

Podstawy matematyki stosowanej czyli wstęp do równań różniczkowych

część 2: Metody numeryczne w równaniach różniczkowych

dr inż. Łukasz Błaszczyk

Wydział Matematyki i Nauk Informatycznych
Politechnika Warszawska

rok akademicki 2019/2020 (semestr letni)

Co już wiemy?

Do tej pory omówiliśmy:

- metody rozwiązywania niektórych typów równań,
- metody badania istnienia i jednoznaczności,
- zachowanie się rozwiązań w zależności od parametrów.

Po co nam są metody numeryczne?

Większości równań zwyczajnych **nie da się** rozwiązać analitycznie, tzn. nie da się zapisać rozwiązania za pomocą znanych funkcji.

Motywacja podjęcia się tematu

Przykład 1.

Równanie

$$x'(t) = \sin(t) - x(t), \quad t \in \mathbb{R},$$

ma rozwiązanie ogólne

$$x(t) = Ae^{-t} + \frac{1}{2} \sin(t) - \frac{1}{2} \cos(t), \quad A \in \mathbb{R}.$$

Motywacja podjęcia się tematu

Przykład 2.

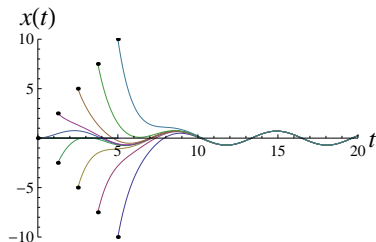
Z drugiej strony (bardzo podobne) równanie

$$x'(t) = \sin(t) - \frac{1}{10}x^3(t), \quad t \in \mathbb{R},$$

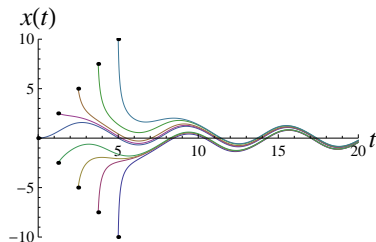
ma rozwiązania, których nie da się zapisać za pomocą znanych formuł.

Co więcej: rozwiązania obu równań dla tych samych warunków początkowych zachowują się bardzo podobnie.

Motywacja podjęcia się tematu



$$x'(t) = \sin(t) - x(t)$$



$$x'(t) = \sin(t) - \frac{1}{10}x^3(t)$$

Czym się zajmiemy?

Rozważamy równania różniczkowe zwyczajne **pierwszego rzędu**:

$$\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x}(t)), \quad t > t_0,$$

gdzie \mathbf{x} może być funkcją wektorową lub skalarną.

Uwaga (przypomnienie).

Ponieważ każde równanie różniczkowe zwyczajne da się sprowadzić do układu równań różniczkowych zwyczajnych pierwszego rzędu, ograniczymy się tylko do takich równań.

Czym się zajmiemy?

Technikami omawianymi na tym przedmiocie **nie da się** przybliżyć rozwiązań ogólnych. Do równania różniczkowego zawsze będziemy dodawać **warunek początkowy**:

$$\mathbf{x}(t_0) = \mathbf{x}_0.$$

Zakładamy ponadto, że rozwiązania szukamy w **skończonym** przedziale czasu:

$$t_0 \leq t \leq t_K, \quad t_0 < t_K < +\infty$$

(nawet jeśli rozwiązanie istnieje dla wszystkich $t > t_0$).

Omawiane metody

Zajmiemy się dwiema klasami metod:

- *liniowymi metodami wielokrokowymi,*
- *metodami Runge-Kutty.*

Każda z tych metod jest uogólnieniem *metody Eulera*.

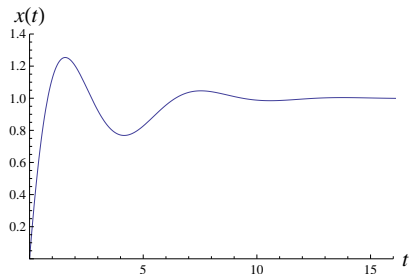
Uwaga.

Nie przybliżamy krzywej $t \mapsto \mathbf{x}(t)$, ale (zazwyczaj) ciąg wartości

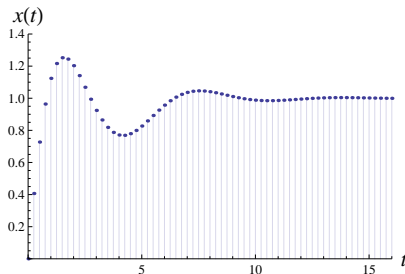
$$\mathbf{x}(t_0), \mathbf{x}(t_0 + h), \mathbf{x}(t_0 + 2h), \dots, \mathbf{x}(t_0 + nh), \dots,$$

gdzie $h > 0$ jest tzw. **wielkością siatki** lub **długością kroku**.

Omawiane metody



NIE



TAK

Omawiane metody

Uwaga (dlaczego zazwyczaj?) #1.

W ogólności można rozważać siatki nierównomierne, tzn.

$$t_0 < t_1 < t_2 < \dots < t_n < \dots,$$

a wówczas długość kroku będzie się zmieniać w czasie:

$$h_n := t_n - t_{n-1}, \quad n = 1, 2, \dots$$

Będziemy zajmować się głównie metodami **stałokrokowymi**, dla których $h_n =: h = \text{const}$.

Omawiane metody

Uwaga (dlaczego zazwyczaj?) #2.

Niektóre twierdzenia (o oszacowaniach błędu przybliżenia) będą wymagały stworzenie pewnych krzywych, które przechodzą przez punkty naszego przybliżenia.

W najprostszym przypadku będą to po prostu łamane łączące punkty przybliżenia.

Omawiane metody

Pytanie.

Wydaje się (i słusznie), że metody stałokrokowe łatwiej będzie zaimplementować.

Do czego w takim razie mogą przydać się metody ze zmienną długością kroku?

Konkretne przykłady na laboratorium.

Notacja – przypomnienie z Algebry liniowej

Przypomnijmy **normy** wektorowe i macierzowe :

dla wektorów $\mathbf{x} = (x_1, \dots, x_N)^T \in \mathbb{R}^N$ definiujemy:

$$\|\mathbf{x}\|_p = \begin{cases} \left(\sum_{i=1}^N |x_i|^p \right)^{1/p}, & \text{dla } p \in [1, \infty), \\ \max_{i=1, \dots, N} |x_i|, & \text{dla } p = \infty \end{cases}$$

(dla dowolnego $p \in [1, \infty]$ jest to norma, nazywana też **normą** ℓ_p).

Większość przedstawionych wyników jest prawdziwa dla dowolnej normy ℓ_p , dlatego będziemy wówczas pisać po prostu $\|\cdot\|$.

Notacja – przypomnienie z Algebry liniowej

Przypomnijmy **normy** wektorowe i macierzowe :

dla macierzy $\mathbf{A} \in \mathbb{R}^{M \times N}$ definiujemy:

$$\|\mathbf{A}\|_{p \rightarrow q} = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{Ax}\|_q}{\|\mathbf{x}\|_p} = \sup_{\|\mathbf{x}\|_p=1} \|\mathbf{Ax}\|_q$$

(jest to najmniejsza liczba taka, że $\|\mathbf{Ax}\|_q \leq \|\mathbf{A}\|_{p \rightarrow q} \cdot \|\mathbf{x}\|_p$).

Tak jak wcześniej, jeśli wyniki będą prawdziwe dla dowolnych p i q , to będziemy pisać po prostu $\|\mathbf{A}\|$.

Dla $p = q = 2$ mamy $\|\mathbf{A}\| = \sqrt{\text{największa wartość własna } \mathbf{A}^T \mathbf{A}}$.

