

Jawny schemat Eulera

$$u(t = 0) = u_0$$
$$\frac{du(t)}{dt} = f(t, u)$$

$$u(t + \Delta t) = u(t) + \Delta t \frac{du(t)}{dt} + \frac{\Delta t^2}{2} \frac{d^2 u(t)}{dt^2} + \frac{\Delta t^3}{6} \frac{d^3 u(t)}{dt^3} + O(\Delta t^4)$$

$$u(t + \Delta t) = u(t) + f(t, u) \Delta t + O(\Delta t^2)$$

przepis na pojedynczy krok
z $u(t)$ do $u(t + \Delta t)$

można wyliczyć bo znamy
 t i $u(t)$

błąd lokalny jawnego Eulera w kroku $t_{n-1} \rightarrow t_n$
wg tw. Taylora

$$l_n = \frac{\Delta t^2}{2} u''(\xi_n)$$

$$\xi_n \in (t_{n-1}, t_n)$$

błąd lokalny schematu różnicowego
definicja:

odchylenie wyniku numerycznego
od dokładnego uzyskane w pojedynczym
kroku, w którym na starcie wstawiono
dokładne rozwiązanie
(dokładny warunek początkowy)

Jawny schemat Eulera

$$u(t = 0) = u_0$$
$$\frac{du(t)}{dt} = f(t, u)$$

$$u(t + \Delta t) = u(t) + f(t, u)\Delta t + O(\Delta t^2)$$

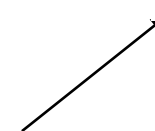
stosowany wielokrotnie:

$$u_1 = u_0 + \Delta t f(0, u_0)$$

$$u_2 = u_1 + \Delta t f(\Delta t, u_1)$$

...

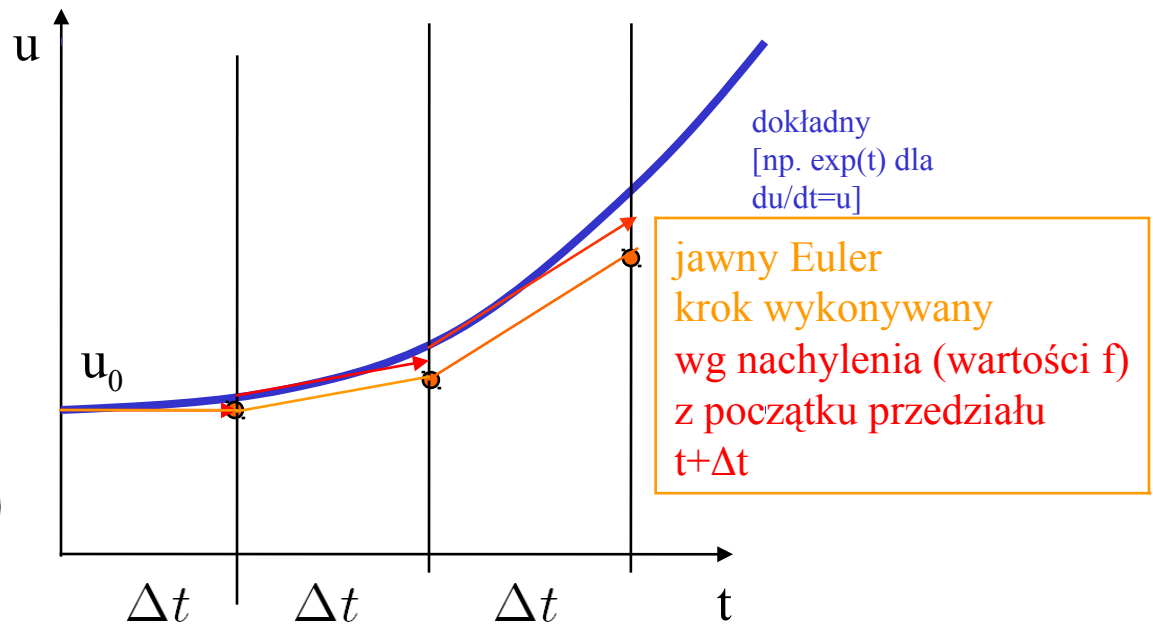
$$u_n = u_{n-1} + \Delta t f(t_{n-1}, u_{n-1})$$

$$t_{n-1} = (n - 1)\Delta t$$


krok wcale nie musi być taki sam dla każdego n ,
ale tak przyjmiemy do analizy

Jawny schemat Eulera

$$\begin{aligned}u_1 &= u_0 + \Delta t f(0, u_0) \\u_2 &= u_1 + \Delta t f(\Delta t, u_1) \\u_n &= u_{n-1} + \Delta t f(t_{n-1}, u_{n-1})\end{aligned}$$



*każdy krok wykonywany z nachyleniem
branym z chwili, w której krok się zaczyna*

W każdym kroku wprowadzamy nowy błąd. Błędy się akumulują.
Zmniejszamy krok Δt : Błąd lokalny zmaleje, ale wykonamy więcej kroków.
Czy opłaca się zmniejszać kroki czasowe?

Definicja: Błąd globalny e = rozwiązanie dokładne w chwili t minus rozwiązanie numeryczne w chwili t

„Czy się opłaca” znaczy: Czy błąd globalny maleje gdy Δt maleje?
a jeśli tak - czy maleje do zera? („czy możliwe jest dokładne rozwiązanie
równania różniczkowego uzyskane jako
granica schematu różnicowego”)

