $y_m(x)$ $m = 1, 2, \dots, M$:

- Jeżeli wraz ze wzrostem wartości x dokładna wartość funkcji $y_m(x)$ $m = 1, 2, \dots, M$ oznaczana tutaj $Y_m(x)$ $m = 1, 2, \dots, M$ **maleje** to błąd globalny ε_n nie może rosnać wraz ze wzrostem n
- Jeżeli wraz ze wzrostem wartości x dokładna wartość funkcji $y_m(x)$ $m = 1, 2, \dots, M$ oznaczana tutaj $Y_m(x)$ $m = 1, 2, \dots, M$ **rośnie** to błąd globalny ε_n nie może rosnać szybciej wraz ze wzrostem n niż wartość dokładna


Stabilność metody zależy od jej parametrów np. przyjętego kroku całkowania h . Stabilność jest parametrem opisującym wpływ kumulacji błędów obliczeń wartości funkcji $f_m(x, y_1, \dots, y_M)$ oraz $y_m(x)$ $m = 1, 2, \dots, M$ na rozwiązanie w n -tym kroku metody.

Zbieżność metod:

Będziemy wymagali od metod rozwiązywania zadania początkowego zbieżności tzn. takiej cechy, że wraz ze zmniejszaniem się kroku h wartości funkcji $y_m(x)$ obliczane w kolejnych punktach x_n były coraz bliższe wartości prawdziwej.

Zbieżność metod będziemy wyrażać w postaci pewnego parametru zwanego *rzędem metody*. Przez rząd metody rozwiązania zagadnienia początkowego rozumieć będziemy liczbę p wynikającą rozwinięcia w szereg Taylora wyrażenia na błąd lokalny (będący funkcją kroku całkowania) wokół wartości $h=0$:

$$T_{m,n+1} = C_{m,p} h^{p+1} + O(h^{p+2})$$

 Błąd lokalny dla m -tego równania i $n+1$ -ego punktu w którym obliczamy wartość y_m

$$T_{m,n+1} = C_{m,p} h^{p+1} + O(h^{p+2})$$

$$C_{m,p} = \frac{T_{m,n+1}^{(p)}(0)}{(p+1)!}$$

p-ta pochodna
błędu lokalnego
dla m-tej funkcji
w punkcie n

Wzór ten i oparta o niego definicja rzędu metody zakłada spełnienie warunków takich samych jakie nakładane są na krok metody w przypadku całkowania i różniczkowania: ***zakładamy, że krok jest dostatecznie mały.***

Sens rzędu metody rozwiązania jest podobny jak w przypadku zbieżności metod całkowania. Im większy jest rząd metody (i spełnione jest założenie o dostatecznie małym kroku) tym błąd lokalny metody jest mniejszy, a przybliżenie uzyskane tą metodą dokładniejsze.

W elektronice zagadnienie początkowe wykorzystywane jest najczęściej do rozwiązywania problemów odpowiedzi czasowych układów nieliniowych przy zadanych warunkach początkowych i w związku z tym zmienna x ma najczęściej wymiar czasu t .

Ponieważ taka konwencja została przyjęta w podręczniku również na tym wykładzie dalej będziemy się posługiwali zmienną t zamiast x .

Należy jednak pamiętać, że problem zagadnienia początkowego może mieć też zastosowanie do innych problemów technicznych, gdzie zmienna x będzie miała inne znaczenia np. znaczenie odległości.

METODY RÓŻNICOWE:

1. Jednokrokowe:

Metody jednokrokowe opisywane są wzorem ogólnym:

$$\mathbf{y}_{n+1} = \mathbf{y}_n + h_n \underbrace{\Phi(t, \mathbf{y}, h)}_{\text{Wektor m funkcji } \Phi_m \text{ specyficznych dla danej metody i ją definiujących}}$$

Wektor m funkcji Φ_m
specyficznych dla danej metody
i ją definiujących

Cechą charakterystyczną metod jednokrokowych jest to, że do wyznaczenia przybliżenia wektora \mathbf{y} w $n+1$ punkcie wykorzystują one wartości obliczone w tylko jednym poprzednim punkcie n . Przyjmując wymiar czasu dla x możemy powiedzieć, że obliczenie wektora funkcji \mathbf{y} w $n+1$ chwili wymaga znajomości tylko wartości w chwili poprzedniej.

Metody jednokrokowe występują w dwóch odmianach:

- Odmiana otwarta:

Metoda otwarta to metoda w której:

$$\Phi(t, y, h) = f(t_n, y_n, h_n)$$

tzn. że do obliczenia y_{n+1} wykorzystuje się tylko wartości w n-tej chwili czasowej i wartość y_{n+1} można uzyskać przez proste podstawienie do wzoru metody odpowiednich wartości wyliczonych wcześniej

Przykładem prostej metody otwartej jest tzw. *metoda otwarta Eulera*, która powstaje przez przybliżenie w równaniu różniczkowym:

$$\begin{cases} \frac{dy_1}{dx} = f_1(x, y_1, \dots, y_M) \\ \vdots \\ \frac{dy_M}{dx} = f_M(x, y_1, \dots, y_M) \end{cases}$$

pochodnych przez ich numeryczne przybliżenie różnicą progresywną:

$$\frac{dy_m}{dt}(t_n) = y'_m(t_n) = \frac{y_m(t_n + h_n) - y_m(t_n)}{h_n}$$

gdzie: $t_n + h_n = t_{n+1}$

Uzyskujemy wtedy metodę jednokrokową o wzorze (w zapisie wektorowym):

$$\mathbf{y}_{n+1} = \mathbf{y}_n + h_n \underbrace{\mathbf{y}'(t_n)}_{\Phi}$$

Zatem wykorzystanie metody otwartej Eulera polega na:

Jest to funkcja Φ występująca we wzorze ogólnym dla metody otwartej Eulera:

$$\Phi(t, y, h) = f(t_n, y_n, h_n) = \mathbf{y}'(t_n)$$

gdzie:

$\mathbf{y}'(t_n)$ jest pochodną w chwili t_n

1. Obliczeniu wartości pochodnej/pochodnych funkcji w \mathbf{y}'_m chwili czasowej t_n poprzez przyjęcie $f(t_n, y_n, h_n) = \mathbf{y}'(t_n)$.
2. Podstawieniu do powyższego wzoru i obliczeniu w ten sposób wartości funkcji y_m w $n+1$ chwili czasowej

Przykład:

Znaleźć rozwiązanie równania różniczkowego:

$$\frac{dy}{dt} = -2ty(t)$$

W chwili czasowej $t=1$ mając dany warunek początkowy $y(0)=1$ oraz przyjmując stały krok $h=0.2$ przy pomocy otwartej metody Eulera:

1. Mamy tylko jedno równanie, więc $m=1$
2. Stosując pierwszy krok metody otwartej Eulera piszemy:

$$y'(t_n) = -2t_n y(t_n)$$

Przy czym: $y(t_n) = y_n$

3. Wstawiamy tę wartość do wzoru metody otwartej Eulera:

$$y_{n+1} = y_n + h_n y'(t_n)$$

zatem:

$$y_{n+1} = y_n - h_n 2t_n y_n$$

Czyli:

- Mamy warunek początkowy $t=0$ $y_0=1$, wykonujemy krok $h=0.2$:

$$t_1 = t_0 + h = 0.2$$

$$y_1 = y_0 - 2ht_0 y_0 = 1 - 2 \cdot 0.2 \cdot 0 \cdot 1 = 1$$

- sadas

- Kolejny krok:

$$t_2 = t_1 + h = 0.4$$

$$y_2 = y_1 - 2ht_1y_1 = 1 - 2 \cdot 0.2 \cdot 0.2 \cdot 1 = 0.2$$

itd. aż do obliczenia y_5 , która to wartość odpowiada
 $t_5 = t_0 + 5h = 1$

- Odmiana zamknięta

Metoda zamknięta to metoda w której:

$$\Phi(t, y, h) = f(t_n, y_n, h_n, t_{n+1}, y_{n+1}, h_{n+1})$$

tzn. że do obliczenia y_{n+1} wykorzystuje się tylko wartości w n-tej i n+1 chwili czasowej. W takiej metodzie wartość y_{n+1} uzyskuje się poprzez podstawienie funkcji $\Phi(t, y, h)$ do wzoru ogólnego metod jednokrokowych:

$$y_{n+1} = y_n + h_n \Phi(t, y, h)$$

Uzyskując w ten sposób w ogólności układ równań nieliniowych, który następnie trzeba rozwiązać dowolną poprawną dla danego układu metodą np. metodą wielowymiarową Newtona.

Przykładem prostej metody zamkniętej jest tzw. *metoda zamknięta Eulera*, która powstaje przez przybliżenie w równaniu różniczkowym:

$$\begin{cases} \frac{dy_1}{dx} = f_1(x, y_1, \dots, y_M) \\ \vdots \\ \frac{dy_M}{dx} = f_M(x, y_1, \dots, y_M) \end{cases}$$

pochodnych przez ich numeryczne przybliżenie różnicą wsteczną:

$$\frac{dy_m}{dt}(t_{n+1}) = y'_m(t_{n+1}) = \frac{y_m(t_n + h_n) - y_m(t_n)}{h_n}$$

gdzie: $t_n + h_n = t_{n+1}$

Uzyskujemy wtedy metodę jednokrokową o wzorze (w zapisie wektorowym):

$$\mathbf{y}_{n+1} = \mathbf{y}_n + h_n \underbrace{\mathbf{y}'(t_{n+1})}_{\Phi}$$

Jest to funkcja Φ występująca we wzorze ogólnym dla metody zamkniętej Eulera:

$$\Phi(t, y, h) = \mathbf{f}(t_{n+1}, y_{n+1}, h_{n+1}) = \mathbf{y}'(t_{n+1})$$

gdzie:

$\mathbf{y}'(t_{n+1})$ jest pochodną w chwili t_{n+1}

Zatem wykorzystanie metody zamkniętej Eulera polega na:

1. Obliczeniu wartości pochodnej/pochodnych funkcji w y'_m chwili czasowej t_{n+1} poprzez przyjęcie $\mathbf{f}(t_{n+1}, y_{n+1}, h_{n+1}) = \mathbf{y}'(t_{n+1})$.
2. Podstawieniu do powyższego wzoru i stworzeniu w ten sposób układ równań opisujący wartości funkcji y_m w $n+1$ chwili czasowej

3. Rozwiązania tego układu równań ze względu na zmienną $y(t_{n+1})$ (traktując ten wektor jako wektor niewiadomych)

Przykład:

Znaleźć rozwiązanie równania różniczkowego:

$$\frac{dy}{dt} = -2ty(t)$$

W chwili czasowej $t=1$ mając dany warunek początkowy $y(0)=1$ oraz przyjmując stały krok $h=0.2$ przy pomocy zamkniętej metody Eulera.