Notacja – przypomnienie z Analizy 1

Przypomnijmy tzw. **notację Landaua** (czyli *notację dużego \mathcal{O}*): mówimy, że $z = \mathcal{O}(h^p)$ (jest co najwyżej rzędu $\mathcal{O}(h^p)$) dla pewnego $p \in \mathbb{Z}_+$, jeśli istnieją dodatnie stałe h_0 i C takie, że

$$|z| \leq Ch^p \quad \text{dla} \quad 0 < h < h_0$$

(w przypadku wielowymiarowym zamiast $|\cdot|$ należy wpisać $\|\cdot\|$).

Wynika stąd, że $z \rightarrow 0$, gdy $h \rightarrow 0$ i **rzęd zbieżności** jest równy p . Wprowadzamy tę notację, jeśli bardziej interesuje nas informacja o p , a nie o C . Będziemy czasem mówić, że z jest p -tego rzędu.

Wzór Taylora – przypomnienie z Analizy 1

Notacja dużego \mathcal{O} pojawiła się w kontekście *wzoru Taylora*.

Przykład 3 (wniosek ze wzoru Taylora).

Stosując wzór Maclaurina do funkcji $f(h) = e^h$ otrzymamy

$$\begin{aligned}e^h &= 1 + \mathcal{O}(h) \\ &= 1 + h + \mathcal{O}(h^2) \\ &= 1 + h + \frac{1}{2!}h^2 + \mathcal{O}(h^3) \\ &= \dots\end{aligned}$$

Założenia na resztę wykładu

Notację dużego \mathcal{O} stosujemy, żeby porównywać tempo zbieżności – wyrazy rzędu $\mathcal{O}(h^3)$ zbiegają szybciej do zera niż rzędu $\mathcal{O}(h^2)$.

Założenie.

Rozwiązanie RRZ (czyli $t \mapsto \mathbf{x}(t)$) jest odpowiednio gładkie, tzn. ma tak wiele ciągłych pochodnych na (t_0, t_K) jak potrzebujemy.

Dzięki temu będziemy mogli skorzystać ze wzoru Taylora tak wysokiego rzędu jak chcemy.

Wzór Taylora (znowu)

Rozpoczniemy od przypadku skalarnego i rozwinięcia rozwiązania x wyjściowego równania ze wzoru Taylora wokół punktu t :

$$x(t+h) = x(t) + h \cdot x'(t) + R_1(h;t),$$

gdzie resztę R_1 będziemy nazywać **lokalnym błędem obcięcia** (ang. *local truncation error* – LTE).

Jeśli x jest 2-krotnie różniczkowalna na (t_0, t_K) , to możemy zapisać

$$R_1(h;t) = \frac{1}{2!} h^2 x''(\xi), \quad \xi \in (t, t+h).$$

Wzór Taylora (znowu)

Założmy, że $\exists M > 0 \forall \xi \in (t_0, t_K) |x''(\xi)| \leq M$, wówczas:

$$|R_1(h; t)| \leq \frac{1}{2} M h^2,$$

czyli $R_1(h; t) = \mathcal{O}(h^2)$.

Inne uzasadnienie:

Wiemy z Tw. o wzorze Taylora (z resztą w postaci Peano), że jeśli x ma pochodne $x', \dots, x^{(m-1)}$ wokół t oraz istnieje $x^{(m)}(t)$, to

$$x(t+h) = \sum_{k=1}^m \frac{x^{(k)}(t)}{k!} h^k + o(h^m),$$

a stąd automatycznie (biorąc $m = 1$) otrzymujemy to, co wcześniej.

Wprowadzenie metody Eulera

Mamy zatem

$$x(t+h) = x(t) + hx'(t) + \mathcal{O}(h^2),$$

co po wstawieniu równania $x'(t) = f(t, x(t))$ daje

$$x(t+h) = x(t) + hf(t, x(t)) + \mathcal{O}(h^2).$$

Wprowadźmy punkty siatki $t = t_n$, gdzie $t_n = t_0 + nh$, $n = 1, \dots, N$, a $N = \lfloor \frac{t_K - t_0}{h} \rfloor$ jest liczbą kroków o długości h potrzebnych by dotrzeć (ale nie przekroczyć) do czasu $t = t_K$.

Wyprowadzenie metody Eulera

Stąd dla $t = t_n$ ($n < N$) mamy

$$x(t_{n+1}) = x(t_n) + hf(t_n, x(t_n)) + \mathcal{O}(h^2), \quad n = 0, \dots, N - 1,$$

oraz $x(t_0) = x_0$.

Dla odpowiednio małego h ostatni wyraz może być dowolnie mały, a pomijając go otrzymujemy **metodę Eulera**:

$$x_{n+1} = x_n + hf(t_n, x_n), \quad n = 0, \dots, N - 1,$$

gdzie x_n jest przybliżeniem numerycznym wartości $x(t_n)$.

Interpretacja geometryczna

Metoda Eulera (w przypadku wielowymiarowym):

$$\mathbf{x}_{n+1} = \mathbf{x}_n + h\mathbf{f}(t_n, \mathbf{x}_n), \quad n = 0, \dots, N - 1,$$

Co opisuje (dla ustalonego punktu (t_n, \mathbf{x}_n)) powyższe równanie?

Mając dany punkt (t_n, \mathbf{x}_n) prowadzimy z niego odcinek *styczny do rozwiązania* przechodzącego przez ten punkt:

$$\tau \mapsto (t_n + \tau, \mathbf{x}_n + \tau\mathbf{f}(t_n, \mathbf{x}_n)), \quad \tau \in (0, h).$$

Interpretacja geometryczna

Przykład 4.

Rozważamy równanie

$$x'(t) = (1 - 2t)x(t), \quad t > 0,$$

z warunkiem $x(0) = 1$, a rozwiązania szukamy dla $0 \leq t \leq 3$.

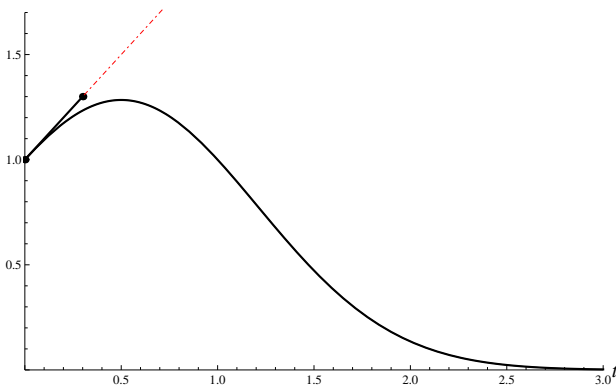
Zagadnienie to ma znane rozwiązanie:

$$x(t) = \exp\left(\frac{1}{4} - \left(\frac{1}{2} - t\right)^2\right),$$

co pozwala nam ocenić poziom dokładności przybliżenia.

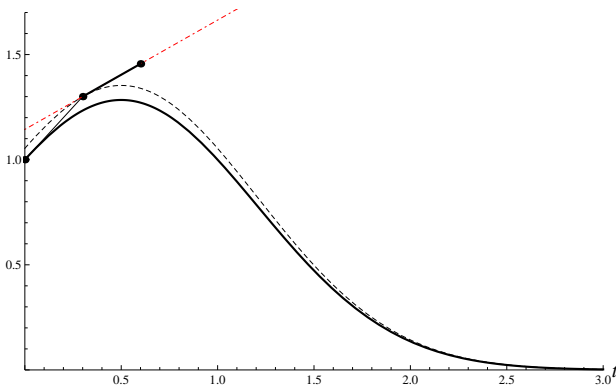
Interpretacja geometryczna

Numeryczne przybliżenie rozwiązania dla $h = 0,3$:

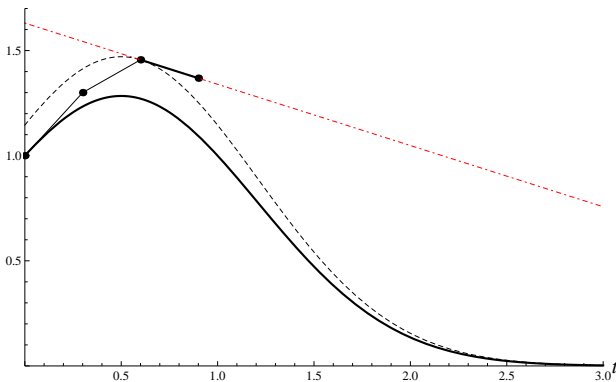


Interpretacja geometryczna

Numeryczne przybliżenie rozwiązania dla $h = 0,3$:

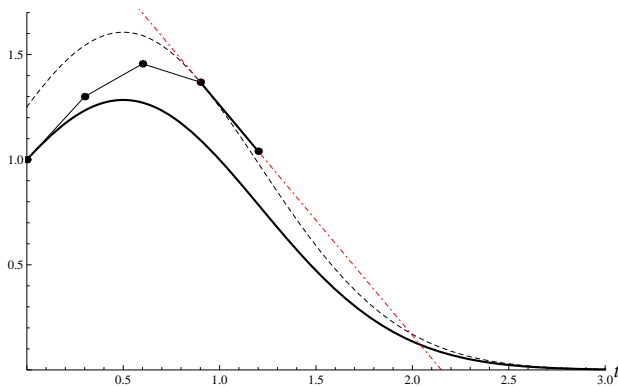


Interpretacja geometryczna

Numeryczne przybliżenie rozwiązania dla $h = 0,3$:

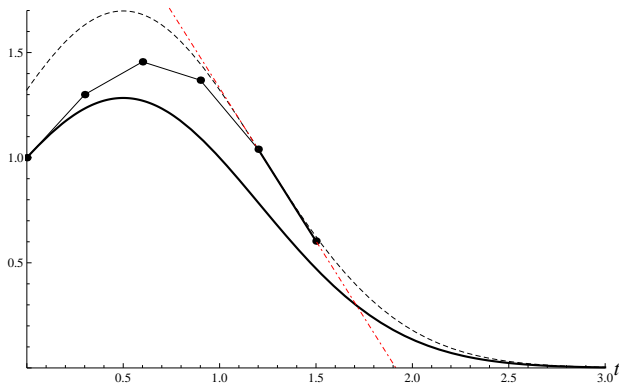
Interpretacja geometryczna

Numeryczne przybliżenie rozwiązania dla $h = 0,3$:



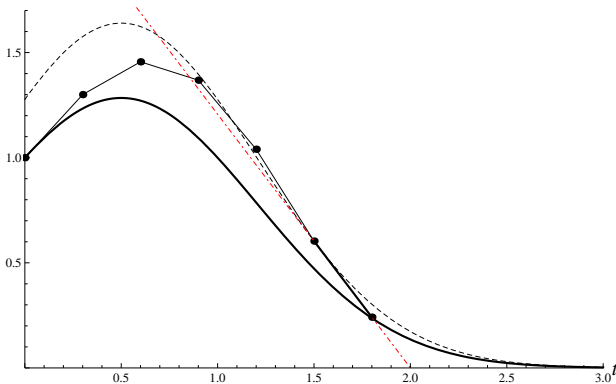
Interpretacja geometryczna

Numeryczne przybliżenie rozwiązania dla $h = 0,3$:



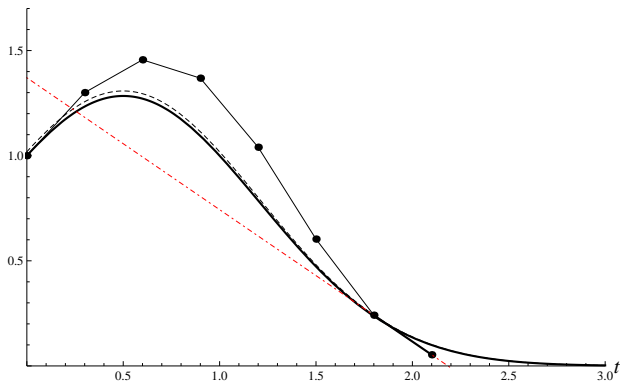
Interpretacja geometryczna

Numeryczne przybliżenie rozwiązania dla $h = 0,3$:



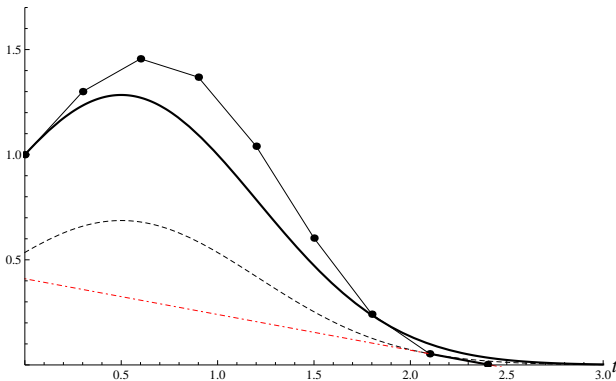
Interpretacja geometryczna

Numeryczne przybliżenie rozwiązania dla $h = 0,3$:



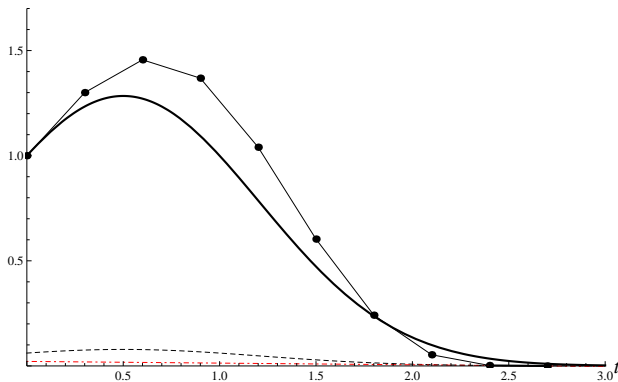
Interpretacja geometryczna

Numeryczne przybliżenie rozwiązania dla $h = 0,3$:



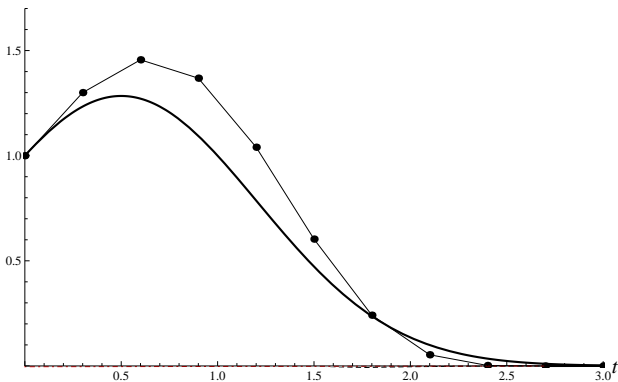
Interpretacja geometryczna

Numeryczne przybliżenie rozwiązania dla $h = 0,3$:



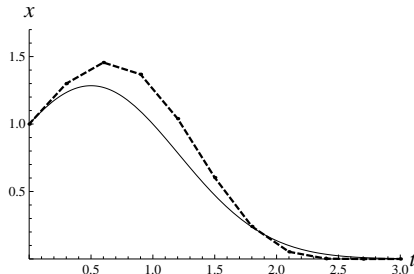
Interpretacja geometryczna

Numeryczne przybliżenie rozwiązania dla $h = 0,3$:



Zbieżność?

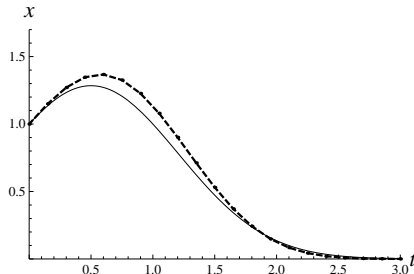
Otrzymane przybliżenie jest dość zgrubne.



Dla przypomnienia: $h = 0,3$.

Zbieżność?