Jawny schemat Eulera

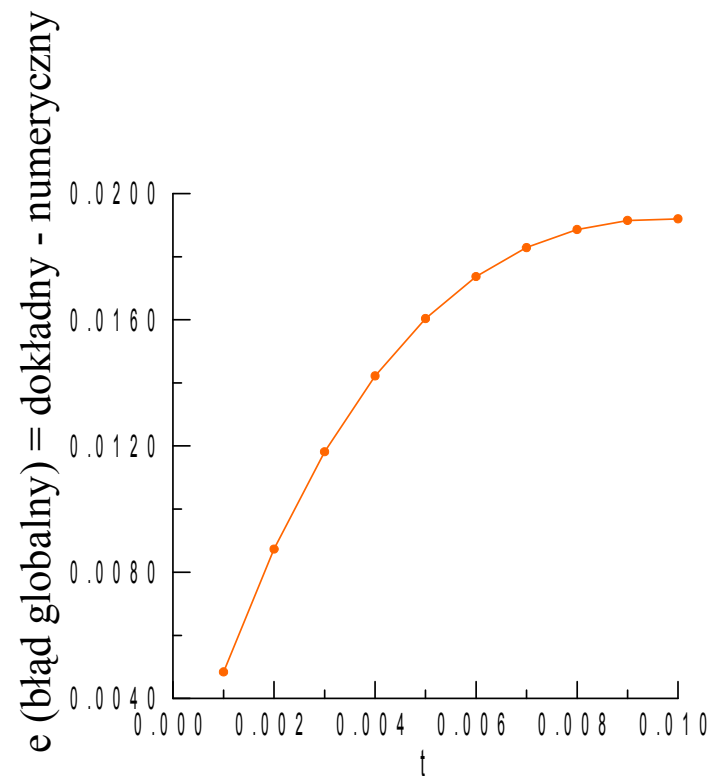
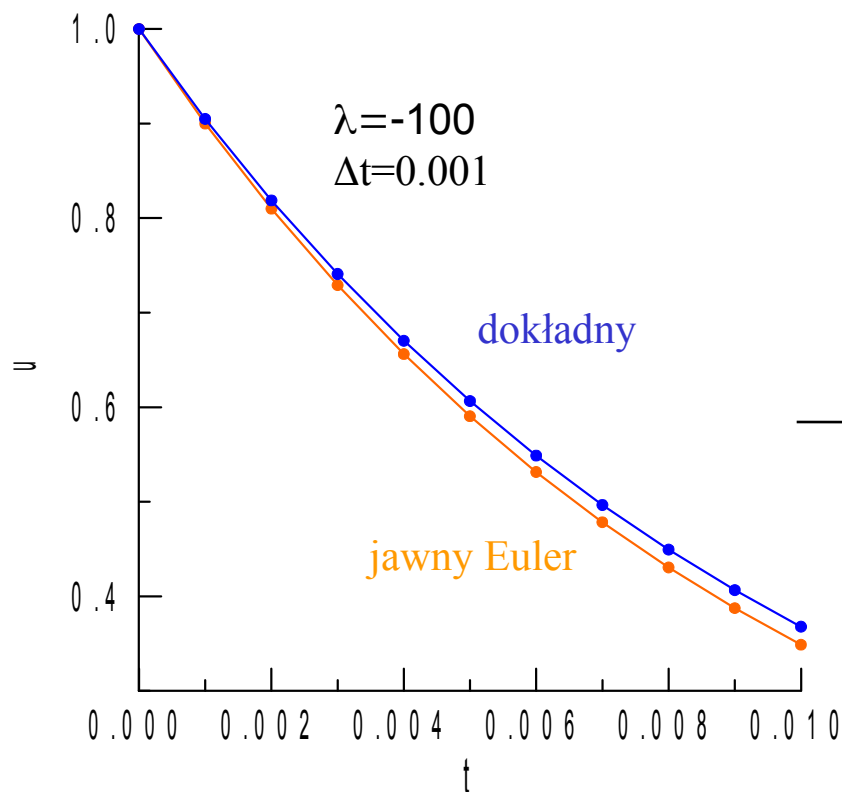
Czy błąd całkowity maleje gdy Δt maleje ?
Czy maleje do zera?

eksperyment numeryczny

problem początkowy:

$$u' = \lambda u, \quad u(0) = 1$$

z rozwiązaniem dokładnym $u(t) = \exp(\lambda t)$



Jawny schemat Eulera

Czy błąd globalny maleje gdy Δt maleje ?
Czy maleje do zera?

eksperyment numeryczny

problem początkowy:

$$u' = \lambda u, \quad u(0) = 1$$

z rozwiązaniem dokładnym $u(t) = \exp(\lambda t)$

$$\lambda = -100$$

zmniejszajmy krok czasowy, jaki wynik w chwili $t=0.01$? [$1/e = .3678794$]

n	Δt	u_n	$\exp(-1) - u_n$
10	10^{-3}	0.34867	1.920×10^{-2}
10^2	10^{-4}	0.36603	1.847×10^{-3}
10^3	10^{-5}	0.36769	1.840×10^{-4}
10^4	10^{-6}	0.36784	1.839×10^{-5}

błąd globalny w chwili $t=0.01$ wydaje się zmieniać liniowo z krokiem czasowym

$$u(t + \Delta t) = u(t) + f(t, u)\Delta t + O(\Delta t^2)$$

interpretacja: błąd lokalny rzędu Δt^2
popołniony $n = t/\Delta t$ razy daje błąd globalny rzędu Δt

zmniejszajmy krok czasowy, jaki wynik w chwili $t=0.01$? [$1/e=.3678794$]

n	Δt	u_n	$\exp(-1)-u_n$
10	10^{-3}	0.34867	1.920×10^{-2}
10^2	10^{-4}	0.36603	1.847×10^{-3}
10^3	10^{-5}	0.36769	1.840×10^{-4}
10^4	10^{-6}	0.36784	1.839×10^{-5}

błąd globalny w chwili $t=0.01$ wydaje się zmieniać liniowo z krokiem czasowym

Definicja:

Metody różnicowa jest **zbieżna** jeśli błąd globalny

$$\lim_{n \rightarrow \infty, \Delta t \rightarrow 0, n \Delta t = T} |e_n| = 0$$

znika do zera w chwili T gdy z Δt do 0

Definicja:

Metody różnicowa jest **zbieżna** jeśli błąd globalny

$$\lim_{n \rightarrow \infty, \Delta t \rightarrow 0, n\Delta t = T} |e_n| = 0$$

znika do zera w chwili T gdy z Δt do 0

rzęd zbieżności metody jest k , jeśli w granicy $\Delta t \rightarrow 0$ błąd globalny znika jak $O(\Delta t^k)$

Zazwyczaj rząd zbieżności metody jest o jeden niższy niż rząd błędu lokalnego (popelnianego w jednym kroku).

Rząd zbieżności metody Eulera – pierwszy.

Metoda Eulera jest metodą pierwszego rzędu.

Nie ma wolniej zbieżnej metody

Arytmetyka zmiennoprzecinkowa: $-1^z \times m \times 2^w$

32-bitowa pojedyncza precyzja:

rozmiar	nazwa
---------	-------

1 bit	znak (z)
-------	--------------

8 bits	wykładnik (w)
--------	-------------------

23 bits	mantysa (m)
---------	-----------------

mantysa w binarnym zapisie $m=10101 \dots$

oznacza $m=1*2^0+0*2^{-1}+1*2^{-2}+0*2^{-3}+1*2^{-4}+\dots$

Wykładnik: $w=2^8=256$ wartości. Kodowane z tzw. Bias -127, tak, że 2^w w zakresie od 2^{-127} do 2^{127} , w przybliżeniu 10^{-38} do 10^{38}

Liczby zmiennoprzecinkowe są położone dyskretnie na osi x .

Liczby rzeczywiste reprezentowane są przez zmiennoprzecinkowe w przybliżeniu, z błędem zaokrągleń.