1. Przyjmujemy:

$$y'(t_{n+1}) = -2t_{n+1}y(t_{n+1})$$

2. Wstawiając do wzoru zamkniętej metody Eulera:

$$y_{n+1} = y_n + h_n y'(t_{n+1})$$

zatem:

$$y_{n+1} = y_n - 2h_n t_{n+1} y_{n+1}$$

Pierwszy krok:

$$t_1 = t_0 + h = 0.2$$

$$y_1 = y_0 - 2h t_1 y_1 = 1 - 2 \cdot 0.2 \cdot y_1$$

Rozwiązując to równanie (w tym przypadku proste równanie liniowe) otrzymamy:

$$y_1 \approx 0.7143$$

itd.

RZĄD I ZBIEŻNOŚĆ METOD JEDNOKROKOWYCH:

Można wykazać, że metoda jednokrokowa jest zbieżna jeżeli spełniony jest warunek:

$$\phi(t, \mathbf{y}, 0) = \mathbf{f}(t, \mathbf{y})$$

tzn. jeżeli wektor funkcji Φ charakterystycznych danej metody dla kroku $h=0$ będzie równy wektorowi funkcji \mathbf{f} będącej prawą stroną równania różniczkowego.

Jeżeli metoda jednokrokowa jest zbieżna to jej rząd wyznacza się przez analizę lokalnego błędu odcięcia:

$$T(t, \mathbf{y}, h) \equiv \frac{\mathbf{y}(t+h) - \mathbf{y}(t)}{h} - \phi(t, \mathbf{y}, h)$$

Przeprowadzając taką analizę dla metod Eulera okazuje się, że:

$$T(t, \mathbf{y}, h) \equiv O(h)$$

Są one więc metodami rzędu pierwszego.

Na podstawie tej analizy można zbudować takie metody jednokrokowe (dobierając odpowiednio funkcję Φ), aby zwiększyć rząd metody. Tego typu rozwiązania mają postać ogólną:

$$\phi(t, \mathbf{y}, h) = \alpha_1 \mathbf{f}(t, \mathbf{y}) + \alpha_2 \mathbf{f}(t + \gamma_1 h, \mathbf{y} + \gamma_2 h \mathbf{f}(t, \mathbf{y}))$$

Są to tzw. metody jednokrokowe wyższych rzędów (metody typu Rungego-Kutty – o nich dalej).

2. Wielokrokowe:

Metody wielokrokowe opisywane są wzorem ogólnym:

$$\mathbf{y}_{n+1} = \sum_{k=1}^K \alpha_k \mathbf{y}_{n+1-k} + h \sum_{k=0}^K \beta_k \mathbf{f}_{n+1-k} \quad \text{dla } n = K-1, \dots, N-1,$$

Cechą charakterystyczną metod wielokrokowych jest to, że do wyznaczenia przybliżenia wektora \mathbf{y} w $n+1$ punkcie wykorzystują one wartości obliczone w K poprzednich punktach. Przyjmując wymiar czasu dla x możemy powiedzieć, że obliczenie wektora funkcji \mathbf{y} w $n+1$ chwili wymaga znajomości wartości wektora \mathbf{y} w K poprzednich chwilach czasowych.

Rząd metod wielokrokowych jest w ogólnym przypadku wyższy niż prostych metod jednokrokowych (ale niekoniecznie metod Rungego-Kutty)

Metody wielokrokowe również można podzielić na metody otwarte i zamknięte:

$$\mathbf{y}_{n+1} = \sum_{k=1}^K \alpha_k \mathbf{y}_{n+1-k} + h \sum_{k=0}^K \beta_k \mathbf{f}_{n+1-k} \quad \text{dla } n = K-1, \dots, N-1,$$

Metodę nazwiemy otwartą kiedy współczynnik $\beta_0 = 0$, co powoduje, że prawa strona równania nie zawiera czynników o indeksie $n+1$ i w związku z tym wyznaczenie \mathbf{y}_{n+1} wymaga jedynie podstawienia danych z k ostatnich kroków $k=1, 2, \dots, K$ do wzoru danej metody.

Metodę nazwiemy zamkniętą, kiedy $\beta_0 \neq 0$, co powoduje, że równanie jest w ogólności równaniem nieliniowym i wymaga rozwiązania odpowiednimi metodami.

Większość metod wielokrokowych powstaje z rozwiązania układu równań różniczkowych:

$$\begin{cases} \frac{dy_1}{dx} = f_1(x, y_1, \dots, y_M) \\ \vdots \\ \frac{dy_M}{dx} = f_M(x, y_1, \dots, y_M) \end{cases}$$

za pomocą całki:

$$y_m(t_{n+1}) = y_m(t_{n-M}) + \int_{t_{n-M}}^{t_{n+1}} f_m(t, \mathbf{y}(t)) dt$$

gdzie całka obliczana jest numerycznie za pomocą jednej z poznanych formuł interpolacyjnych opartych o $L+2$ węzły interpolacji:

$$t_k: k = n - L, n - L + 1, \dots, n, n + 1 \quad \text{gdzie } t_k \in [t_n - M, t_{n+1}]$$

Jeżeli zastosujemy kwadraturę Newtona-Cotesa (opartą o interpolację wielomianami Lagrange'a) to otrzymamy metody Adamsa-Bashforth (otwarte) i Adamsa-Moultona (zamknięte) opisane równaniami:

| METODY ADAMSA-BASHFORTH | p | A_{p+1} |
|--|-----|-----------|
| $y_{n+1} = y_n + hf_n$ | 1 | 1/2 |
| $y_{n+1} = y_n + h(3f_n - f_{n-1})/2$ | 2 | 5/12 |
| $y_{n+1} = y_n + h(23f_n - 16f_{n-1} + 5f_{n-2})/12$ | 3 | 3/8 |
| $y_{n+1} = y_n + h(55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3})/24$ | 4 | 251/720 |
| METODY ADAMSA-MOULTONA | p | A_{p+1} |
| $y_{n+1} = y_n + hf_{n+1}$ | 1 | -1/2 |
| $y_{n+1} = y_n + h(f_{n+1} + f_n)/2$ | 2 | -1/12 |
| $y_{n+1} = y_n + h(5f_{n+1} + 8f_n - f_{n-1})/12$ | 3 | -1/24 |
| $y_{n+1} = y_n + h(9f_{n+1} + 19f_n - 5f_{n-1} + f_{n-2})/24$ | 4 | -19/720 |

METODY RUNGEGO-KUTTY

Okazuje się, że można skonstruować metodę jednokrokową o wysokim rzędzie (takim jak w metodach wielokrokowych), ale okupione jest to zwiększeniem nakładów obliczeniowych. Metody tego typu nazywane są metodami Rungego-Kutty, wyrażone są one wzorem ogólnym:

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \sum_{i=1}^R w_i \mathbf{K}_i$$

gdzie:

$$\mathbf{K}_1 = hf(x_n, \mathbf{y}_n), \quad \mathbf{K}_i = hf\left(x_n + a_i h, \mathbf{y}_n + \sum_{j=1}^{i-1} b_{ij} \mathbf{K}_j\right)$$

w_i, a_i, b_{ij} – są stałymi zdefiniowanymi dla każdej z metod tego typu

Zatem zwiększenie rzędu metody jednokrokowej okupione jest wzrostem nakładów obliczeniowych związanym z koniecznością wielokrotnego obliczania współczynnika \mathbf{K}_i

Podobnie jak poprzednio występują metody otwarte i zamknięte:

- Metoda Rungego-Kutty jest otwarta, jeżeli $b_{ij} = 0$ dla $j \geq i$ oraz $r = 1, \dots, R$
- W przeciwnym przypadku metoda jest metodą zamkniętą, a wyznaczenie funkcji $\Phi(t, y, h) = \sum_{i=1}^R w_i \mathbf{K}_i$ wymaga rozwiązania układu równań nieliniowych ze względu na \mathbf{K}_i

Maksymalny rząd otwartej metody Rungego-Kutty, korzystającej z R wartości funkcji wynosi $p(R) = R$, przy czym dla $R=1,2,3,4$ rząd ten jest maksymalny i wynosi $p(R)=R$. Metody wyższych stopni niekoniecznie mają rząd maksymalny

Maksymalny rząd zamkniętej metody Rungego-Kutty może być równy $2R$, jednak ze względu na konieczność rozwiązywania nieliniowych równań algebraicznych koszt wykonania jednego kroku jest dla takiej metody znacznie większy niż dla odpowiedniej metody otwartej. Najczęściej stosowana jest metoda Rungego-Kutty czwartego rzędu ze współczynnikami:

$$\Phi(t_n, \mathbf{y}_n, h) = \frac{1}{6} (k_1 + 2k_2 + 2k_3 + k_4)$$

$$k_1 = \mathbf{f}(t_n, \mathbf{y}_n)$$

$$k_2 = \mathbf{f}(t_n + h/2, \mathbf{y}_n + hk_1/2)$$

$$k_3 = \mathbf{f}(t_n + h/2, \mathbf{y}_n + hk_2/2)$$

$$k_4 = \mathbf{f}(t_n + h, \mathbf{y}_n + hk_3)$$

Odmianami metod Rungego-Kutty często spotykanymi w literaturze są metody:

Metoda Heuna:

$$\alpha_1 = \alpha_2 = \frac{1}{2}, \gamma_1 = \gamma_2 = 1$$

$$\phi(t, \mathbf{y}, h) = \frac{1}{2} \left(\mathbf{f}(t, \mathbf{y}) + \mathbf{f}(t + h, \mathbf{y} + h\mathbf{f}(t, \mathbf{y})) \right)$$

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \frac{h}{2} \left(\mathbf{f}(t, \mathbf{y}_n) + \mathbf{f}(t + h, \mathbf{y}_n + h\mathbf{f}(t, \mathbf{y}_n)) \right)$$

Metoda zmodyfikowana Eulera:

$$\alpha_1 = 0, \alpha_2 = 0, \gamma_1 = \gamma_2 = \frac{1}{2}$$

$$\phi(t, \mathbf{y}, h) = \mathbf{f}(t + h, \mathbf{y}_n + h\mathbf{f}(t, \mathbf{y}))$$

$$\mathbf{y}_{n+1} = \mathbf{y}_n + h\mathbf{f}\left(t + \frac{h}{2}, \mathbf{y}_n + \frac{h}{2}\mathbf{f}(t, \mathbf{y}_n)\right)$$

Są to metody rzędu 2.