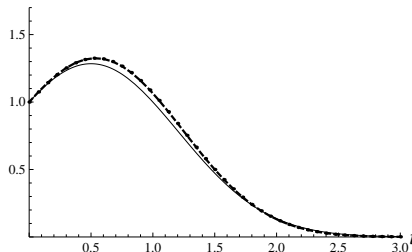
Co się stanie, jeśli zmniejszymy długość kroku siatki?



Teraz: $h = 0,15$.

Zbieżność?

Co się stanie, jeśli zmniejszymy długość kroku siatki?



Teraz: $h = 0,075$.

Błąd globalny rozwiązania i rząd zbieżności

Warto zwrócić uwagę jak zmienia się błąd rozwiązania (**błąd globalny**: $e_n = |x(t_n) - x_n|$) w pewnych charakterystycznych punktach dla wszystkich tych przybliżeń.

h	$x_n (t_n = 0,9)$	błąd glob.	$x_n (t_n = 1,5)$	błąd glob.
0,3	$x_3 \approx 1,367$	$e_3 \approx 0,274$	$x_5 \approx 0,603$	$e_5 \approx 0,131$
0,15	$x_6 \approx 1,227$	$e_6 \approx 0,133$	$x_{10} \approx 0,531$	$e_{10} \approx 0,058$
0,075	$x_{12} \approx 1,159$	$e_{12} \approx 0,065$	$x_{20} \approx 0,500$	$e_{20} \approx 0,028$
	$x(0,9) \approx 1,094$	(dokładne)	$x(1,5) \approx 0,472$	(dokładne)

Czy daje się zauważyć jakąś zależność?

Błąd globalny rozwiązania i rząd zbieżności

Wyprowadzając metodę Eulera otrzymaliśmy:

$$x(t_{n+1}) = x(t_n) + hf(t_n, x(t_n)) + \mathcal{O}(h^2), \quad n = 0, \dots, N - 1$$

(w przypadku skalarnym).

Zmniejszając krok siatki *dwukrotnie* błąd globalny również zmniejsza się (co do modułu) *dwukrotnie* (około). Sugeruje to, że błąd globalny tej metody jest proporcjonalny do h (a sama metoda jest zbieżna).

Spróbujemy to sformalizować.

Sformalizowanie dotychczasowych rozważań

Definicja 1.

Rozw. numeryczne \mathbf{x}_n **zbiega** do rozwiązania \mathbf{x} zag. początkowego

$$\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x}(t)), \quad t > 0, \quad \mathbf{x}(0) = \mathbf{x}_0,$$

w punkcie $t = t^*$, jeśli $\mathbf{e}_n = \mathbf{x}(t_n) - \mathbf{x}_n$ w punkcie $t_n = t^*$ spełnia

$$\|\mathbf{e}_n\| \xrightarrow{h \rightarrow 0} 0$$

(czyli dla $n \rightarrow \infty$ i $nh = t_n = t^*$).

Zbieżność jest rzędu p , jeśli $\mathbf{e}_n = \mathcal{O}(h^p)$ dla pewnego $p \in \mathbb{Z}_+$ (przyjmujemy, że rzędem metody jest największe takie p).

Sformalizowanie dotychczasowych rozważań

Mówimy, że metoda rozwiązywania jest zbieżna (rzędu p), jeśli dla dowolnego zagadnienia

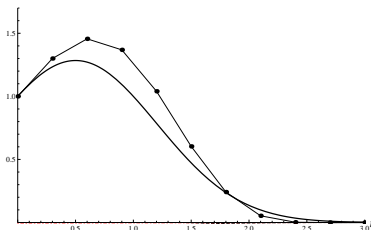
$$\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x}(t)), \quad t > 0, \quad \mathbf{x}(0) = \mathbf{x}_0,$$

(spełniającego właściwe założenia) rozwiązanie numeryczne otrzymane tą metodą zbiega (z rzędem p) do rozwiązania zagadnienia.

Rozwiązanie otrzymane metodą Eulera

Zauważmy, że metoda Eulera daje nam tzw. **łamaną Eulera**:

$$\mathbf{x}_h(t) = \mathbf{x}_n + (t - t_n) \cdot \mathbf{f}(t_n, \mathbf{x}_n) \quad \text{dla } t_n \leq t \leq t_{n+1},$$
$$n = 0, \dots, N - 1.$$



Zbieżność metody Eulera

Rozważamy zagadnienie

$$\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x}(t)), \quad t > t_0, \quad \mathbf{x}(t_0) = \mathbf{x}_0.$$

Oznaczenia.

$$\mathbf{D}_{\mathbf{x}}\mathbf{f} = \begin{pmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix},$$
$$\mathbf{f}_t = \frac{\partial \mathbf{f}}{\partial t} = \begin{pmatrix} \frac{\partial f_1}{\partial t} & \cdots & \frac{\partial f_n}{\partial t} \end{pmatrix}^T.$$

Zbieżność metody Eulera

Twierdzenie 1.

Przypuśćmy, że w pewnym otoczeniu rozwiązania \mathbf{x} wyjściowego zagadnienia zachodzi

$$\|\mathbf{f}\| \leq A, \quad \|\mathbf{D}_x \mathbf{f}\| \leq L, \quad \|\mathbf{f}_t\| \leq M,$$

a także niech \mathbf{x}_h będzie łamaną Eulera (uzyskaną z krokiem h).
Wówczas dla odpowiednio małych $h > 0$ mamy

$$\|\mathbf{x}(t) - \mathbf{x}_h(t)\| \leq \frac{M + AL}{L} \left(e^{L(t-t_0)} - 1 \right) \cdot h.$$

Zbieżność metody Eulera

Uwagi.

- Dość mocne założenia na funkcję f – musi być różniczkowalna.
- Zakładamy, że rozwiązanie x istnieje (i jest jednoznaczne)...
- ... ale przy tych założeniach na f to właściwie nie jest założenie, tylko wniosek (np. z Tw. Picarda-Lindelöfa).
- Twierdzenie mówi, że łamana Eulera zbiega jednostajnie do rozwiązania, a sama metoda Eulera jest zbieżna (rzędu 1).
- Jest to dość słaby rezultat (sformułujemy *lepsze* twierdzenie).

Dlaczego metoda Eulera jest taka zła?

Przykład 5.

Rozważamy równanie

$$x'(t) - x(t) = -\frac{1}{2}e^{t/2} \sin(5t) + 5e^{t/2} \cos(5t), \quad t > 0,$$

z warunkiem $x(0) = 0$, a rozwiązania szukamy dla $0 \leq t \leq 5$.

Zagadnienie to ma znane rozwiązanie:

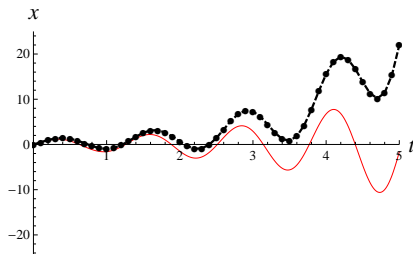
$$x(t) = e^{t/2} \sin(5t).$$

Przybliżenia szukamy metodą Eulera z krokiem $h = 0,1$ (a następnie $h = 0,05$ i $h = 0,01$).

Istnieją lepsze metody

Dlaczego metoda Eulera jest taka zła?

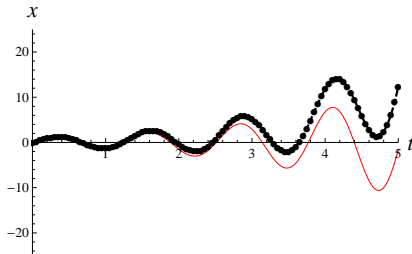
Otrzymane przybliżenie jest zupełnie bezsensowne.



Na początek: $h = 0,1$.

Dlaczego metoda Eulera jest taka zła?

Zmniejszenie kroku siatki nie daje zbyt wiele.