Arytmetyka zmiennoprzecinkowa:

$$-1^z \times m \times 2^w$$

rozmiar	nazwa
1 bit	znak (z)
8 bits	wykładnik (w)
23 bits	mantysa (m)



zbieżność metody różnicowej a błędy zaokrągleń
(skończona dokładność arytmetyki
zmiennoprzecinkowej)

błędy zaokrągleń a zbieżność

do tej pory zakładaliśmy, że błędy zaokrągleń nie ma (że arytmetyka dokładna)
 arytmetyka zmiennoprzecinkowa nie jest dokładna.

pojedyncza precyzja: 32 bity
 podwójna : 64 bity

Joseph E. Flaherty
 Ordinary Differential Equations

arytmetyka 21 - bitowa →

błąd minimalny →

zmniejszanie kroku
 czasowego nie poprawi już
 wyniku

k	$N = 2^k$	$h = 1/N$	\tilde{y}_N	\tilde{e}_N
0	1	1.00000000	2.00000000	-0.71828181
1	2	0.50000000	2.25000000	-0.46828184
2	4	0.25000000	2.44140625	-0.27687559
3	8	0.12500000	2.56578445	-0.15249738
4	16	0.06250000	2.63792896	-0.08035287
5	32	0.03125000	2.67698956	-0.04129227
6	64	0.01562500	2.69734669	-0.02093514
7	128	0.00781250	2.70773602	-0.01054581
8	256	0.00390625	2.71297836	-0.00530347
9	512	0.00195312	2.71561337	-0.00266846
10	1024	0.00097656	2.71694279	-0.00133904
11	2048	0.00048828	2.71764278	-0.00063904
12	4096	0.00024414	2.71795559	-0.00032624
13	8192	0.00012207	2.71811104	-0.00017079
14	16384	0.00006104	2.71814919	-0.00013264
15	32768	0.00003052	2.71804428	-0.00023755
16	65536	0.00001526	2.71732903	-0.00095280

Table 2.1.4: Solutions of $y' = y$, $y(0) = 1$, at $t = 1$ obtained by Euler's method with 21-bit rounded arithmetic.

błędy zaokrągleń a metody różnicowe

$u(t_n)$ rozwiązanie równania różniczkowego w chwili t_n

u_n rozwiązanie równania różnicowego z dokładną arytmetyką

\tilde{u}_n rozwiązanie uzyskane z arytmetyką skończonej dokładności

$$\tilde{e}_n = u(t_n) - \tilde{u}_n \quad \text{błąd całkowity}$$

$$e_n = u(t_n) - u_n \quad \text{błąd globalny (jak wcześniej zdefiniowano)}$$

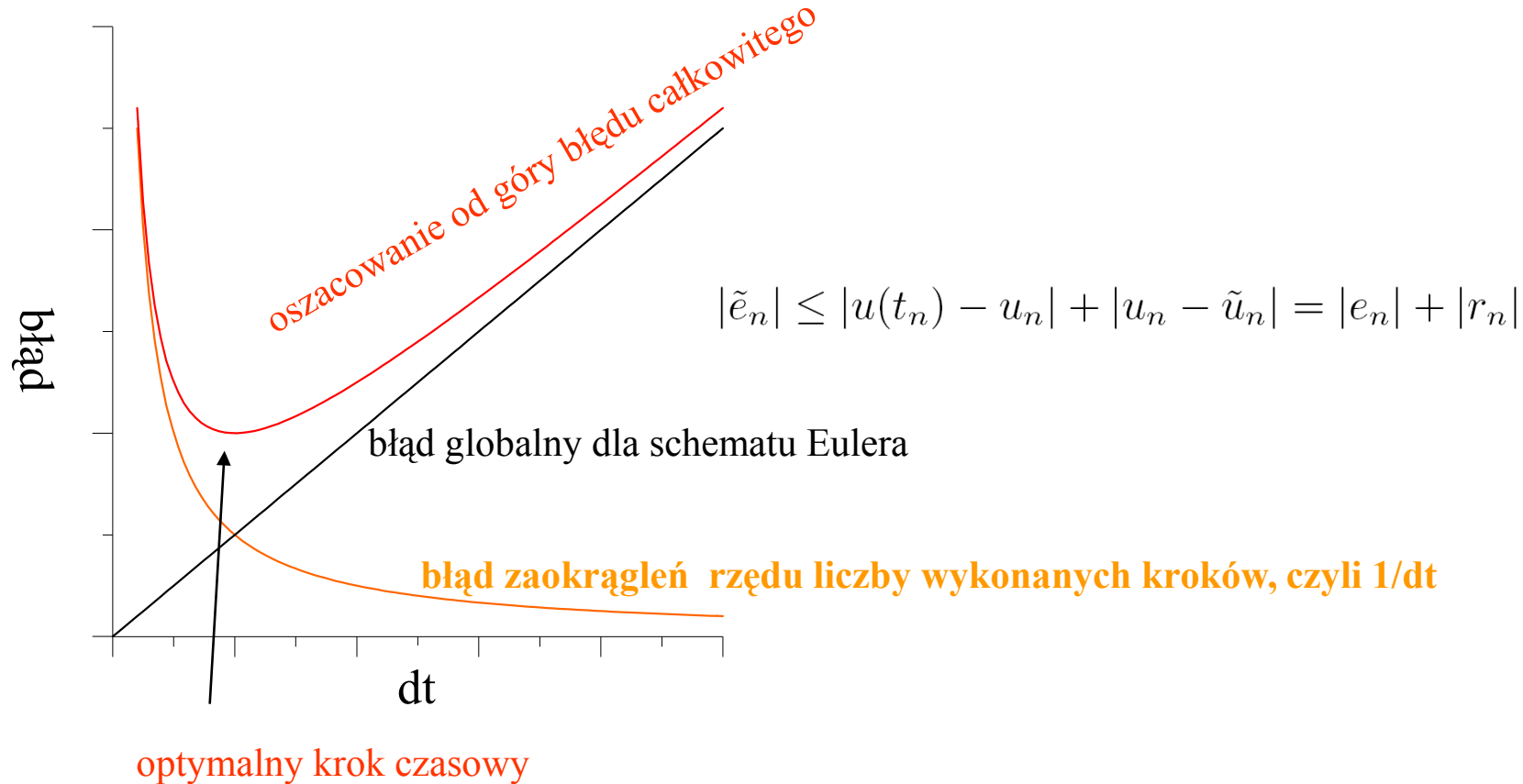
$$r_n = u_n - \tilde{u}_n \quad \text{błąd zaokrąglenia}$$