Podobnie jak w przypadku metod różniczkowania obliczenie każdego kolejnego stosując daną metodę Rungego-Kutty punkty obarczone jest następującymi błędami:

- a) Błędem globalnym wynikającym z zastosowanej metody i niedokładności danych wejściowych (warunku początkowego). Dla metody będącej rzędu p podlega on oszacowaniu jako:

$$\varepsilon_n \approx Kh^p$$

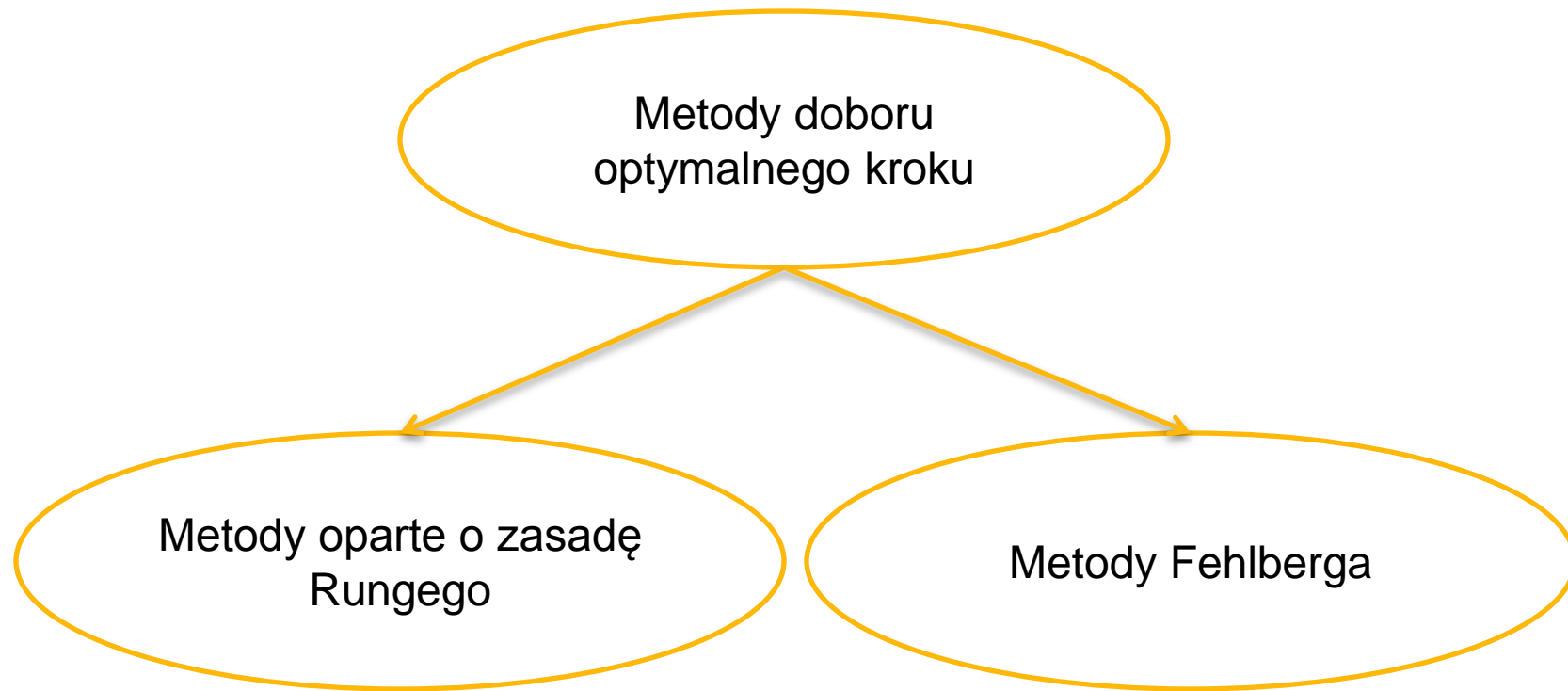
gdzie: K to pewna stała, h to krok metody.

- b) Błędem zaokrągleń numerycznych. Błąd ten podlega oszacowaniu (dla małych wartości kroku):

$$|\eta_{y,N}| \xrightarrow{h \rightarrow 0} N \sum_{n=0}^{N-1} \eta_{+,n} \leq N \cdot eps = \frac{T}{h} \cdot eps$$

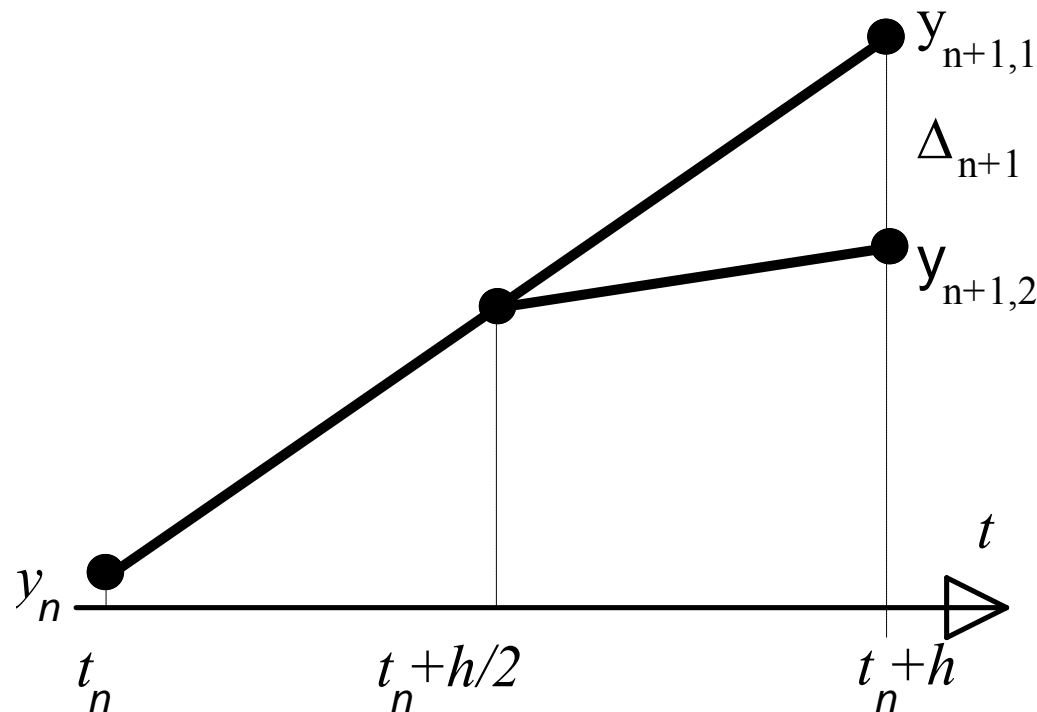
Dlatego też teoretycznie istnieje optymalna wartość kroku dająca najdokładniejsze rozwiązanie.

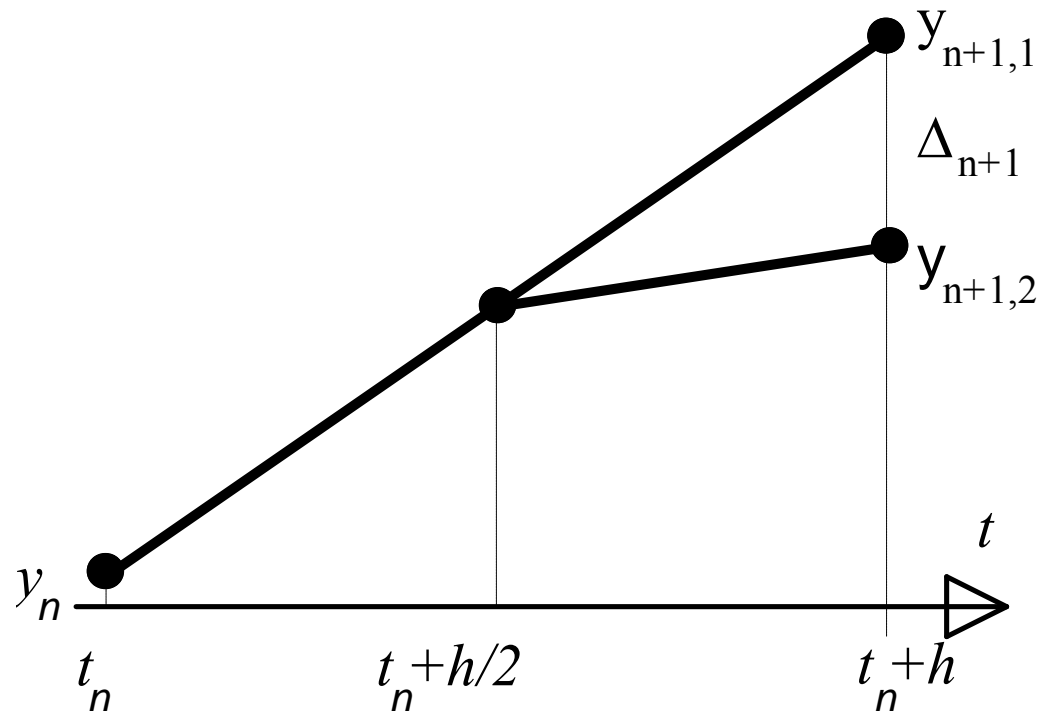
Praktycznie jednak wartość kroku optymalnego zależy od rozwiązywanego układu równań. Dlatego problem doboru optymalnego kroku rozwiązuje się przez oszacowanie w każdym kroku lokalnego błędu obcięcia i automatycznego wyboru kroku w taki sposób, aby osiągnąć założoną dokładność rozwiązania przy jak najmniejszych nakładach obliczeniowych.



METODY OPARTE O ZASADĘ RUNGEGO:

Założmy, że metoda całkowania jest rzędu p i że dominuje lokalny błąd dyskretyzacji rzędu h^{p+1} . Jednym z praktycznie użytecznych sposobów oszacowania tego błędu jest tzw. zasada Rungego, zgodnie z którą następny punkt rozwiązania wyznacza się dwukrotnie: najpierw przy użyciu formuły z krokiem h (wynik oznaczony przez $y_{n+1,1}$), a potem - przez dwukrotne użycie formuły z krokiem $h/2$ (wynik oznaczony przez $y_{n+1,2}$):





Wartości funkcji y_{n+1} wyznaczone każdą z metod wynoszą:

$$y_{n+1,1} = y(t+h) + \underbrace{C_{p+1}h^{p+1} + O(h^{p+2})}_{\text{Błąd dla kroku } h}$$

$$y_{n+1,2} = y(t+h) + \underbrace{2C_{p+1}(h/2)^{p+1} + O(h^{p+2})}_{\text{Błąd dla kroku } h/2}$$

Wartość
rzeczywista

Błąd dla kroku h

Błąd dla kroku $h/2$

Różnica Δ_{n+1} może służyć do oszacowania lokalnego błędu obcięcia. Z dokładnością $O(h^{p+2})$ mamy bowiem:

$$\Delta_{n+1} \equiv y_{n+1,1} - y_{n+1,2}$$

$$\Delta_{n+1} = C_{p+1}h^{p+1} - 2C_{p+1}\left(\frac{h}{2}\right)^{p+1} = C_{p+1}h^{p+1}(1 - 2^{-p})$$

stąd:

$$C_{p+1} \approx \frac{\Delta_{n+1}}{h^{p+1}(1 - 2^{-p})}$$

Wynika stąd oszacowanie błędu metody popełnionego przy obliczaniu y_{n+1} :

$$y_{n+1,1} \approx y(t+h) + C_{p+1}h^{p+1} = y(t+h) + \frac{\Delta_{n+1}}{h^{p+1}(1-2^{-p})}h^{p+1}$$

$$\xi_{n+1,1} = \frac{\Delta_{n+1}}{h^{p+1}(1-2^{-p})}h^{p+1} = \frac{\Delta_{n+1}}{(1-2^{-p})}$$

$$y_{n+1,2} \approx y(t+h) + 2C_{p+1}\frac{h^{p+1}}{2} = y(t+h) + 2\frac{\Delta_{n+1}}{h^{p+1}(1-2^{-p})}\frac{h^{p+1}}{2}$$

$$\xi_{n+1,2} = 2\frac{\Delta_{n+1}}{h^{p+1}(1-2^{-p})}\frac{h^{p+1}}{2} = \frac{\Delta_{n+1}}{2^p(1-2^{-p})}$$