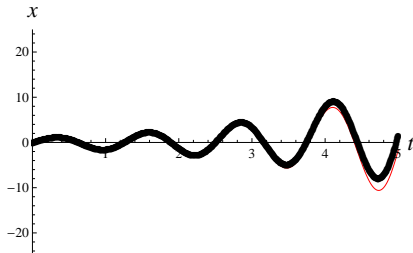


Teraz: $h = 0,05$.

Istnieją lepsze metody

Dlaczego metoda Eulera jest taka zła?

Zmniejszenie kroku siatki nie daje zbyt wiele.



Teraz: $h = 0,01$.

Dlaczego metoda Eulera jest taka zła?

Uwagi.

- Metoda Eulera daje sensowne wyniki, gdy rozwiązania są wolno zmienne.
- Gdy rozwiązania szybko zmieniają się w czasie, błąd bezwzględny staje się rosnącą funkcją czasu.
- Do tej pory badaliśmy jedynie przykłady, w których znaleźliśmy rozwiązanie analityczne – zazwyczaj jednak nie mamy takich informacji.

Wykorzystajmy *historię*

Metoda Eulera wykorzystuje informacje o zaledwie jednym wcześniejszym kroku, zapominając o wcześniejszych przybliżeniach.

Wprowadzimy **linowe metody wielokrokowe**, które pozwolą osiągnąć wyższy rząd zbieżności, wykorzystując dostępną *historię*, tzn. wartości x i x' przybliżone we wcześniejszych krokach.

Zacniemy od prostego przykładu: **2-krokowej metody Adamsa-Bashfortha**.

Wzór Taylora (po raz kolejny)

Zacniemy ponownie od przypadku skalarnego i ponownie od rozwiązania równania x wyjściowego równania ze wzoru Taylora wokół punktu t :

$$x(t+h) = x(t) + h \cdot x'(t) + \frac{1}{2}h^2 \cdot x''(t) + \mathcal{O}(h^3).$$

Pojawia się wyraz z drugą pochodną. Możemy go wyeliminować:

$$\begin{aligned}x'(t-h) &= x'(t) - h \cdot x''(t) + \mathcal{O}(h^2) \\ \Rightarrow h \cdot x''(t) &= x'(t) - x'(t-h) + \mathcal{O}(h^2).\end{aligned}$$

Wzór Taylora (po raz kolejny)

Stąd (pamiętamy, że $h \cdot x''(t) = x'(t) - x'(t-h) + \mathcal{O}(h^2)$)

$$\begin{aligned}x(t+h) &= x(t) + h \cdot x'(t) + \frac{1}{2}h \cdot h \cdot x''(t) + \mathcal{O}(h^3) \\&= x(t) + h \cdot x'(t) + \frac{1}{2}h \cdot (x'(t) - x'(t-h) + \mathcal{O}(h^2)) + \mathcal{O}(h^3) \\&= x(t) + \frac{1}{2}h \cdot (3 \cdot x'(t) - x'(t-h)) + \mathcal{O}(h^3) \\&= x(t) + \frac{1}{2}h \cdot (3 \cdot f(t, x(t)) - f(t-h, x(t-h))) + \mathcal{O}(h^3)\end{aligned}$$

Położmy $t = t_n$. Pomijamy resztę, stosujemy notację $f_n = f(t_n, x_n)$ (gdzie x_n to przybliżenie $x(t_n)$) i otrzymujemy nową metodę.

2-krokowa metoda Adamsa-Bashfortha

Otrzymaliśmy 2-krokową metodę Adamsa-Bashfortha (AB2):

$$x_{n+2} = x_{n+1} + \frac{h}{2} (3f_{n+1} - f_n), \quad n = 0, \dots, N - 2.$$

Uwaga.

- Potrzebujemy znać wartości przybliżeń w **dwóch** poprzednich krokach.
- Wartość x_1 znajdujemy metodą niższego rzędu (np. metodą Eulera albo metodą Runge-Kutty).

Liniowe metody wielokrokowe

W ogólności **liniowe metody wielokrokowe** zapisujemy w postaci

$$\begin{aligned}x_{n+k} + \alpha_{k-1}x_{n+k-1} + \dots + \alpha_0x_n \\ = h \cdot (\beta_k f_{n+k} + \beta_{k-1}f_{n+k-1} + \dots + \beta_0f_n).\end{aligned}$$

Metody jawne i niejawne.

Jeśli $\beta_k \neq 0$, to do obliczenia przybliżenia x_{n+k} trzeba znać $f_{n+k} = f(t_{n+k}, \mathbf{x}_{n+k})$ – trzeba rozwiązań (często nieliniowe) równanie. Takie metody nazywamy **niejawnymi** (ang. *implicit*).

Jeśli $\beta_k = 0$, to metoda jest **jawna** (ang. *explicit*) – np. metoda Eulera, czy metoda AB2.

Jawne metody Adamsa

2-krokowa metoda Adamsa-Bashfortha jest szczególnym przypadkiem **jawnych metod Adamsa**, wyprowadzanych z postaci całkowej

$$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} f(t, x(t)) dt,$$

gdzie nieznane rozwiązanie $x(t)$ na przedziale $[t_n, t_{n+1}]$ przybliża się stosując **wzór interpolacyjny Newtona***:

$$\tilde{f}(t) = \tilde{f}(t_n + sh) = \sum_{j=0}^{k-1} (-1)^j \binom{-s}{j} \nabla^j f_n,$$

gdzie $\nabla^0 f_n = f_n$, $\nabla^{j+1} f_n = \nabla^j f_n - \nabla^j f_{n-1}$.

* D. Kincaid, W. Cheney, *Analiza numeryczna*, WNT 2006.

Wzór interpolacyjny Newtona

Przykład 4 (jeszcze raz).

Wróćmy do przykładu z równaniem

$$x'(t) = (1 - 2t)x(t), \quad t > 0,$$

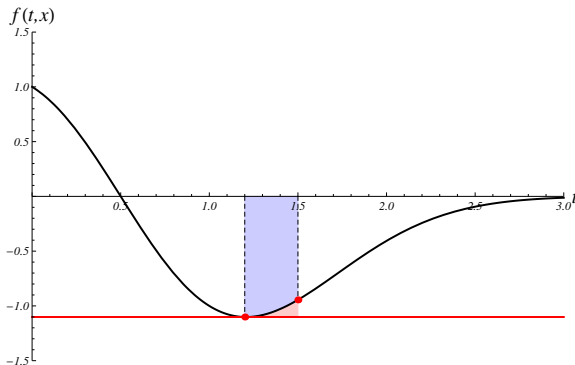
z warunkiem $x(0) = 1$.

Jak będą wyglądały przybliżenia funkcji $f(t, x) = (1 - 2t)x(t)^*$?

* **Uwaga.** Znamy w tym przypadku dokładną postać funkcji x .