$$\tilde{e}_n = u(t_n) - u_n + u_n - \tilde{u}_n$$

$$|\tilde{e}_n| \leq |u(t_n) - u_n| + |u_n - \tilde{u}_n| = |e_n| + |r_n|$$

oszacowanie od góry błędu całkowitego

błędy zaokrągleń a metody różnicowe

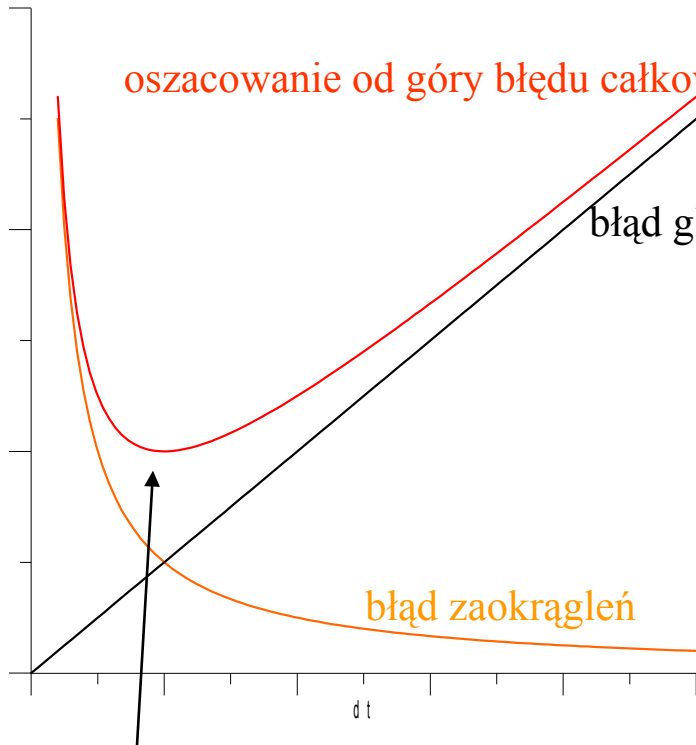


błędy zaokrągleń dają o sobie znać gdy wykonamy zbyt wiele kroków

remedium: używać się schematów o wyższym rzędzie zbieżności niż pierwszy.

Poniżej pewnego dt nie warto schodzić – nie tylko ze względu na czas obliczeń, ale i dokładność

błędy zaokrągleń a metody różnicowe



$$\tilde{e}_n = u(t_n) - \tilde{u}_n \quad \text{błąd całkowity}$$

$$e_n = u(t_n) - u_n \quad \text{błąd globalny}$$

$$r_n = u_n - \tilde{u}_n \quad \text{błąd zaokrąglenia}$$

optymalny krok czasowy

Definicja:

Metody różnicowa jest **zbieżna** jeśli błąd globalny

$$\lim_{n \rightarrow \infty, \Delta t \rightarrow 0, n\Delta t = T} |e_n| = 0$$

znika do zera w chwili T gdy Δt do 0

uwaga:

definicja zbieżności
dotyczy błędu globalnego
a nie całkowitego

stabilność bezwzględną wzoru trapezów

problem modelowy:

$$\frac{du}{dt} = \lambda u \quad \text{WP: } u(t=0)=1.$$

rozwiązanie $u=\exp(\lambda t)$

$$u_n = u_{n-1} + \frac{\Delta t}{2} \lambda (u_{n-1} + u_n)$$

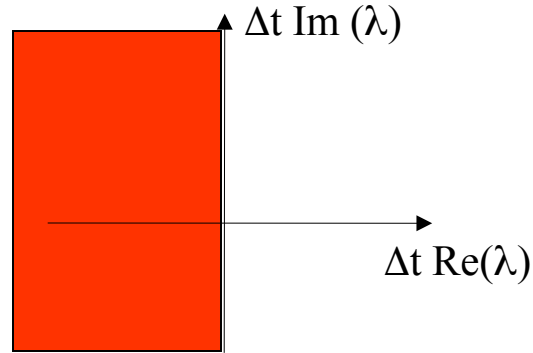
$$u_n = \frac{1 + \frac{\Delta t}{2} \lambda}{1 - \frac{\Delta t}{2} \lambda} u_{n-1}$$

$$u_n = \left(\frac{1 + \frac{\Delta t}{2} \lambda}{1 - \frac{\Delta t}{2} \lambda} \right)^n u_0$$
$$\left| \frac{1 + \lambda \frac{\Delta t}{2}}{1 - \lambda \frac{\Delta t}{2}} \right| \leq 1$$

$$|2 + z| \leq |2 - z|$$

zbiór punktów na p. Gaussa, które są nie dalej od $(-2,0)$ niż od $(0,2)$

region bzwz. stabilności wzoru trapezów

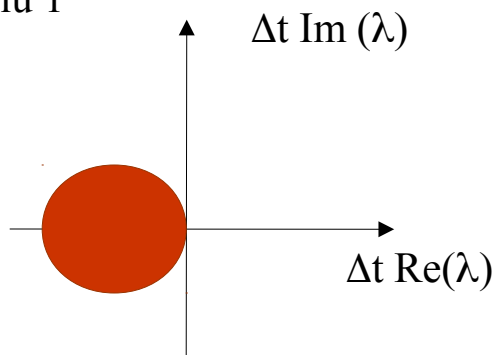


Wniosek: dla $\lambda < 0$ wzór trapezów bezwzględnie stabilny dla dowolnego kroku czasowego ! A-stabilny

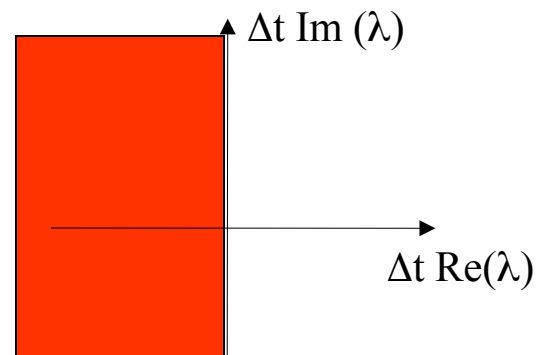
druga bariera Dahlquist: maksymalny rząd dokładności metody A-stabilnej = 2
schemat trapezów jest najdokładniejszą metodą A-stabilną spośród liniowych metod wielokrokowych

Implementowana np. w SPICE.

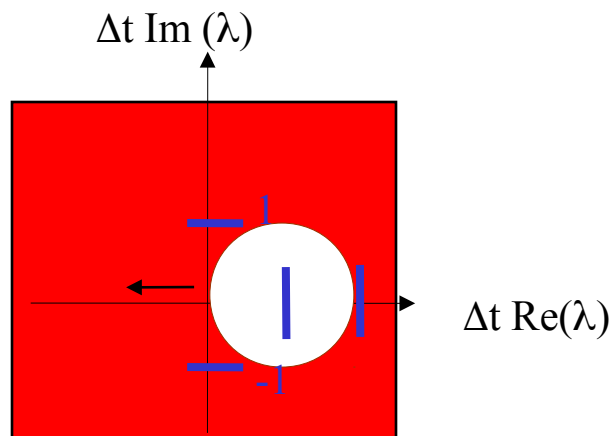
region bzwz. stabilności Eulera:
 koło o promieniu 1
 i środku $(-1,0)$



region bzwz. stabilności wzoru trapezów



niejawna metoda Eulera:
 region bezwzględnej stabilności



$$\left| \frac{du}{dt} = f \right|$$

poznane metody:

$$1) u_{n+1} = u_n + \Delta t f(t_n, u_n) + O(\Delta t^2) \quad \left| \right.$$

$$2) u_{n+1} = u_n + \Delta t f(t_{n+1}, u_{n+1}) + O(\Delta t^2)$$

$$3) u_{n+1} = u_n + \frac{\Delta t}{2} (f(t_{n+1}, u_{n+1}) + f(t_n, u_n)) + O(\Delta t^3) \quad \left| \right.$$