Algorytm wyboru kroku stosujący metodę Rungego wygląda zatem następująco:

W każdym kroku:

- I. Przyjąć zadowalającą użytkownika dokładność na poziomie $e = \frac{\varepsilon h}{K}$, gdzie ε jest założonym błędem globalnym, a K przyjętym współczynnikiem (bezpieczeństwa)
- II. Obliczyć dla $n+1$ punktu błędy ξ_{n+1} (oba błędy $\xi_{n+1,1}$ oraz $\xi_{n+1,2}$)
- III. Jeżeli $\xi_{n+1} \leq e$ ($\xi_{n+1,1} \leq e$ oraz $\xi_{n+1,2} \leq e$) to krok dobrany jest prawidłowo i następny punkt obliczamy z takim samym krokiem
- IV. Jeżeli $\xi_{n+1} \ll e$ to krok należy wydłużyć aby zmniejszyć koszt obliczeń. Krok wydłużamy wg wybranej strategii np. podwajamy, bądź wg wzoru:

$$h_{n+2} = h_{n+1} \min \left\{ 10, \frac{1}{\sqrt[p+1]{\frac{\xi_{n+1}}{e}}} \right\}$$

- V. Jeżeli $\xi_{n+1} > e$ to obliczenia w tym kroku należy powtórzyć zmniejszając krok wg wybranej strategii np. o połowę lub:

$$h_{n+1}^{(nowy)} = \frac{h_{n+1}^{(stary)}}{\min \left\{ 2, \sqrt[p+1]{\frac{\xi_{n+1}}{e}} \right\}}$$

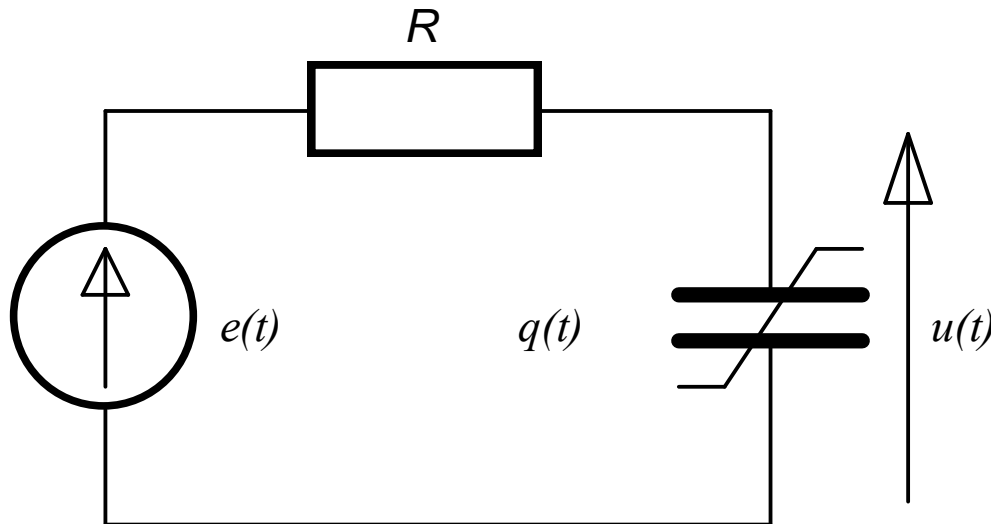
- VI. Jeżeli zachodzi przypadek graniczny tzn. : $\xi_{n+1,1} > e$ oraz $\xi_{n+1,2} < e$ to należy przyjąć wybraną strategię z poprzednich. Zazwyczaj zakłada się kontynuowanie obliczeń z bieżącą wartością kroku (ewentualny wzrost błędu wykryty będzie w kolejnych krokach i skorygowany)

PRZYKŁAD:

Wyznaczyć odpowiedź układu z poniższego rysunku metodą otwartą Eulera dla:

$$R=1; C=1; E_0=1; t=[0,5]$$

Obliczenia przeprowadzić ze stałym i zmiennym krokiem przy założeniu błędu globalnego $\varepsilon < 0.005$



PRZYKŁAD c.d.

Porównanie błędów globalnych rozwiązania w kolejnych krokach:

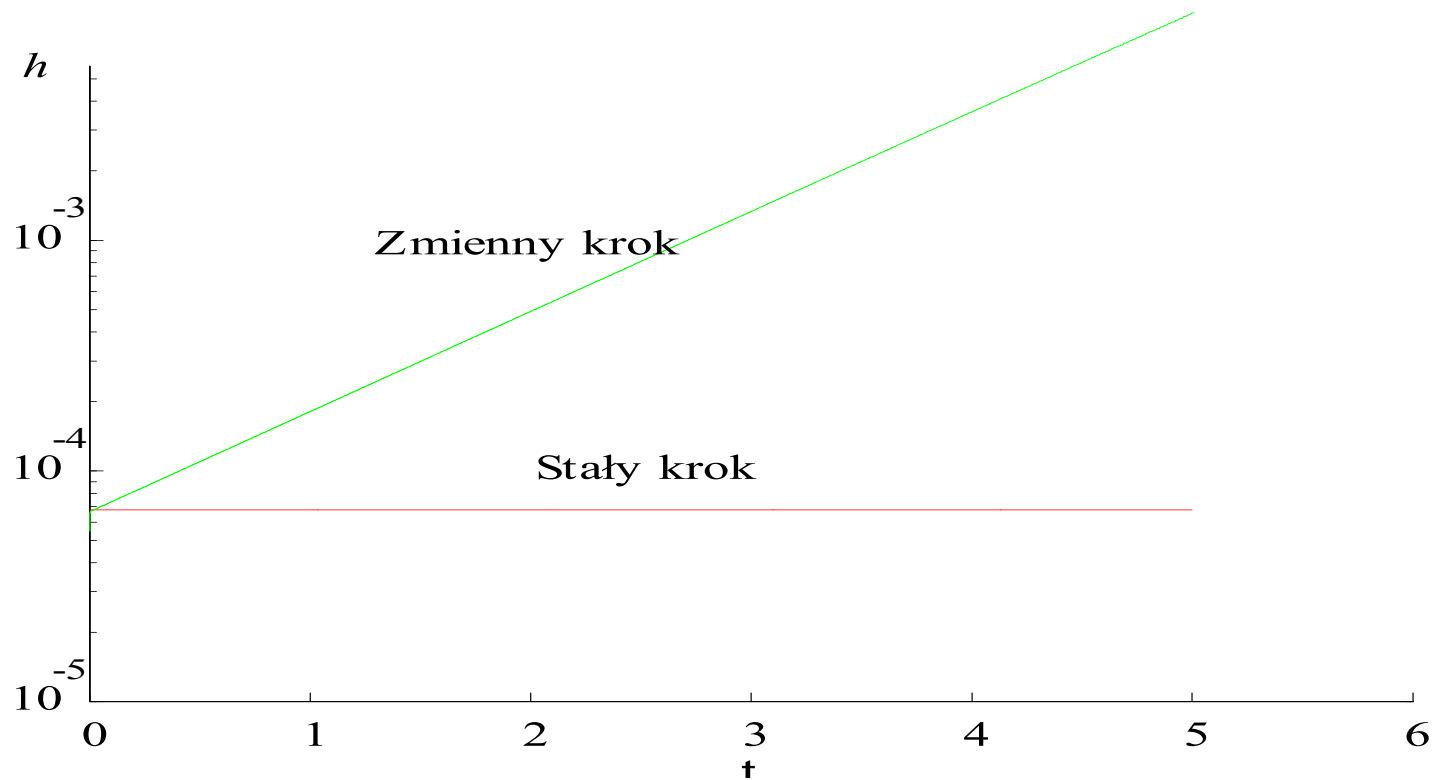


Metoda zmiennokrokowa utrzymuje błąd na zadanym poziomie i do obliczenia odpowiedzi w chwili $t=5$ wykonała $N=373$ kroków.

Metoda stałokrokowa dzięki zastosowaniu odpowiednio małego kroku powoduje spadek błędu w kolejnych punktach. Ale obliczenie wyniku w chwili $t=5$ wymagało zrobienia $N=73707$ kroków

PRZYKŁAD c.d.

Porównanie długości kroku w kolejnych punktach:



METODA FEHLBERGA:

W metodzie Fehlberga do estymacji błędu w $n+1$ kroku wykorzystuje się nie dwukrotne obliczenie wartości w t_{n+1} tą samą metodą, ale z różnym krokiem, ale obliczenie jej w chwili t_{n+1} z tym samym krokiem, ale dwoma metodami o różnym rzędzie.

Dla takiego podejścia można przeprowadzić podobną analizę jak przedstawiona poprzednio i wykorzystać ją do estymacji błędu ξ_{n+1} .

Najczęściej do tego celu przyjmuje się pary metod Rungego-Kutty:

- metody 2 i 3 rzędu (w MATLABIE zaimplementowane jako funkcja ode23)
- Metody 4 i 5 rzędu (w MATLABIE zaimplementowane jako funkcja ode45)

Metody takie noszą miano metod Rungego-Kutty- Fehlberga.

Działanie metody Rungego-Kutty- Fehlberga najprościej zrozumieć na przykładzie pary Dormanda-Prince'a (metody RK 4 i 5 rzędu):

Metoda RK4:

$$y_{n+1}^{(4)} = y_n + \sum_{i=1}^4 c_i v_i$$

Metoda RK5:

$$y_{n+1}^{(5)} = y_n + \sum_{i=1}^5 c_i v_i$$

gdzie:

$$v_0 = f(x_n, y_n)$$

$$v_i = f(x_n + a_i h, y_n + h \sum_{j=0}^{i-1} b_{i,j} v_j)$$

Współczynniki a , b , c i d (współczynniki Fehlberga):

| i | a_i | $b_{i,0}$ | $b_{i,1}$ | $b_{i,2}$ | $b_{i,3}$ | $b_{i,4}$ |
|-----|-------|-----------|------------|------------|-----------|-----------|
| 1 | 1/4 | 1/4 | | | | |
| 2 | 3/8 | 3/32 | 9/32 | | | |
| 3 | 12/13 | 1932/2197 | -7200/2197 | 7296/2197 | | |
| 4 | 1 | 439/216 | -8 | 3680/513 | -845/4104 | |
| 5 | 1/2 | -8/27 | 2 | -3544/2565 | 1859/4104 | -11/40 |

| i | 0 | 1 | 2 | 3 | 4 | 5 |
|-------|--------|---|------------|-------------|-------|------|
| c_i | 25/216 | 0 | 1408/2565 | 2197/4104 | -1/5 | |
| d_i | 16/135 | 0 | 6656/12825 | 28561/56430 | -9/50 | 2/55 |

Podstawą dla oszacowania błędu lokalnego ξ_{n+1} jest:

$$\xi = y_{n+1}^{(5)} - y_{n+1}^{(4)}$$

Metoda ta szacuje błąd lokalny, na który nakładamy warunki takie same jak poprzednio. Algorytm jest analogiczny.