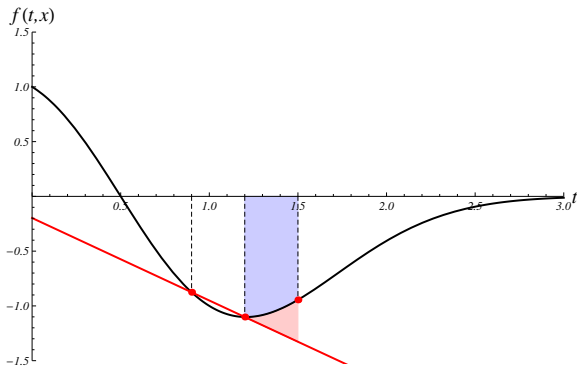
Wzór interpolacyjny Newtona

Szukamy przybliżenia w punkcie $t_5 = 5h$ ($h = 0,3$) na podstawie wartości w punkcie t_4 ($k = 1$, metoda Eulera):



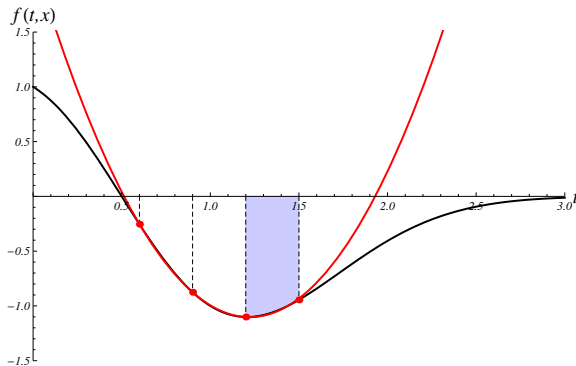
Wzór interpolacyjny Newtona

Szukamy przybliżenia w punkcie $t_5 = 5h$ ($h = 0,3$) na podstawie wartości w punktach t_4 i t_3 ($k = 2$, metoda Adamsa-Bashfortha):



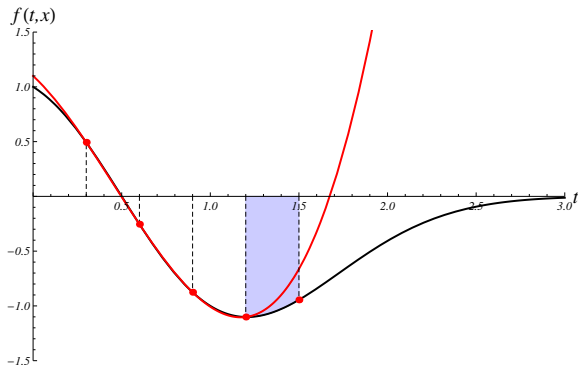
Wzór interpolacyjny Newtona

Szukamy przybliżenia w punkcie $t_5 = 5h$ ($h = 0,3$) na podstawie wartości w punktach t_4 , t_3 i t_2 ($k = 3$):



Wzór interpolacyjny Newtona

Szukamy przybliżenia w punkcie $t_5 = 5h$ ($h = 0,3$) na podstawie wartości w punktach t_4, t_3, t_2 i t_1 ($k = 4$):



Jawne metody Adamsa

Specjalne przypadki jawnych metod Adamsa wyrażają się wzorami:

$$k = 1: \quad x_{n+1} = x_n + hf_n \quad (\text{metoda Eulera})$$

$$k = 2: \quad x_{n+2} = x_{n+1} + h \left(\frac{3}{2} f_{n+1} - \frac{1}{2} f_n \right) \quad (\text{metoda AB2})$$

$$k = 3: \quad x_{n+3} = x_{n+2} + h \left(\frac{23}{12} f_{n+2} - \frac{16}{12} f_{n+1} + \frac{5}{12} f_n \right)$$

$$k = 4: \quad x_{n+4} = x_{n+3} + h \left(\frac{55}{24} f_{n+3} - \frac{59}{24} f_{n+2} + \frac{37}{24} f_{n+1} - \frac{9}{24} f_n \right)$$

Inny rodzaj (jawnych) metod wielokrokowych: Nyströma.

Co z metodami niejawnymi?

Wprowadzenie do formalizmu

Liniowa metoda wielokrokowa:

$$\begin{aligned} \alpha_k x_{n+k} + \alpha_{k-1} x_{n+k-1} + \dots + \alpha_0 x_n \\ = h \cdot (\beta_k f_{n+k} + \beta_{k-1} f_{n+k-1} + \dots + \beta_0 f_n). \quad (\star) \end{aligned}$$

Definicja 2.

Liniowym operatorem różnicowym \mathcal{L}_h związanym z liniową metodą k -krokową (\star) nazywamy operator przypisujący dowolnej funkcji z klasy \mathcal{C}^1 funkcję

$$\mathcal{L}_h z(t) = \sum_{j=0}^k (\alpha_j z(t + jh) - h\beta_j z'(t + jh)).$$

Wprowadzenie do formalizmu

Liniowa metoda wielokrokowa:

$$\begin{aligned} \alpha_k x_{n+k} + \alpha_{k-1} x_{n+k-1} + \dots + \alpha_0 x_n \\ = h \cdot (\beta_k f_{n+k} + \beta_{k-1} f_{n+k-1} + \dots + \beta_0 f_n). \quad (*) \end{aligned}$$

Definicja 3.

Liniowy operator różnicowy \mathcal{L}_h jest **zgodny rzędu p** , jeśli

$$\mathcal{L}_h z(t) = \mathcal{O}(h^{p+1})$$

dla $p \in \mathbb{Z}_+$ i dla dowolnej funkcji odpowiednio gładkiej z . Metodę nazywamy **zgodną**, jeśli jest zgodna rzędu 1.

Wprowadzenie do formalizmu

Liniowa metoda wielokrokowa:

$$\begin{aligned}\alpha_k x_{n+k} + \alpha_{k-1} x_{n+k-1} + \dots + \alpha_0 x_n \\ = h \cdot (\beta_k f_{n+k} + \beta_{k-1} f_{n+k-1} + \dots + \beta_0 f_n). \quad (*)\end{aligned}$$

Definicja 4.

Pierwszym oraz **drugim wielomianem charakterystycznym** metody k -krokowej $(*)$ nazywamy odpowiednio

$$\begin{aligned}\rho(r) &= \alpha_k r^k + \alpha_{k-1} r^{k-1} + \dots + \alpha_0, \\ \sigma(r) &= \beta_k r^k + \beta_{k-1} r^{k-1} + \dots + \beta_0.\end{aligned}$$

Zgodność liniowych metod wielokrokowych

Twierdzenie 2.

Liniowa metoda wielokrokowa (\star) jest zgodna rzędu p wtedy i tylko wtedy, gdy

$$\sum_{j=0}^k \alpha_j = 0 \quad \text{oraz} \quad \sum_{j=0}^k \alpha_j j^q = q \sum_{j=0}^k \beta_j j^{q-1} \quad \text{dla } q = 1, \dots, p.$$

Wniosek.

Metoda jest zgodna wtedy i tylko wtedy, gdy

$$\rho(1) = 0 \quad \text{oraz} \quad \rho'(1) = \sigma(1).$$

Zbieżność liniowych metod wielokrokowych

Aby móc mówić o zbieżności rozważanych metod, musimy założyć, że zagadnienie

$$\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0,$$

posiada jednoznaczne rozwiązanie.

Będziemy więc zakładać, że spełnione są założenia Twierdzenia Picarda-Lindelöfa.

Będziemy oznaczać przez \mathbf{x}_h funkcję określoną na tym samym przedziale, co \mathbf{x} i taką, że $\mathbf{x}_h(t_n) = \mathbf{x}_n$ we wszystkich punktach siatki (np. łamana łącząca punkty).

Zbieżność liniowych metod wielokrokowych

Definicja 5a.

Liniowa metoda k -krokowa jest **zbieżna**, jeśli dla dowolnych warunków początkowych spełniających wymienione wcześniej założenia zachodzi

$$\mathbf{x}(t) - \mathbf{x}_h(t) \xrightarrow{h \rightarrow 0^+} \mathbf{0} \quad \text{dla wszystkich } t \in [0, T],$$

o ile wartości startowe metody spełniały

$$\mathbf{x}(t_n) - \mathbf{x}_h(t_n) \xrightarrow{h \rightarrow 0^+} \mathbf{0} \quad \text{dla } n = 0, 1, \dots, k - 1.$$

Zbieżność liniowych metod wielokrokowych

Definicja 5b.

Liniowa metoda k -krokowa jest **zbieżna rzędu p** , jeśli dla dowolnych warunków początkowych i dla dowolnej funkcji \mathbf{f} odpowiednio gładkiej istnieją stałe $C, h > 0$ takie, że

$$\|\mathbf{x}(t) - \mathbf{x}_h(t)\| \leq Ch^p \quad \text{dla } h \leq h_0,$$

o ile wartości startowe metody spełniały

$$\|\mathbf{x}(t_n) - \mathbf{x}_h(t_n)\| \leq Ch^p \quad \text{dla } h \leq h_0 \text{ i } n = 0, 1, \dots, k - 1.$$

Zbieżność liniowych metod wielokrokowych

Mamy następujący **pierwszy warunek konieczny** zbieżności.