Poznane metody: jednokrokowe (1-3), jawna (1) i niejawne (2-3), pierwszego (1-2) i drugiego (3) rzędu dokładności
Metody (2-3) A stabilne, metoda (2) nadstabilna

jawne metody różnicowe wysokiej dokładności ??

jawne metody jednokrokowe wyższego rzędu dokładności niż jawny Euler

$$u' = f(t, u), \quad u(0) = u_0$$

rozwiniecie Taylora ponownie:

$$u(t_n) = u(t_{n-1}) + \Delta t u'(t_{n-1}) + \frac{\Delta t^2}{2} u''(t_{n-1}) + \dots + \frac{\Delta t^k}{k!} u^{(k)}(t_{n-1}) + O(\Delta t^{k+1})$$

liczymy pochodne:

$$u'(t_{n-1}) = f(t_{n-1}, u(t_{n-1})) \quad \text{z RR.}$$

RR różniczkujemy po czasie

$$u''(t_{n-1}) = [f'_t + f'_u u']_{(t_{n-1}, u_{n-1})}$$

czyli

$$u''(t_{n-1}) = [f'_t + f'_u f]_{(t_{n-1}, u_{n-1})}$$

podobnie

$$u'''(t_{n-1}) = [f''_{tt} + 2f''_{tu} f + f''_{uu} f^2 + (f'_u)^2 f + f'_u f'_t + f''_{tu} f]_{(t_{n-1}, u_{n-1})}$$

Zależnie od tego gdzie się zatrzymamy uzyskamy błąd lokalny danego rzędu

$$u(t_n) = u(t_{n-1}) + \Delta t u'(t_{n-1}) + \frac{\Delta t^2}{2} u''(t_{n-1}) + \dots + \frac{\Delta t^k}{k!} u^{(k)}(t_{n-1}) + O(\Delta t^{k+1})$$

Zależnie od tego gdzie się zatrzymamy uzyskamy błąd lokalny zadanego rzędu np.

$$u(t_n) = u(t_{n-1}) + \Delta t f(t_{n-1}, u_{n-1}) + \frac{\Delta t^2}{2} [f'_t + f'_u f]_{(t_{n-1}, u_{n-1})} + \frac{\Delta t^3}{6} u'''(\xi_n)$$

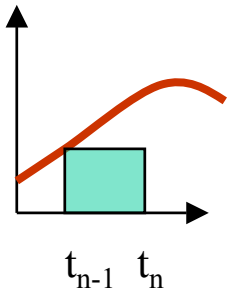
pomysł: mało przydatny w praktyce ze względu na konieczność analitycznego wyliczenia pochodnych cząstkowych f .

Dla metod ogólnych: nie powinniśmy liczyć, że f jest dane wzorem

podjęcie alternatywne: inspirowane całkowaniem

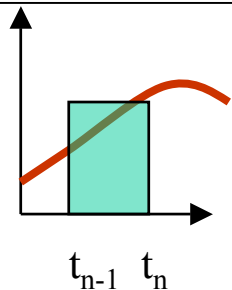
prawa strona = funkcja tylko t $u' = f(t), u(0) = u_0$ |

z rozwiązaniem: $u(t) = u_0 + \int_0^t f(\tau) d\tau$ |



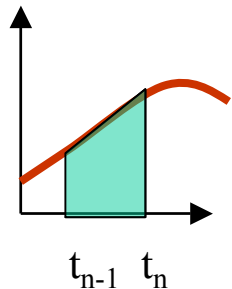
jeśli zastąpimy całkę kwadraturą
prostokątów z wywołaniem funkcji w lewym końcu przedziału

$u(t_n) = u(t_{n-1}) + \Delta t f(t_{n-1}) + O(\Delta t^2)$ - rozpoznajemy jawny schemat Eulera



kwadratura prostokątów z wywołaniem funkcji w prawym końcu przedziału

$u(t_n) = u(t_{n-1}) + \Delta t f(t_n) + O(\Delta t^2)$ - rozpoznajemy niejawny schemat Eulera

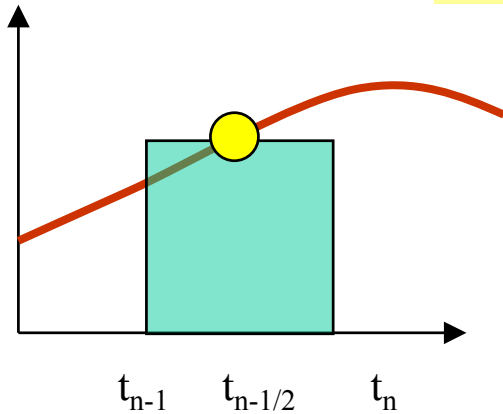


kwadratura trapezów

$u(t_n) = u(t_{n-1}) + \Delta t f(t_n)/2 + \Delta t f(t_{n-1})/2 + O(\Delta t^3)$

- rozpoznajemy niejawny schemat trapezów

reguła punktu środkowego



wzór prostokątów z wywołaniem
funkcji w środku przedziału
(dokładny dla funkcji liniowej, znoszenie błędów)

$$u_n = u_{n-1} + \Delta t f(t_{n-1} + \Delta t/2) + O(\Delta t^3)$$

uogólniony wzór na równanie różniczkowe $u' = f(t, u)$

$$u_n = u_{n-1} + \Delta t f\left(t_{n-\frac{1}{2}}, u_{n-\frac{1}{2}}\right)$$

↑

np. ze schematu Eulera:

ale - skąd rozwiązanie w środku przedziału?

$$u_{n-\frac{1}{2}} = u_{n-1} + \frac{\Delta t}{2} f(t_{n-1}, u_{n-1})$$

błąd lokalny Eulera $O(\Delta t^2)$, czy reguła punktu środkowego zachowa trzeci rząd błędu lokalnego?