STABILNOŚĆ

Jeżeli dana metoda różnicowa (jednokrokowa lub wielokrokowa) jest zbieżna i ma potencjalnie wysoki rząd to jeszcze nie znaczy, że jest to metoda użyteczna. Może się bowiem okazać, że w skutek kumulacji błędów popełnianych w poprzednich krokach wynik nią otrzymany będzie obciążony bardzo dużym błędem całkowitym.

Na czym polega praktyczna różnica i skąd się bierze problem stabilności?

Otóż wysoki rząd metody mówi nam, że przybliżenie w kolejnym punkcie jest dokładniejsze jeżeli chodzi o błąd lokalny (tzn. dla dokładnych danych wejściowych). Jeżeli dane wejściowe nie są dokładne to rozwiązanie będzie obciążone dodatkowym błędem, który w niesprzyjających okolicznościach będzie się sumował. Wtedy dla pewnego n -tego kroku (potencjalnie nawet dużego) błąd całkowity będzie na tyle duży, że metoda nie będzie użyteczna.

Powstaje pytanie – które metody i kiedy są stabilne, a które nie?

Badanie stabilności metod przeprowadza się wykorzystując specjalny rodzaj równania różniczkowego zwany równaniem liniowym (lub testowym) ponieważ wnioski płynące z analizy rozwiązania tego rodzaju równania daną metodą można uogólnić na szeroką klasę równań różniczkowych.

Liniowe zadanie testowe ma postać:

$$\mathbf{y}'(t) = \mathbf{A}\mathbf{y}(t) + \mathbf{g}(t)$$

Przy czym ograniczymy się tutaj do przypadków, kiedy macierz \mathbf{A} powyższego równania ma różne wartości własne $\lambda_1, \lambda_2, \dots \in \mathbb{C}$. Jeżeli tak jest to istnieje macierz kwadratowa \mathbf{T} taka, że:

$$\mathbf{A} = \mathbf{T}^{-1} \cdot \mathbf{\Lambda} \cdot \mathbf{T}$$

gdzie $\mathbf{\Lambda}$ jest specjalną macierzą diagonalną:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \lambda_M \end{bmatrix}$$

Wtedy nasz układ równań różniczkowych można przedstawić w postaci:

$$\mathbf{y}'(t) = \mathbf{T}^{-1} \cdot \mathbf{\Lambda} \cdot \mathbf{T}\mathbf{y}(t) + \mathbf{g}(t)$$

I mnożąc przez \mathbf{T}

$$\mathbf{T}\mathbf{y}'(t) = \mathbf{\Lambda} \cdot \mathbf{T}\mathbf{y}(t) + \mathbf{T}\mathbf{g}(t)$$

Dla macierzy $\mathbf{\Lambda}$ spełniających podany warunek jest to więc nowy układ, w którym $\mathbf{z}'(t) = \mathbf{T}\mathbf{y}'(t)$ oraz $\mathbf{z}(t) = \mathbf{T}\mathbf{y}(t)$

$$\mathbf{z}'(t) = \mathbf{\Lambda} \cdot \mathbf{z}(t) + \mathbf{T}\mathbf{g}(t)$$

Powyższe równanie dalej będzie zapisywane stosując nazwę zmiennej \mathbf{y}' jako:

$$\mathbf{y}'(t) = \mathbf{\Lambda} \cdot \mathbf{y}(t) + \mathbf{g}(t)$$

METODY WIELOKROKOWE:

Aby zanalizować stabilność metod wielokrokowych należy wyprowadzać wzór na błąd globalny ε_n . Dla układu równań:

$$\varepsilon_n = Y_n - y_n$$

ponieważ metoda wielokrokowa opisana jest równaniem:

$$y_{n+1} = \sum_{k=1}^K \alpha_k y_{n+1-k} + h \sum_{k=0}^K \beta_k f_{n+1-k} \quad \text{dla } n = K-1, \dots, N-1,$$

W ogólności:

$$Y_{n+1} = \sum_{k=1}^K \alpha_k Y_{n+1-k} + h \sum_{i=0}^K \beta_k f_{n+1-k} + T_n$$

Błąd lokalny

$$y_{n+1} = \sum_{k=1}^K \alpha_k y_{n+1-k} + h \sum_{i=0}^K \beta_k f_{n+1-k} - \delta_n$$

Błąd wynikający z niedokładnego rozwiązania

Odejmując stronami:

$$\boldsymbol{\varepsilon}_{n+1} = \mathbf{Y}_n - \mathbf{y}_n$$

$$\boldsymbol{\varepsilon}_{n+1} = \sum_{k=1}^K \alpha_k \boldsymbol{\varepsilon}_{n+1-k} + h \sum_{i=0}^K \beta_k \boldsymbol{\eta}_{n+1-k} + \mathbf{B}_n$$

Gdzie:

$$\boldsymbol{\eta}_{n+1-k} = \mathbf{f}_{n+1-k}^{(Y)} - \mathbf{f}_{n+1-k}^{(y)} \quad \mathbf{B}_n = \mathbf{T}_n + \boldsymbol{\delta}_n$$

Jeżeli funkcje \mathbf{f}_n posiadają ciągłe pochodne (w większości praktycznych przypadków), to z twierdzenia o wartości średniej:

$$\boldsymbol{\eta}_{n+1-k} = \boldsymbol{\varepsilon}_{n+1-k} \cdot \mathbf{f}_y(x_{n+1-k}, \mathbf{y}_{n+1-k}, \boldsymbol{\varepsilon}_{n+1-k}, \boldsymbol{\Theta}_{n+1-k})$$

$$\mathbf{f}_y(x, \mathbf{y}, \boldsymbol{\varepsilon}, \boldsymbol{\Theta}) = \left[\frac{\delta f_i(x, \mathbf{y} + \Theta_i \boldsymbol{\varepsilon})}{\delta y_j} \right]_{i,j=1,2,\dots,M}$$

Równanie dla błędu całkowitego wygląda więc tak:

$$\begin{aligned} & [I - hb_0 f_y(x_{n+1}, y_{n+1}, \varepsilon_{n+1}, \theta_{n+1})] \varepsilon_{n+1} \\ &= \sum_{k=1}^K [\alpha_k I + h\beta_k f_y(x_{n+1-k}, y_{n+1-k}, \varepsilon_{n+1-k}, \theta_{n+1-k})] \varepsilon_{n+1-k} + B_n \end{aligned}$$

Wstawiając teraz nasze równanie testowe:

$$y'(t) = \Lambda \cdot y(t) + g(t)$$

Wiedząc, że będziemy badać stabilność (a więc zależność błędu od niedokładności znajomości rozwiązania $y(t)$) zakładamy, że czynnik $g(t)$ jest dokładny – błąd globalny nie zależy od niego. Dla zbadania stabilności wystarczy więc zbadać zachowanie błędu globalnego dla równania testowego:

$$y'(t) = \underbrace{\Lambda \cdot y(t)}_{f(x, y)}$$

$$\mathbf{f}_y(x, \mathbf{y}, \boldsymbol{\varepsilon}, \boldsymbol{\theta}) = \left[\frac{\delta f_i(x, \mathbf{y} + \Theta_i \boldsymbol{\varepsilon})}{\delta y_j} \right]_{i,j=1,2,\dots,M} = \boldsymbol{\Lambda}$$

A ostatecznie równanie dla błędu całkowitego:

$$[\mathbf{I} - hb_0 \boldsymbol{\Lambda}] \boldsymbol{\varepsilon}_{n+1} = \sum_{k=1}^K [\alpha_k \mathbf{I} + h\beta_k \boldsymbol{\Lambda}] \boldsymbol{\varepsilon}_{n+1-k} + \mathbf{B}_n$$

Zazwyczaj czynie się jeszcze jedno założenie mówiące, że składowa błędu \mathbf{B}_n jest stała i wtedy błąd globalny można przedstawić jako:

$$\boldsymbol{\varepsilon}_n = \boldsymbol{\varepsilon}_n^1 + \boldsymbol{\varepsilon}_n^2$$

Składowa
błędu dla $\mathbf{B}_n=0$

Czynnik
dodatkowy

$$\varepsilon_n = \varepsilon_n^1 + \varepsilon_n^2$$

Decydująca o stabilności jest składowa ε_n^1 , tak więc dla $\mathbf{B}_n=0$:

$$[\mathbf{I} - hb_0\mathbf{A}]\varepsilon_{n+1} = \sum_{k=1}^K [\alpha_k \mathbf{I} + h\beta_k \mathbf{A}]\varepsilon_{n+1-k}$$

1. Przypadek $M=1$ (pojedyncze równanie różniczkowe)

Dla pojedynczego równania różniczkowego: $\mathbf{A} = \lambda$

$$[1 - hb_0\lambda]\varepsilon_{n+1} = \sum_{k=1}^K [\alpha_k + h\beta_k\lambda]\varepsilon_{n+1-k}$$

Z równaniem tym kojarzymy wielomian $w(z)$ (szukamy rozwiązania w formie $\varepsilon_{n+1} = z^n$) jest tw. wielomian charakterystyczny:

$$[1 - hb_0\lambda]z^n = \sum_{k=1}^K [\alpha_k + h\beta_k\lambda]z^{n-k}$$

Z równaniem tym kojarzymy wielomian $w(z)$ (szukamy rozwiązania w formie $\varepsilon_{n+1} = z^n$) jest tzw. wielomian charakterystyczny:

$$w(z) = z^n - \frac{1}{[1 - hb_0\lambda]} \sum_{k=1}^K [\alpha_k + h\beta_k\lambda] z^{n-k}$$

Położenie miejsc zerowych tego wielomianu ma zasadnicze znaczenia dla stabilności. **Proszę zwrócić uwagę, że pierwiastki tego wielomianu zależą od λ (a więc danych, typu równania) oraz od wartości kroku h .**

Warunkiem koniecznym i wystarczającym zbieżności metody jest aby dla $h=0$ wszystkie pierwiastki spełniały nierówności:

a) Pierwiastki jednokrotne: $|z_j| \leq 1$

b) Pierwiastki wielokrotne: $|z_j| < 1$

Ponieważ położenie pierwiastków jest zależne od kroku i wartości λ to rozróżniamy kilka rodzajów stabilności:

- Każda metoda spełniająca powyższe warunki dla $h=0$ nazywana jest **metodą D-stabilną** (stabilną w sensie Dahlquista) – jeżeli metoda jest D-stabilna to jeszcze za mało, żeby była użyteczna ponieważ w praktyce wykorzystujemy kroki $h>0$
- Metoda, która jest stabilna dla danego obszaru $h\lambda \in \Omega$ jest nazywana **absolutnie stabilną** jeżeli powyższe warunki są spełnione dla każdego $h\lambda \in \Omega$
- Metoda absolutnie stabilna w której obszar Ω to lewa półpłaszczyzna płaszczyzny zespolonej $\Omega: \{h\lambda: \operatorname{Re}(h\lambda) \leq 0\}$ nazywana jest **A-stabilną**
- Metoda absolutnie stabilna w której obszar Ω to część lewej półpłaszczyzny płaszczyzny zespolonej nazywana jest **S-stabilną**

Przykład 1:

Przypuśćmy, że mamy daną metodę wielokrokową otwartą rzędu trzeciego:

$$y_{n+1} = -4y_n + 5y_{n-1} + h(4y'_n + 2y'_{n-1})$$

Jaki jest jej obszar stabilności?