Twierdzenie 2.

Zbieżna liniowa metoda wielokrokowa (\star) jest zgodna.

Czy jest to warunek wystarczający?

Zgodność to za mało

Przykład 6.

Rozważmy metodę dwukrokową

$$x_{n+2} + 4x_{n+1} - 5x_n = h(4f_{n+1} + 2f_n).$$

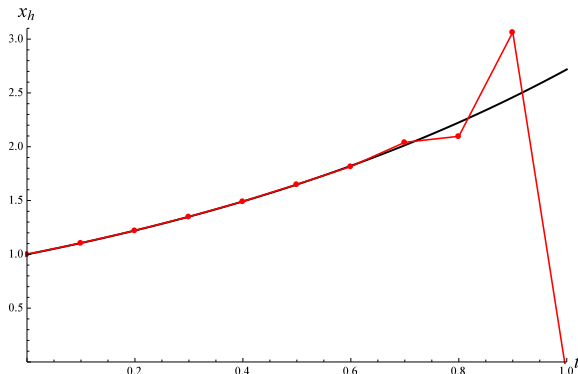
Można łatwo sprawdzić, że jest to metoda zgodna rzędu 3.

Zastosujmy ją do zagadnienia

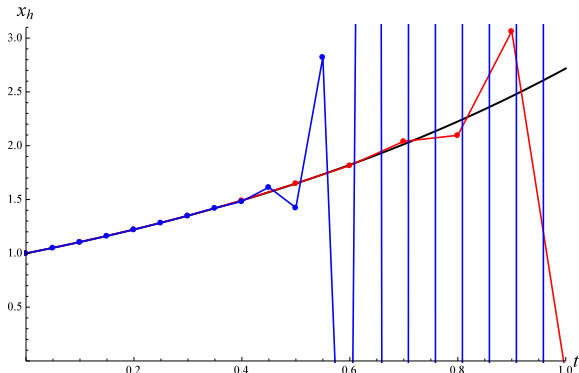
$$x'(t) = x(t), \quad x(0) = 1,$$

które ma znane rozwiązanie $x(t) = e^t$. Ponieważ metoda jest dwukrokowa, musimy znać x_0 oraz x_1 – przyjmujemy $x_0 = 1$ oraz $x_1 = e^h$.

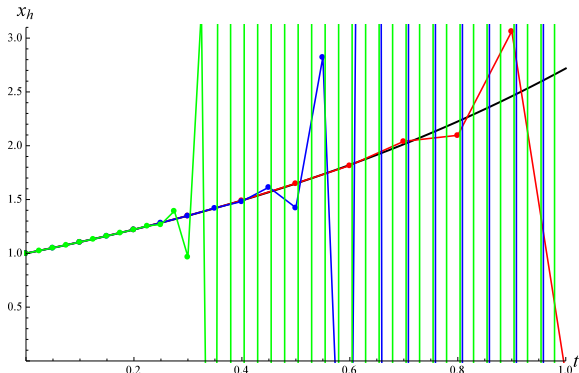
Zgodność to za mało

Numeryczne przybliżenie rozwiązania dla $h = 0,1$:

Zgodność to za mało

Numeryczne przybliżenie rozwiązania dla $h = 0,05$:

Zgodność to za mało

Numeryczne przybliżenie rozwiązania dla $h = 0,025$:

Zgodność to za mało

W dalszej części wykażemy, że aby metoda wielokrokowa była zbieżna, to nie wystarczy by była zgodna, potrzebny jest jeszcze dodatkowy warunek, tzn. *stabilność*:

zbieżność = zgodność + stabilność.

Stabilność liniowych metod wielokrokowych

Liniowa metoda wielokrokowa:

$$\begin{aligned} \alpha_k x_{n+k} + \alpha_{k-1} x_{n+k-1} + \dots + \alpha_0 x_n \\ = h \cdot (\beta_k f_{n+k} + \beta_{k-1} f_{n+k-1} + \dots + \beta_0 f_n). \quad (*) \end{aligned}$$

Definicja 6.

Liniowa metoda k -krokowa $(*)$ jest **0-stabilna** (**D-stabilna** lub **stabilna w sensie Dahlquista**), jeśli wszystkie pierwiastki pierwszego wielomianu charakterystycznego $\rho(r) = \alpha_k r^k + \alpha_{k-1} r^{k-1} + \dots + \alpha_0$ leżą wewnątrz lub na brzegu koła jednostkowego, a pierwiastki leżące na brzegu są pojedyncze.

Stabilność liniowych metod wielokrokowych

W przypadku jawnych metod Adamsa, tzn.

$$x_{n+1} = x_n + h \sum_{j=0}^{k-1} \gamma_j \nabla^j f_n$$

mamy zawsze

$$\rho(r) = r^k - r^{k-1} = r^{k-1}(r - 1),$$

a zatem jawne metody Adamsa są 0-stabilne.

Stabilność liniowych metod wielokrokowych

Przykład 6 (jeszcze raz).

Zauważmy, że dla metody

$$x_{n+2} + 4x_{n+1} - 5x_n = h(4f_{n+1} + 2f_n)$$

mamy $\rho(r) = r^2 + 4r - 5 = (r - 1)(r + 5)$ i $\sigma(r) = 4r + 2$.

Jest to więc metoda zgodna, ale niestabilna.

Zbieżność liniowych metod wielokrokowych

W ogólności, stabilność jest **drugim warunkiem koniecznym** zbieżności.

Twierdzenie 3.

Zbieżna liniowa metoda wielokrokowa (\star) jest 0-stabilna.

Stabilność, w połączeniu ze zgodnością metody, stanowi również **warunek wystarczający**.

Zbieżność liniowych metod wielokrokowych

Twierdzenie 4 (Dahlquist, 1956).

Jeśli liniowa metoda wielokrokowa (\star) jest 0-stabilna i zgodna rzędu p , to jest zbieżna rzędu p .

Jaki jest jednak maksymalny rząd zbieżności dla metody k -krokowej?

Maksymalny rząd zbieżności

Twierdzenie 5 (Pierwsza bariera Dahlquist, 1959).

Rząd p 0-stabilnej liniowej metody k -krokowej spełnia warunki:

- 1 $p \leq k + 2$ jeśli k jest parzyste,
- 2 $p \leq k + 1$ jeśli k jest nieparzyste,
- 3 $p \leq k$ jeśli $\beta_k/\alpha_k \leq 0$ (w szczególności dla metod jawnych).

Czy zbieżna to już użyteczna?

Zgodność i stabilność stanowią minimalne wymagania wobec użytecznej metody całkowania równań zwyczajnych, jednak nie uwzględniają one w ogóle długości kroku siatki.

To znacznie ogranicza użyteczność takich metod.

Zbieżna to jeszcze nie użyteczna

Przykład 7.

Rozważmy zagadnienie początkowe

$$x'(t) = -x(t), \quad t > 0,$$

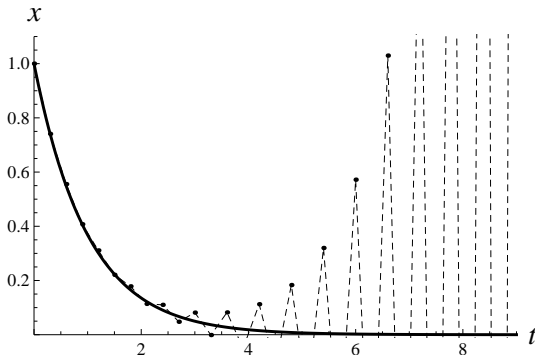
z warunkiem $x(0) = 1$, a także tzw. *metodę punktu środkowego* (ang. *midpoint rule* lub *leapfrog method*):

$$x_{n+2} - x_n = 2hf_{n+1}.$$

Zauważmy, że $\rho(r) = r^2 - 1$ oraz $\sigma(r) = 2r$, zatem metoda jest zgodna rzędu 2 oraz stabilna, a więc zbieżna.