sprawdźmy to rozważając bardziej ogólny schemat:

$$u_n = u_{n-1} + \Delta t (b_1 k_1 + b_2 k_2) \quad \Bigg|$$

$$k_1 = f(t_{n-1}, u_{n-1}) \quad \Bigg| \text{ obliczone na początku kroku}$$

$$k_2 = f(t_{n-1} + c\Delta t, u_{n-1} + a\Delta t k_1) \quad \Bigg|$$

obliczone gdzieś
w środku przedziału
(t_{n-1}, t_n)

z odpowiednio oszacowanym
rozwiązaniem u dla tego t (wzór typu Eulera)

jest to jawny dwustopniowy schemat **Rungego-Kutty**. potencjalna wyższa dokładność od jawnego Eulera kosztem dwóch wywołań f (podobnie jak we wzorze trapezów, ale RK: jawny)

b_1, b_2, a, c – parametry metody – jakie muszą być aby RK2 (2 = rząd dokładności)

reguła punktu środkowego:

należy do tej klasy z
 $b_1=0, b_2=1, c=1/2, a=1/2$

$$u_n = u_{n-1} + \Delta t f\left(t_{n-\frac{1}{2}}, u_{n-\frac{1}{2}}\right) \quad \Bigg|$$

$$u_{n-\frac{1}{2}} = u_{n-1} + \frac{\Delta t}{2} f(t_{n-1}, u_{n-1})$$

$$(*) \left\{ \begin{array}{l} u_n = u_{n-1} + \Delta t (b_1 k_1 + b_2 k_2) \\ k_1 = f(t_{n-1}, u_{n-1}) \\ k_2 = f(t_{n-1} + c\Delta t, u_{n-1} + a\Delta t k_1) \end{array} \right.$$

jak dobrać b_1, b_2, c, a ? – *metodą brutalnej siły - tak aby rozwinięcie Taylora metody zgadzało się z rozwinięciem Taylora dokładnego równania różniczkowego do wyrazów tak wysokiego rzędu jak to tylko możliwe*

$$u' = f(t, u)$$

przypominamy: rozwinięcie Taylora dla funkcji dwóch zmiennych

$$\begin{aligned} g(x + \Delta x, y + \Delta y) &= g(x, y) \\ &+ \left(\Delta x \frac{\partial g(x, y)}{\partial x} + \Delta y \frac{\partial g(x, y)}{\partial y} \right) \\ &+ \frac{1}{2!} \left(\Delta x^2 \frac{\partial^2 g(x, y)}{\partial x^2} + 2\Delta x \Delta y \frac{\partial^2 g(x, y)}{\partial x \partial y} + \Delta y^2 \frac{\partial^2 g(x, y)}{\partial y^2} \right) + \dots \end{aligned}$$

wstawiamy rozwiązanie dokładne $u(t_n)$, $u(t_{n-1})$ do (*) i rozwijamy względem t_{n-1} , u_{n-1}

$$u(t_n) = u(t_{n-1}) + \Delta t (b_1 k_1 + b_2 k_2)$$

$$k_1 = f(t_{n-1}, u(t_{n-1}))$$

$$k_2 = f(t_{n-1} + c\Delta t, u(t_{n-1}) + a\Delta t k_1) \leftarrow \text{to trzeba rozwinąć}$$

$$k_2 = f + c\Delta t f'_t + a\Delta t f f'_u + \frac{1}{2} (c^2 \Delta t^2 f''_{tt} + a^2 \Delta t^2 f^2 f''_{uu} + 2ac \Delta t^2 f f'_{tu}) + O(\Delta t^3)$$

wstawmy k_2 do rozwinięcia.

Zachowajmy człony do Δt^2 :

(wszystko liczone
w t_{n-1}, u_{n-1})

$$u(t_n) = u(t_{n-1}) + \Delta t [b_1 f + b_2 (f + c\Delta t f'_t + a\Delta t f f'_u)] + O(\Delta t^3)$$

rozwinięcie Taylora rozwiązania dokładnego uzyskaliśmy kilka slajdów wcześniej

$$u(t_n) = u(t_{n-1}) + \Delta t f(t_{n-1}, u_{n-1}) + \frac{\Delta t^2}{2} [f'_t + f'_u f]_{(t_{n-1}, u_{n-1})} + \frac{\Delta t^3}{6} u'''(\xi_n)$$

czyli:

rzęd Δt : $b_1 + b_2 = 1$, rzęd Δt^2 : $b_2 c = b_2 a = 1/2$

czyli reguła punktu środkowego: $b_1=0, b_2=1, c=1/2, a=1/2$ ma błąd lokalny rzędu $O(\Delta t^3)$
mamy metodę równie dokładną co wzór trapezów – ale jawną (co ma swoje zalety i wady)

Wyższy rzęd błędu do uzyskania tylko w metodach o większej niż 2 liczbie stopni

cztery parametry i trzy równania

$$b_1 + b_2 = 1$$

$$b_2 c = b_2 a = 1/2$$

- pozostaje swoboda w wyborze parametrów

reguła punktu środkowego RK2

$$b_1 = 0, b_2 = 1, c = 1/2, a = 1/2$$

dwa zastosowania jawnego schematu Eulera

$$u_{n-\frac{1}{2}} = u_{n-1} + \frac{\Delta t}{2} f(t_{n-1}, u_{n-1})$$

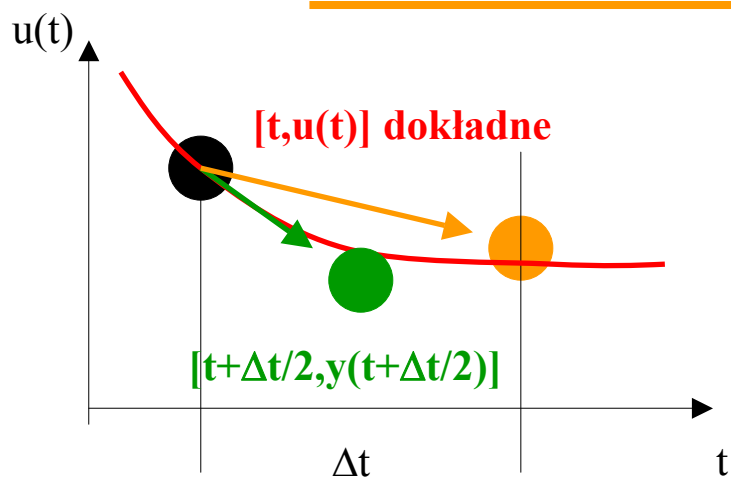
oszacowanie wstępne w punkcie pośrednim
(błąd lokalny rzędu drugiego)

$$u_n = u_{n-1} + \Delta t f(t_{n-\frac{1}{2}}, u_{n-\frac{1}{2}})$$

oszacowanie docelowe
(błąd lokalny oszacowania: rzędu trzeciego)

albo (przesunięty indeks)

$$u_{n+1} = u_n + \Delta t f\left(t_n + \frac{\Delta t}{2}, u_n + \frac{\Delta t}{2} f(t_n, u_n)\right)$$



- 1) Szacujemy metodą Eulera punkt środkowy $[t + \Delta t/2, u(t + \Delta t/2)]$ korzystając z $f(t, u)$ w lewym końcu przedziału
- 2) Wykorzystujemy wartość f w tym punkcie do wyliczenia zmiany y na całym przedziale Δt