Nasze równanie testowe dla $M=1$:

$$y'(t) = \lambda \cdot y(t)$$

Zatem:

$$y_{n+1} = -4y_n + 5y_{n-1} + h\lambda(4y_n + 2y_{n-1})$$

Odpowiadający mu wielomian charakterystyczny:

$$w(z) = z^2 + 4z(1 - \lambda h) - (5 + 2\lambda h) = 0$$

$$z_{1,2} = 2(\lambda h - 1) \pm 2\sqrt{4(\lambda h)^2 - 6\lambda h + 9}$$

Przykład 1 c.d.:

$$z_{1,2} = 2(\lambda h - 1) \pm 2\sqrt{4(\lambda h)^2 - 6\lambda h + 9}$$

Najpierw sprawdzimy czy metoda jest D-stabilna: $h=0$

$$z_{1,2}^{(h=0)} = -2 \pm 2\sqrt{9} = \begin{cases} 1 \\ 4 \end{cases}$$

Ponieważ nie wszystkie pierwiastki spełniają podane warunki metoda jest niestabilna w sensie D-stabilności (ani żadnym innym)

Przykład 2:

Jaka jest stabilność otwartej metody Eulera?

$$y_{n+1} = y_n + h_n y'(t_n)$$

Nasze równanie testowe dla $M=1$:

$$y'(t) = \lambda \cdot y(t)$$

Zatem:

$$y_{n+1} = y_n + h\lambda y_n$$

Odpowiadający mu wielomian charakterystyczny:

$$z - (1 + h\lambda) = 0$$

Zatem z warunku stabilności metoda jest stabilna jeżeli: $|1 + h\lambda| \leq 1$.

Proszę zwrócić uwagę, że krok jest liczbą dodatnią ($h>0$), tak więc metoda Eulera jest zbieżna jeżeli $\lambda < 0$ (warunek absolutnej zbieżności) jeżeli $\lambda \in \mathbb{R}$. Dla ustalonego λ zwiększając krok spowodujemy jej niestabilność.

Metodami A-stabilnymi są np. :

- Zamknięta metoda Eulera

$$y_{n+1} = y_n + hf_{n+1}$$

$$(1 - h\lambda)z = 1$$

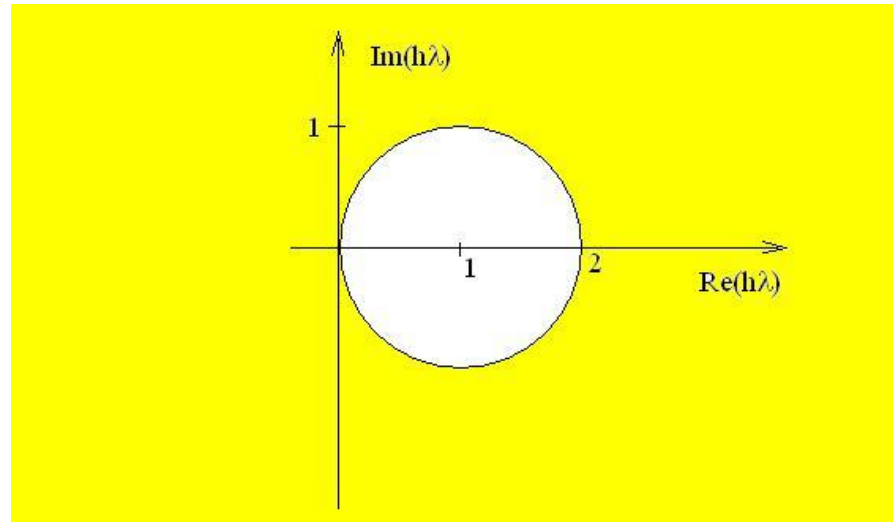
- Metoda Adamsa-Moultona rzędu 2-go (zamknięta metoda trapezów)

$$y_{n+1} = y_n + h(f_{n+1} + f_n)/2$$

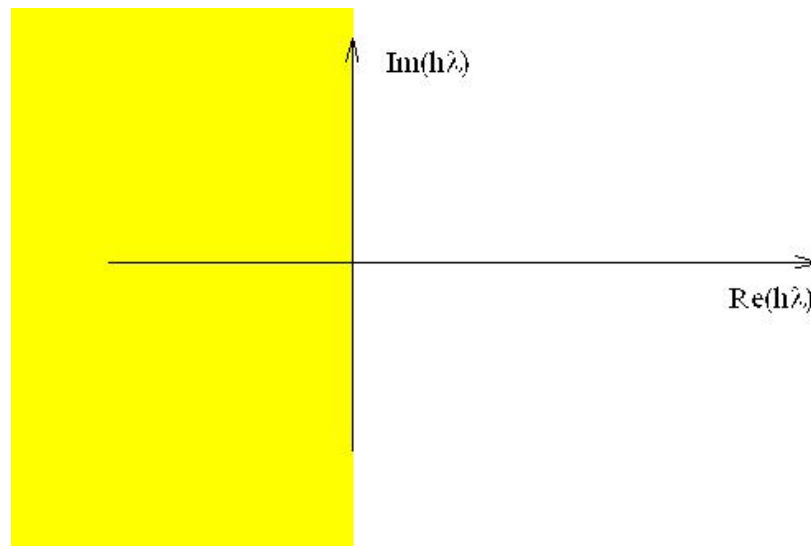
$$(2 - h\lambda)z = (2 + h\lambda)$$

Obszary stabilności:

Zamknięta metoda Eulera (A-stabilna)

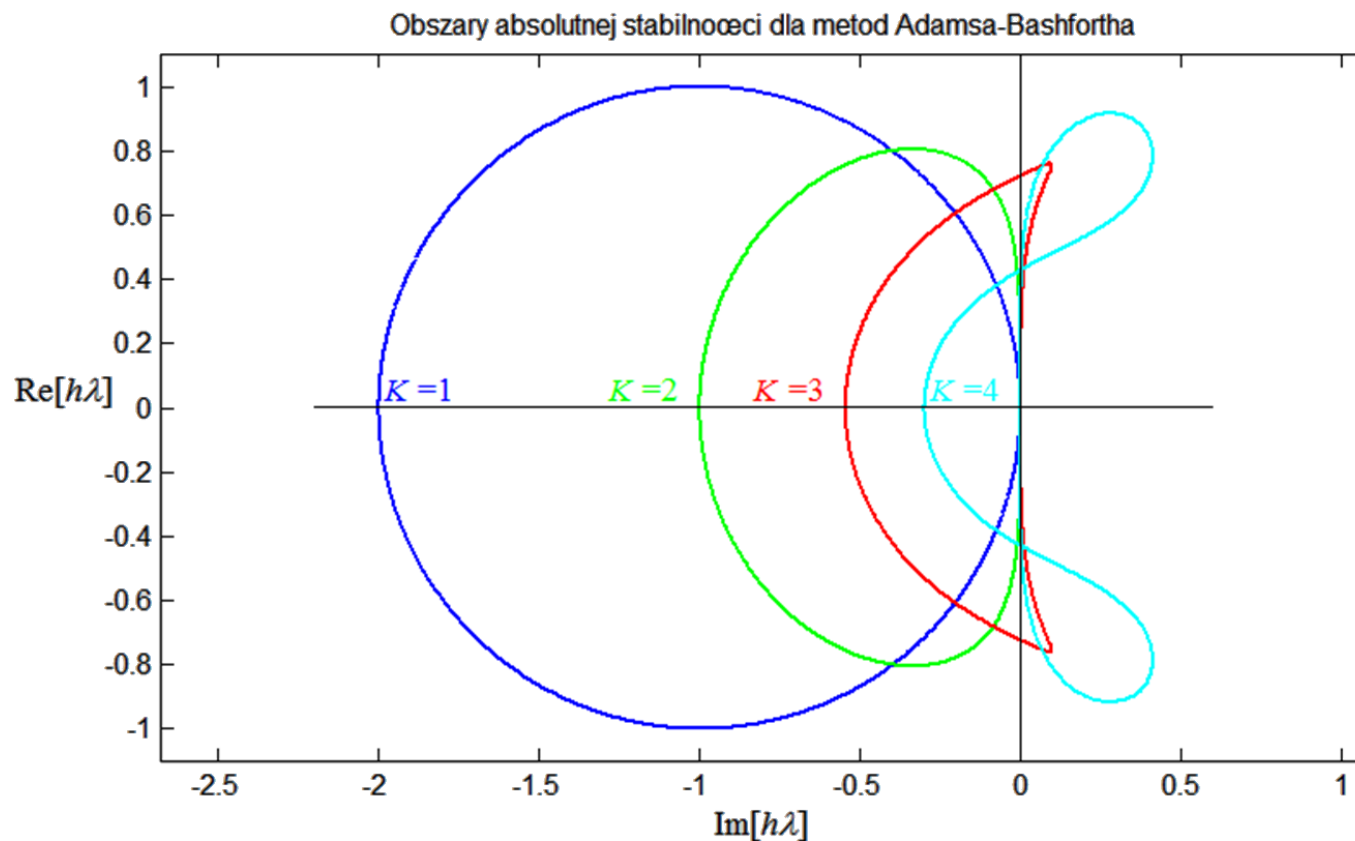


Metoda Adamsa-Moultona rzędu 2-go (A-stabilna)