RK punktu środkowego: $b_1 + b_2 = 1$,
 $b_2 c = b_2 a = 1/2$

inny wybór:

$b_1 = b_2 = 1/2$, wtedy musi $a = c = 1$

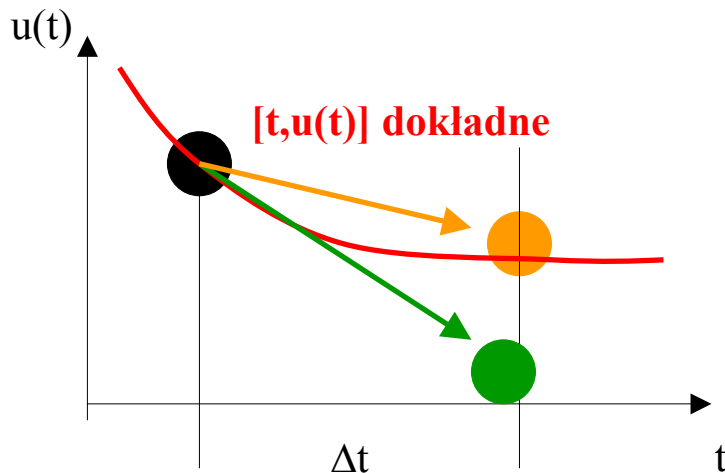
$$u_n = u_{n-1} + \Delta t (b_1 k_1 + b_2 k_2)$$

$$k_1 = f(t_{n-1}, u_{n-1})$$

$$k_2 = f(t_{n-1} + c\Delta t, u_{n-1} + a\Delta t k_1)$$

metoda podobna do wzoru trapezów (ale jawna)

$$u_{n+1} = u_n + \frac{\Delta t}{2} f(t_n, u_n) + \frac{\Delta t}{2} f(t_n + \Delta t, u_n + \Delta t f(t_n, u_n))$$



metoda RK2 trapezów

- 1) Szacujemy metodą Eulera punkt końcowy $[t + \Delta t, u(t + \Delta t)]$ korzystając z $f(t, u)$ w lewym końcu przedziału
- 2) krok z t do $t + \Delta t$ wykonujemy biorąc średnią arytmetyczną z f na początku i końcu

dla błędu lokalnego $O(\Delta t^3)$ potrzeba aby,
rząd Δt : $b_1 + b_2 = 1$, rząd Δt^2 : $b_2 c = b_2 a = 1/2$

punkt środkowy $b_2 = 1$, $b_1 = 0$ [$b_1 + b_2$]=1

czy ma sens $b_1 = 1$, $b_2 = 0$?

$$u(t_n) = u(t_{n-1}) + \Delta t (b_1 k_1 + b_2 k_2)$$

$$k_1 = f(t_{n-1}, u(t_{n-1}))$$

są to metody jednokrokowe, czyli można zapisać:

$$u_n = u_{n-1} + \Delta t \phi(t_{n-1}, u_{n-1}, \Delta t)$$

metoda RK w s-odśłonach (*stage*) (unikamy słowa „krok”)

$$u_n = u_{n-1} + \Delta t \sum_{i=1}^s b_i k_i$$

z

$$k_i = f(t_{n-1} + c_i \Delta t, u_{n-1} + \Delta t \sum_{j=1}^s a_{ij} k_j)$$

wzory przedstawiane w formie tabel Butchera

c **A**
b

c_1	a_{11}	a_{12}	\cdots	a_{1s}
c_2	a_{21}	a_{22}	\cdots	a_{2s}
\vdots	\vdots	\vdots	\ddots	\vdots
c_s	a_{s1}	a_{s2}	\cdots	a_{ss}
	b_1	b_2	\cdots	b_s

Metody Rungego-Kutty, forma ogólna

$$u_n = u_{n-1} + \Delta t \sum_{i=1}^s b_i k_i$$

$$k_i = f(t_{n-1} + c_i \Delta t, u_{n-1} + \Delta t \sum_{j=1}^s a_{ij} k_j)$$

czasem zapisywane w postaci:

$$u_n = u_{n-1} + \Delta t \sum_{i=1}^s b_i f(t_{n-1} + c_i \Delta t, U_i)$$

$$U_i = u_{n-1} + \Delta t \sum_{j=1}^s a_{ij} f(t_{n-1} + c_j \Delta t, U_j)$$

tutaj U_i – przybliżone rozwiązanie w chwili $t_{n-1} + c_i \Delta t$
zazwyczaj niższej dokładności niż rozwiązanie końcowe

jawne metody Rungego-Kutty

$$u_n = u_{n-1} + \Delta t \sum_{i=1}^s b_i k_i$$

$$k_i = f(t_{n-1} + c_i \Delta t, u_{n-1} + \Delta t \sum_{j=1}^s a_{ij} k_j)$$

jawne: $a_{ij}=0$ dla $j \geq i$

c_1	a_{11}	a_{12}	\cdots	a_{1s}
c_2	a_{21}	a_{22}	\cdots	a_{2s}
\vdots	\vdots	\vdots	\ddots	\vdots
c_s	a_{s1}	a_{s2}	\cdots	a_{ss}
	b_1	b_2	\cdots	b_s

obcięte sumowanie:

$$k_i = f(t_{n-1} + c_i \Delta t, u_{n-1} + \Delta t \sum_{j=1}^{i-1} a_{ij} k_j)$$

odsłona i -ta wyliczana na podstawie tylko wcześniejszych odsłon

historycznie wszystkie RK były jawne, uogólnienie okazało się przydatne dla problemów sztywnych

Wyprowadzanie formuł RK (a,b,c)

- 1) Rozwijamy rozwiązanie dokładne w szereg Taylora względem t_{n-1}
- 2) Podstawiamy rozwiązanie dokładne do ogólnej formy RK i rozwijamy względem t_{n-1}
- 3) Wartości parametrów a,b,c uzyskujemy z porównania. zazwyczaj w sposób niejednoznaczny

najbardziej popularne: jawne formuły 4-etapowe RK4:

o 4-tym stopniu zbieżności (4-tym rzędzie dokładności)

i 5-tym rzędzie błędu lokalnego

ogólna tabela Butchera:
dla jawnych RK4

0	0	0	0	0
c_2	a_{21}	0	0	0
c_3	a_{31}	a_{32}	0	0
c_4	a_{41}	a_{42}	a_{43}	0
	b_1	b_2	b_3	b_4

$$u_n = u_{n-1} + \Delta t \sum_{i=1}^s b_i k_i$$

$$k_i = f(t_{n-1} + c_i \Delta t, u_{n-1} + \Delta t \sum_{j=1}^{i-1} a_{ij} k_j)$$

$$c_1 = 0$$

(dla każdej jawnej RK, zaczynamy – k_1
od wyliczenia prawej strony w kroku początkowym)