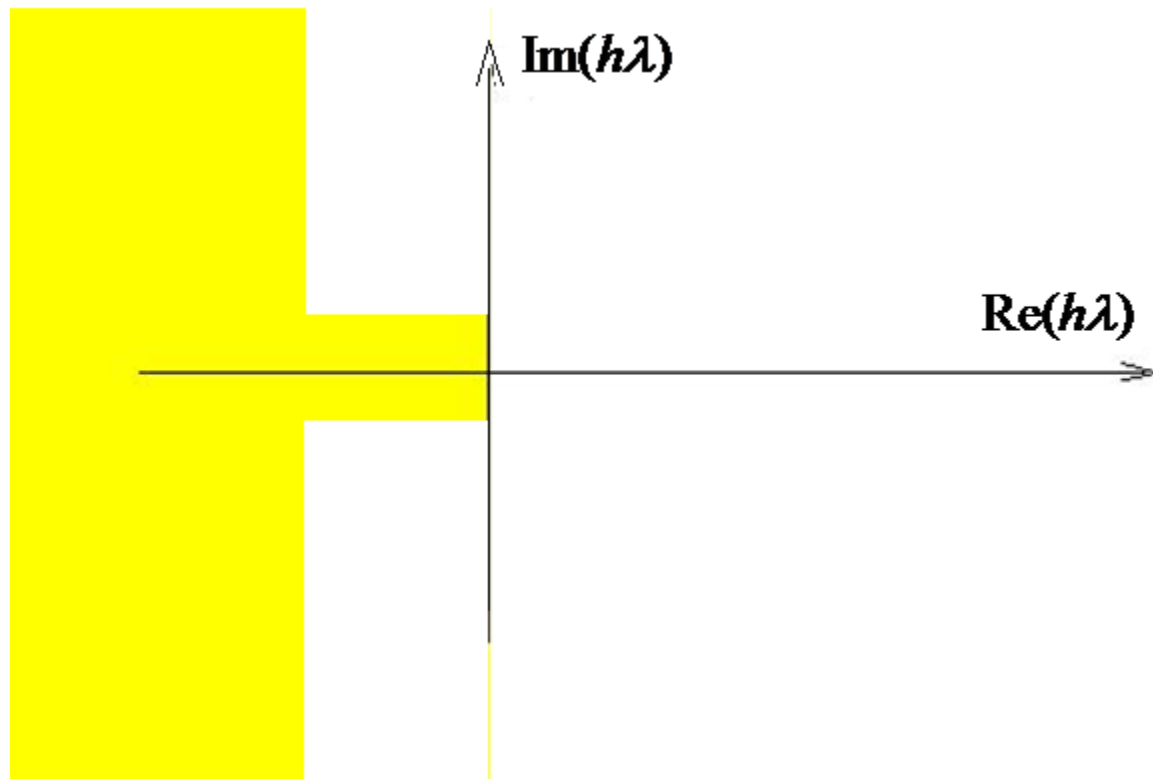
Obszary stabilności:

Metody Adamsa-Bashfortha (absolutnie stabilne)



Obszary stabilności:

Metody Geara (S-stabilne)



STABILNOŚĆ UKŁADÓW RÓWNAŃ:

W praktyce stabilność układu równań sprawdza się przyjmując równanie testowe:

$$\mathbf{y}'(t) = \mathbf{A} \cdot \mathbf{y}(t)$$

Analizę przeprowadza się dla największego $\lambda_n \in \mathbf{A}$ tak jak dla równania pojedynczego.

SCHEMAT PREDYKTOR-KOREKTOR:

Najbardziej praktyczne są metody A-stabilne, ponieważ ich stabilność nie zależy od wielkości kroku (dla $\lambda \leq 0$).

Okazuje się jednak, że metody wielokrokowe rzędów większych niż 2 nie są A-stabilne. Powstaje zatem problem ponieważ jednocześnie chcielibyśmy mieć metodę możliwie wysokiego rzędu (pozwalającą wydłużyć krok przy zadanej dokładności i w ten sposób zmniejszyć ilość obliczeń), z drugiej strony chcielibyśmy mieć metodę, w której wartość kroku może być dowolne (A-stabilną).

Rozsądnym rozwiązaniem tego problemu jest stosowanie metod zamkniętych, które mają wyższy rząd i lepsza stabilność od metod otwartych. W metodach tych trzeba jednak rozwiązać układ równań w każdym kroku potencjalnie metodą iteracyjną. Ilość obliczeń takiej metody (koszt kroku) będzie mniejsza, jeżeli przybliżenie początkowe będzie w miarę dokładne.

Schemat predyktor-korektor to jednoczesne zastosowanie dwóch metod tego samego typu – otwartej i zamkniętej.

1. Najpierw wyznacza się rozwiązanie metodą otwartą (ma ono jednak niższy rząd <większy błąd lokalny>)
2. Uzyskany w p.1 wynik traktuje się jako punkt startowy do rozwiązania układu metodą zamkniętą przy pomocy algorytmu iteracyjnego (np. Newtona-Raphsona)

W pary łączy się:

Metody Adamsa-Bashfortha (otwarte) z metodami Adamsa-Moultona (zamknięte)

Otwarte metody metodami Gear'a z zamkniętymi metodami Gear'a

Wysoki rząd metody oraz A-stabilność można uzyskać również dla niektórych metod zamkniętych Rungego-Kutty (metody otwarte są zawsze absolutnie stabilne, ale nie A-stabilne), które posiadają rząd wyższy od 2. Przykładem może być metoda zamknięta RK-3:

$$\xi_1 = \mathbf{y}_n + \frac{1}{4}h \left[\mathbf{f}(t_n, \xi_1) - \mathbf{f}\left(t_n + \frac{2}{3}h, \xi_2\right) \right]$$

$$\xi_2 = \mathbf{y}_n + \frac{1}{12}h \left[3\mathbf{f}(t_n, \xi_1) + 5\mathbf{f}\left(t_n + \frac{2}{3}h, \xi_2\right) \right]$$

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \frac{1}{4}h \left[\mathbf{f}(t_n, \xi_1) + 3\mathbf{f}\left(t_n + \frac{2}{3}h, \xi_2\right) \right]$$

SZTYWNE UKŁADY RÓWNAŃ RÓŻNICZKOWYCH:

Układ równań różniczkowych nazywamy sztywnym, jeżeli wartości własne macierzy Jakobiego **J** spełniają następujące warunki:

$$J = \begin{bmatrix} \frac{\delta f_1}{\delta y_1} & \frac{\delta f_1}{\delta y_2} & \dots & \frac{\delta f_1}{\delta y_M} \\ \frac{\delta f_2}{\delta y_1} & \frac{\delta f_2}{\delta y_2} & \dots & \frac{\delta f_2}{\delta y_M} \\ \dots & \dots & \dots & \dots \\ \frac{\delta f_M}{\delta y_1} & \frac{\delta f_M}{\delta y_2} & \dots & \frac{\delta f_M}{\delta y_M} \end{bmatrix}$$

$$\frac{\lambda_n}{\lambda_1} \gg 1$$

W układach sztywnych spotykamy się z następującymi problemami:

- Z analizy stabilności tego typu układów wynika, że maksymalna dopuszczalna wartość kroku jest bardzo mała w stosunku do założonej dokładności.
- Ze względu na mały krok obliczenie wyniku (typowymi metodami stosunkowo niskich rzędów) jest bardzo pracochłonne i długo trwa

Do takich układów stosuje się metody zaproponowane przez Geara, które mają wyższy rząd niż metody RK lub metody Adamsa. Metody Geara wyższych rzędów nie są jednak A-stabilne, ale S-stabilne.

Metody Geara:

$$\mathbf{y}_{n+1} = h\beta_0 \mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}) + \sum_{k=1}^K \alpha_k \mathbf{y}_{n+1-k}$$

gdzie α_k, β_0 są dobrane tak, by metoda ta była rzędu K :

$$\beta_0 = \frac{2}{K(K+1)}, \quad \alpha_k = \frac{\beta_0}{k} \prod_{\substack{j=1 \\ j \neq k}}^K \frac{j}{j-k} \quad \text{dla } k = 1, \dots, K$$

Dla $K=1$: A-stabilna zamknięta metoda Eulera.

Dla $K=2$: A-stabilna metoda drugiego rzędu:

$$\mathbf{y}_{n+1} = \frac{2}{3} h \mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}) + \frac{4}{3} \mathbf{y}_n - \frac{1}{3} \mathbf{y}_{n-1}$$

Metody Geara:

Otwarte:

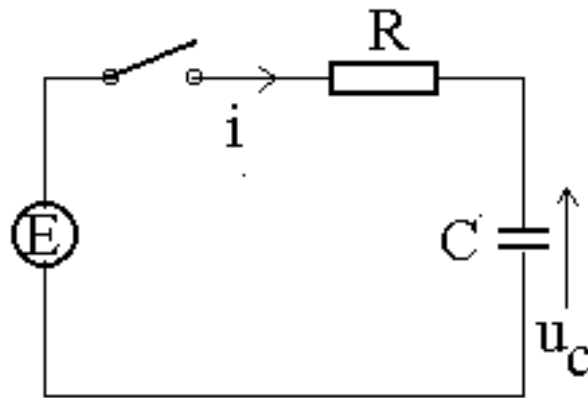
| K | β_0 | α_1 | α_2 | α_3 | α_4 | α_5 | α_6 | p | C_{p+1} |
|----------|-----------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|----------|-----------------------------|
| 1 | 1 | 1 | | | | | | 1 | -1/2 |
| 2 | 2 | 0 | 1 | | | | | 2 | -1/3 |
| 3 | 3 | -3/2 | 3 | -1/2 | | | | 3 | -1/4 |
| 4 | 4 | -10/3 | 6 | -2 | 1/3 | | | 4 | -1/5 |
| 5 | 5 | -65/12 | 10 | -5 | 5/3 | -1/4 | | 5 | -1/6 |
| 6 | 6 | -77/10 | 15 | -10 | 5 | -3/2 | 1/5 | 6 | -1/7 |

Zamknięte:

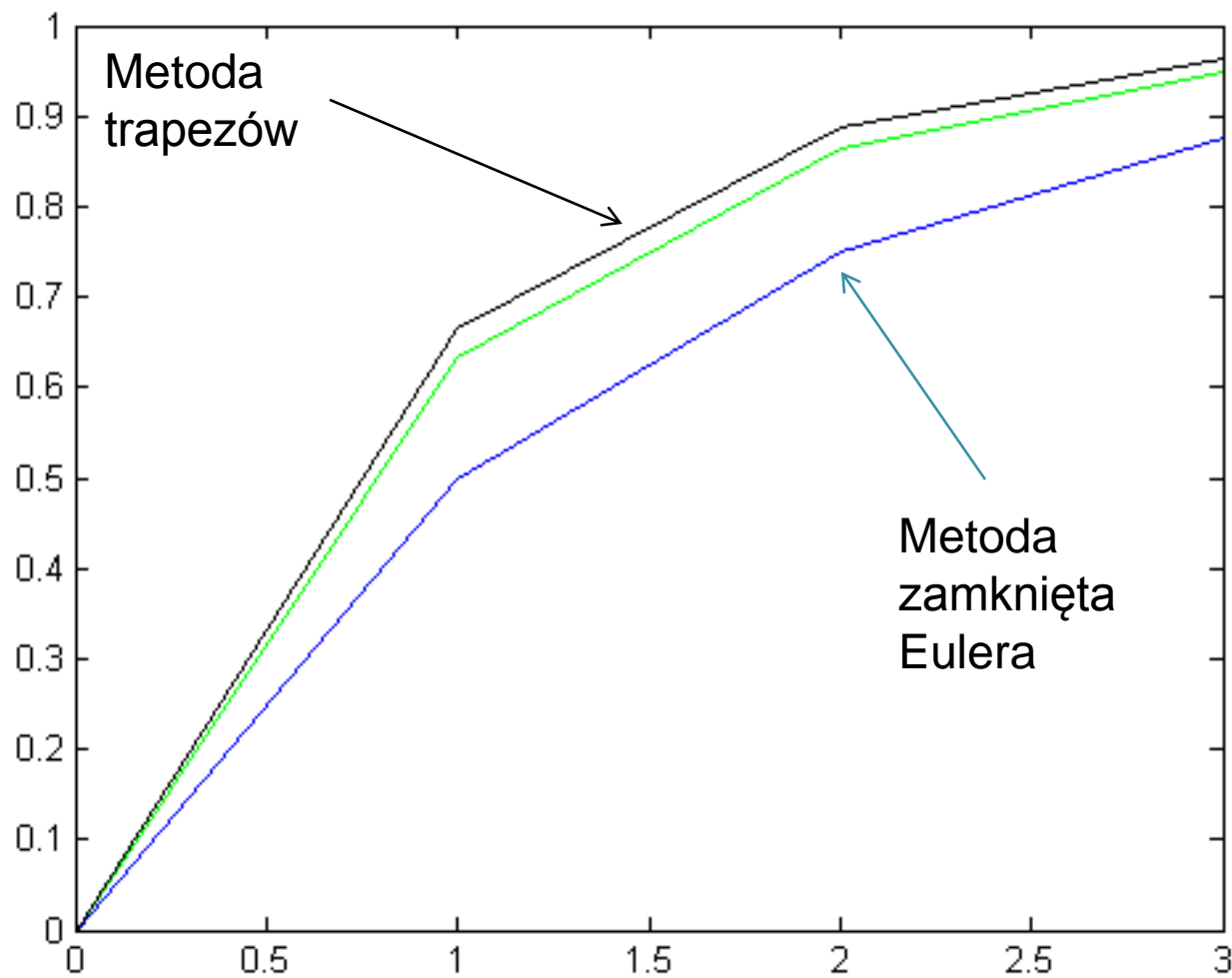
| K | β_0 | α_1 | α_2 | α_3 | α_4 | α_5 | α_6 | p | C_{p+1} |
|---|-----------|------------|------------|------------|------------|------------|------------|---|-----------|
| 1 | 1 | 1 | | | | | | 1 | -1/2 |
| 2 | 2/3 | 4/3 | -1/3 | | | | | 2 | -1/3 |
| 3 | 6/11 | 18/11 | -9/11 | 2/11 | | | | 3 | -1/4 |
| 4 | 12/25 | 48/25 | -36/25 | 16/25 | -3/25 | | | 4 | -1/5 |
| 5 | 60/137 | 300/137 | 0 | 200/137 | -75/137 | 12/137 | | 5 | -1/6 |
| 6 | 60/147 | 360/147 | 0 | 400/147 | 0 | 72/147 | -10/147 | 6 | -1/7 |

Przykład:

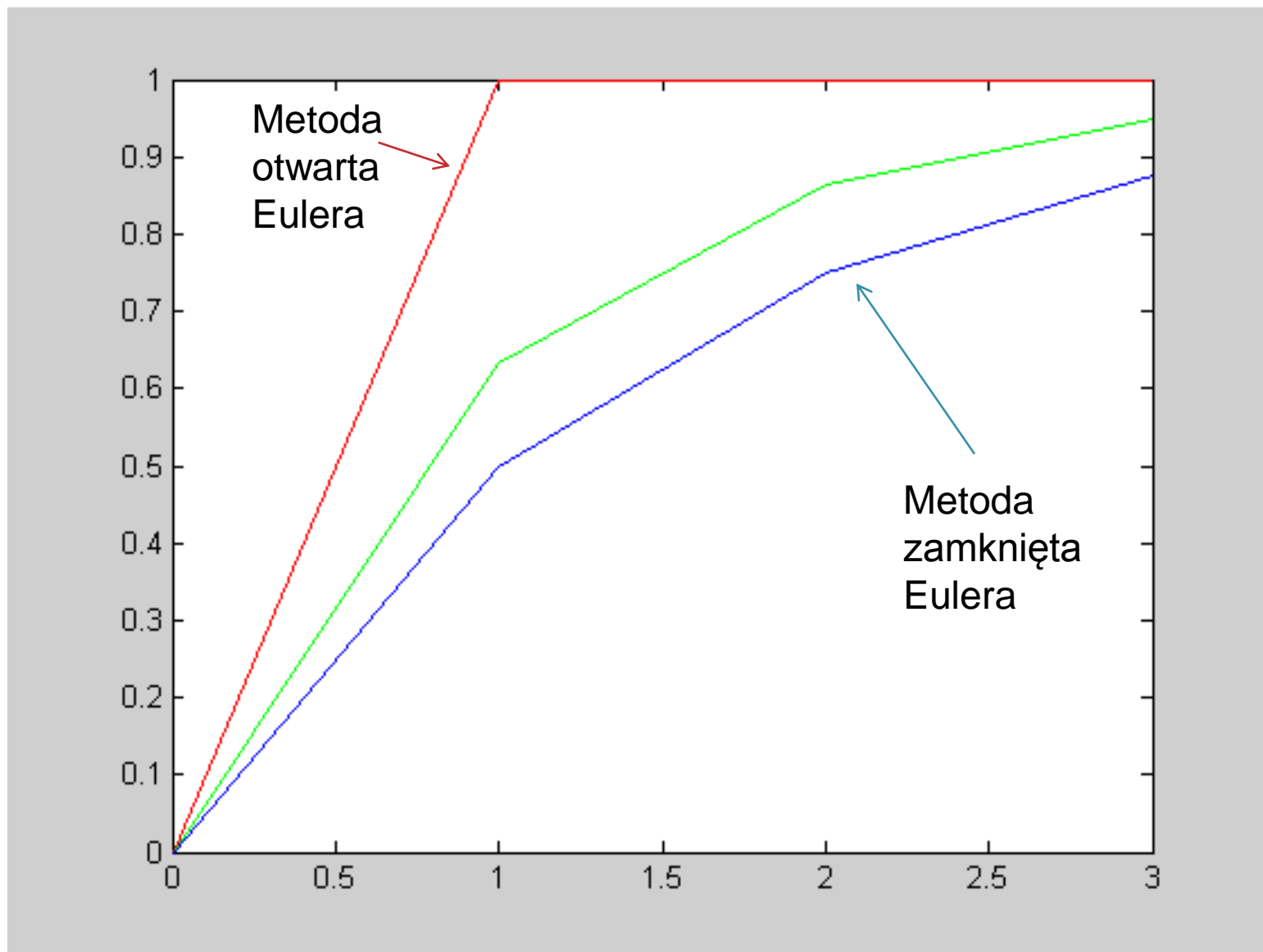
Dla obwodu elektrycznego jak na rysunku: $R=1\text{ M}\Omega$, $C=1\mu\text{F}$, $E=1\text{V}$ wyznaczyć zależność napięcie na kondensatorze w funkcji czasu dla $t \in <0,3> [\text{s}]$ metodami otwarta i zamkniętą Eulera oraz metodą trapezów ze stałym krokiem całkowania. Jako warunek początkowy przyjąć $u_{c0}=0\text{ V}$. Narysować przebiegi napięcia dokładnego i wyznaczonego metodami Eulera oraz metodą trapezów dla kroku $h=(t_k-t_0)/4$



Przykład c.d:

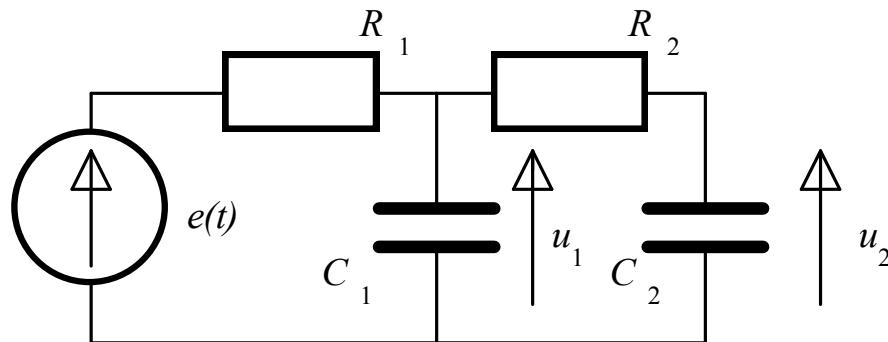


Przykład c.d:



Przykład:

Znaleźć odpowiedź układu metodą otwartą Eulera, przyjmując: $R_1=1\text{k}\Omega$, $R_2=1\text{M}\Omega$, $C_1=C_2=1\mu\text{F}$. $E=1(t)$



Odpowiadający układowi układ równań różniczkowych:

$$\begin{bmatrix} u_1' \\ u_2' \end{bmatrix} = \begin{bmatrix} -\frac{R_1 + R_2}{R_1 R_2 C_1} & \frac{1}{R_2 C_1} \\ \frac{1}{R_2 C_2} & -\frac{1}{R_2 C_2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} \frac{e(t)}{R_1 C_1} \\ 0 \end{bmatrix}$$

Przykład c.d.:

$$\begin{bmatrix} u_1' \\ u_2' \end{bmatrix} = \begin{bmatrix} -\frac{R_1 + R_2}{R_1 R_2 C_1} & \frac{1}{R_2 C_1} \\ \frac{1}{R_2 C_2} & -\frac{1}{R_2 C_2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} \frac{e(t)}{R_1 C_1} \\ 0 \end{bmatrix}$$

Wstawiając podane wartości:

$$\begin{bmatrix} u_1'(t) \\ u_2'(t) \end{bmatrix} = \begin{bmatrix} -1001 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} + \begin{bmatrix} 10^3 \\ 0 \end{bmatrix} \text{ dla } t \in [0, T], \quad \begin{bmatrix} u_1(0) \\ u_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Wartości macierzy Jakobiego:

$$\lambda_1 = -501 + \sqrt{250001} \approx -0.999$$

$$\lambda_2 = -501 - \sqrt{250001} \approx -1001$$

Przykład c.d.:

Jest to zatem układ sztywny o znanych rozwiązaniach ogólnych:

$$u_2(t) = c_1 e^{-\lambda_1 t} + c_2 e^{-\lambda_2 t} + 1, \quad \text{dla } t \geq 0$$

$$c_1 = \lambda_2 / (\lambda_1 - \lambda_2) \approx -1.001$$

$$c_2 = -1 - c_1 \approx 0.001$$

$$u_1(t) = u_2'(t) + u_2(t) = c_1(1 - \lambda_1)e^{-\lambda_1 t} + c_2(1 - \lambda_2)e^{-\lambda_2 t} + 1, \quad \text{dla } t \geq 0$$

Przykład c.d.:

Rozwiązania równań a) dokładne b) przybliżone dla kroku $1.5e-3$ c) przybliżone dla kroku $1.5e-4$