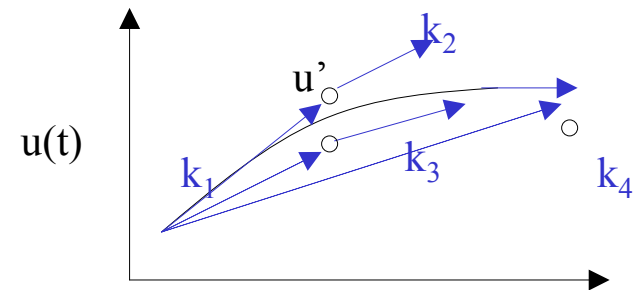
klasyczna formuła RK4: $u_n = u_{n-1} + \frac{\Delta t}{6} (k_1 + 2k_2 + 2k_3 + k_4)$

$$k_1 = f(t_{n-1}, u_{n-1})$$

$$k_2 = f\left(t_{n-1} + \frac{\Delta t}{2}, u_{n-1} + \frac{\Delta t k_1}{2}\right)$$

$$k_3 = f\left(t_{n-1} + \frac{\Delta t}{2}, u_{n-1} + \frac{\Delta t k_2}{2}\right)$$

$$k_4 = f(t_{n-1} + \Delta t, u_{n-1} + \Delta t k_3)$$



4 wywołania f na krok,
błąd lokalny $O(\Delta t^5)$

gdy f tylko funkcja czasu RK4

redukuje się do formuły Simpsona (dokładnie całkuje wielomiany trzeciego stopnia):

$$u_n = u_{n-1} + \frac{\Delta t}{6} (f(t_{n-1}) + 4f(t_{n-1/2}) + f(t_n))$$

Jawne schematy RK dla układu równań różniczkowych

$$\begin{aligned}\frac{du^1}{dt} &= f^1(t, u^1, u^2) \\ \frac{du^2}{dt} &= f^2(t, u^1, u^2)\end{aligned}$$

2 zmienne zależne u^1, u^2 ,
2 prawe strony f^1, f^2

2 równania, s-odstęp
(i=1,2,...,s)

$$\begin{aligned}U_i^1 &= u_{n-1}^1 + \Delta t \sum_{j=1}^{i-1} a_{ij} f^1(t_{n-1} + c_j \Delta t, U_j^1, U_j^2) \\ U_i^2 &= u_{n-1}^2 + \Delta t \sum_{j=1}^{i-1} a_{ij} f^2(t_{n-1} + c_j \Delta t, U_j^1, U_j^2)\end{aligned}$$

$$\begin{aligned}u_n^1 &= u_{n-1}^1 + \Delta t \sum_{i=1}^s b_i f^1(t_{n-1} + c_i \Delta t, U_i^1, U_i^2) \\ u_n^2 &= u_{n-1}^2 + \Delta t \sum_{i=1}^s b_i f^2(t_{n-1} + c_i \Delta t, U_i^1, U_i^2)\end{aligned}$$

zapis wektorowy

$\mathbf{u}_{n-1}, \mathbf{u}_n, \mathbf{f}, \mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_N$ są wektorami o 2 składowych

$$\frac{d\mathbf{u}}{dt} = \mathbf{f}(t, \mathbf{u})$$

$$\mathbf{U}_i = \mathbf{u}_{n-1} + \Delta t \sum_{j=1}^{i-1} a_{ij} \mathbf{f}(t_{n-1} + c_j \Delta t, \mathbf{U}_j)$$

$$\mathbf{u}_n = \mathbf{u}_{n-1} + \Delta t \sum_{i=1}^s b_i \mathbf{f}(t_{n-1} + c_i \Delta t, \mathbf{U}_i)$$

Tabela Butchera dla klasycznej jawnej RK4

$$u_n = u_{n-1} + \frac{\Delta t}{6} (k_1 + 2k_2 + 2k_3 + k_4)$$

$$k_1 = f(t_{n-1}, u_{n-1})$$

$$k_2 = f(t_{n-1} + \frac{\Delta t}{2}, u_{n-1} + \frac{\Delta t k_1}{2})$$

$$k_3 = f(t_{n-1} + \frac{\Delta t}{2}, u_{n-1} + \frac{\Delta t k_2}{2})$$

$$k_4 = f(t_{n-1} + \Delta t, u_{n-1} + \Delta t k_3)$$

0	0	0	0	0
1/2	1/2	0	0	0
1/2	0	1/2	0	0
1	0	0	1	0
	1/6	1/3	1/3	1/6

$$u_n = u_{n-1} + \Delta t \sum_{i=1}^s b_i k_i$$

$$k_i = f(t_{n-1} + c_i \Delta t, u_{n-1} + \Delta t \sum_{j=1}^{i-1} a_{ij} k_j)$$

0	0	0	0	0
c ₂	a ₂₁	0	0	0
c ₃	a ₃₁	a ₃₂	0	0
c ₄	a ₄₁	a ₄₂	a ₄₃	0
	b ₁	b ₂	b ₃	b ₄

Dlaczego RK4 najbardziej popularna:

Liczba kroków a rząd zbieżności jawnych metod RK:

rząd	1	2	3	4	5	6	7	8
minimalna liczba odślon	1	2	3	4	6	7	9	11

RK4 – wyjątkowo opłacalna

RK1 – metoda RK w jednej odślonie

$$u_n = u_{n-1} + \Delta t \sum_{i=1}^s b_i k_i$$

$$k_i = f(t_{n-1} + c_i \Delta t, u_{n-1} + \Delta t \sum_{j=1}^s a_{ij} k_j)$$

$b_1 + b_2 = 1$, przy $b_1 = 1, b_2 = 0$
dostaniemy jawnego Eulera
warunek $a_{12} = c_2 = 1/2$
nie będzie spełniony

jawny schemat Eulera to jawna metoda RK1

$$u_n = u_{n-1} + \Delta t \sum_{i=1}^s b_i k_i$$

jawny Euler

$$k_i = f(t_{n-1} + c_i \Delta t, u_{n-1} + \Delta t \sum_{j=1}^s a_{ij} k_j)$$

0	0
	1

tabela Butchera

RK2 trapezów

$$\mathbf{b}_1=\mathbf{b}_2=1/2, \mathbf{a}=\mathbf{c}=1$$

0	0	0
1	1	0
	$\frac{1}{2}$	$\frac{1}{2}$

$$u_n = u_{n-1} + \Delta t (b_1 k_1 + b_2 k_2)$$

$$k_1 = f(t_{n-1}, u_{n-1})$$

$$k_2 = f(t_{n-1} + c\Delta t, u_{n-1} + a\Delta t k_1)$$

RK2 punktu środkowego

$$u_{n+1} = u_n + \Delta t f\left(t_n + \frac{\Delta t}{2}, u_n + \frac{\Delta t}{2} f(t_n, u_n)\right)$$

0	0	0
1/2	1/2	0
	0	1