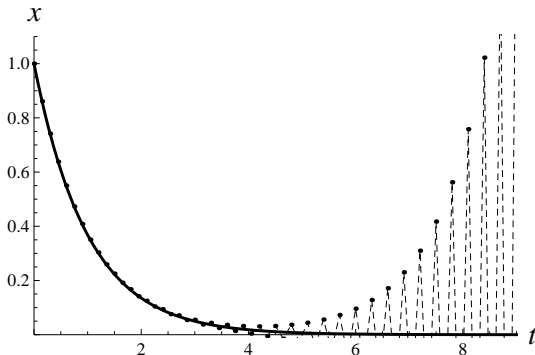
Zbieżna to jeszcze nie *użyteczna*

Numeryczne przybliżenie rozwiązania dla $h = 0,3$:



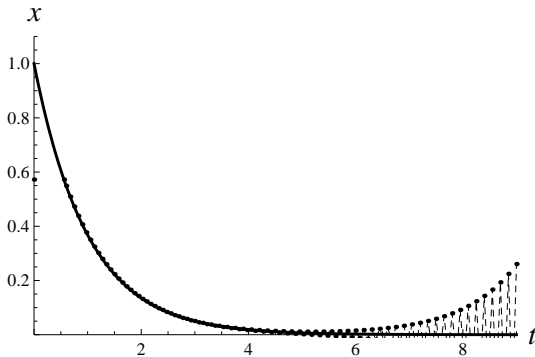
Zbieżna to jeszcze nie użyteczna

Numeryczne przybliżenie rozwiązania dla $h = 0,15$:



Zbieżna to jeszcze nie użyteczna

Numeryczne przybliżenie rozwiązania dla $h = 0,075$:



Zbieżna to jeszcze nie użyteczna

- 1 Pojęcie zbieżności (a także zgodności i stabilności) są pojęciami granicznymi.
- 2 0-stabilność oznacza, że przy przejściu do granicy z długością kroku rozwiązanie jest stabilne...
- 3 ... ale w praktyce nie da się „przejsć do granicy” z długością kroku.
- 4 Wprowadza się dodatkowe wymaganie – **A-stabilność** (lub **absolutną stabilność**) metody...
- 5 ... ale o tym już przy innej okazji :-)

I znów metoda Eulera...

Wróćmy jeszcze do metody, od której zaczęliśmy – metody Eulera (w przypadku skalarnym). Tym razem jednak, zamiast liczyć od razu x_{n+1} na podstawie x_n , wprowadźmy *punkt pośredni*:

$$x_{n+\frac{1}{2}} = x_n + \frac{h}{2} f(t_n, x_n),$$
$$x_{n+1} = x_{n+\frac{1}{2}} + \frac{h}{2} f(t_{n+\frac{1}{2}}, x_{n+\frac{1}{2}}).$$

Spróbujmy zapisać te dwa wyrażenia w jednym wzorze.

... ale w nowym wydaniu

$$x_{n+1} = x_n + h \cdot \left(\underbrace{\frac{1}{2} f(t_n, x_n)}_{\substack{\uparrow \\ b_1}} + \underbrace{\frac{1}{2} f\left(t_n + \frac{1}{2} h, x_n + \frac{1}{2} h f(t_n, x_n)\right)}_{\substack{\uparrow \\ k_1}} \right)$$

\uparrow b_2
 \uparrow k_2

\downarrow a
 \downarrow a
 \downarrow k_1

Nachylenie stycznej do przybliżonego rozwiązania jest **średnią** nachyleń rzeczywistego rozwiązania przechodzącego przez różne punkty (t_n oraz $t_n + ah$).

Dwustopniowa metoda Runge-Kutty

Jawne 2-stopniowe metody Rungego-Kutty możemy zapisać w postaci:

$$k_1 = f(t_n, x_n),$$

$$k_2 = f(t_n + ah, x_n + ahk_1),$$

$$x_{n+1} = x_n + h(b_1k_1 + b_2k_2).$$

Jak dobrać wartości współczynników a , b_1 i b_2 , aby osiągnąć jak najlepszy rząd metody?

Dwustopniowa metoda Runge-Kutty

Skorzystamy (jak zwykle) ze wzoru Taylora, ale zastosujemy go do wyrażenia k_2 :

$$f(t + \alpha h, x + \beta h) = f(t, x) + h(\alpha f_t(t, x) + \beta f_x(t, x)) + \mathcal{O}(h^2).$$

Zauważmy, że u nas $\alpha = a$ oraz $\beta = ak_1$, stąd dla $(t, x) = (t_n, x_n)$:

$$f(t_n + ah, x_n + ahk_1) = f_n + ah(f_t + f f_x)|_{t=t_n, x=x_n} + \mathcal{O}(h^2).$$

Dwustopniowa metoda Runge-Kutty

Po podstawieniu do wzoru na x_{n+1} daje to

$$\begin{aligned}x_{n+1} &= x_n + h(b_1 k_1 + b_2 k_2) = x_n + h b_1 k_1 + h b_2 k_2 \\&= x_n + h b_1 f_n + h b_2 \left(f_n + a h (f_t + f f_x)|_{t=t_n, x=x_n} + \mathcal{O}(h^2) \right) \\&= x_n + h(b_1 + b_2) f_n + h^2 a b_2 (f_t + f f_x)|_{t=t_n, x=x_n} + \mathcal{O}(h^3).\end{aligned}$$

Z drugiej strony (jak zwykle z Twierdzenia Taylora)

$$x(t_n + h) = x(t_n) + h \cdot 1 \cdot f_n + h^2 \cdot \frac{1}{2} \cdot (f_t + f f_x)|_{t=t_n, x=x_n} + \mathcal{O}(h^3).$$

Dwustopniowa metoda Runge-Kutty

Przypuśćmy, że wartość przybliżenia w kroku t_n jest **dokładna** (tzn. mamy $x_n = x(t_n)$), wówczas

$$x(t_{n+1}) - x_{n+1} = h \cdot (1 - b_1 - b_2) \cdot f_n \\ + h^2 \cdot \left(\frac{1}{2} - ab_2 \right) \cdot (f_t + ff_x)|_{t=t_n, x=x_n} + \mathcal{O}(h^3).$$

Powstałą w ten sposób wielkość nazywamy **błędem lokalnym**. W ogólności ta metoda ma błąd lokalny rzędu $\mathcal{O}(h)$ – nie jest to najlepszy rezultat (metoda nie jest nawet zgodna). Ale...

Rząd zbieżności 2-stopniowej metody RK

... zerując kolejne współczynniki otrzymujemy rząd metody równy 1 lub 2.

błąd lokalny		warunki osiągnięcia rzędu metody		rząd*
$\mathcal{O}(h^2)$	jeśli	$b_1 + b_2 = 1$	oraz dowolne a	$p = 1$
$\mathcal{O}(h^3)$	jeśli	$b_1 + b_2 = 1$	oraz $ab_2 = \frac{1}{2}$	$p = 2$

* W przypadku metod RK mówimy, że metoda jest rzędu p , jeśli błąd lokalny jest $\mathcal{O}(h^{p+1})$.

Rząd zbieżności 2-stopniowej metody RK

Przyjmując $b_2 = \theta$ otrzymujemy rodzinę 2-stopniowych jawnych metod RK rzędu 2 (dla $\theta \neq 0$), gdzie $b_1 = 1 - \theta$, $a = \frac{1}{2\theta}$. Najpopularniejsze z nich to:

- 1 *ulepszona metoda Eulera*: $\theta = \frac{1}{2}$,
- 2 *zmodyfikowana metoda Eulera*: $\theta = 1$.

Można pokazać, że nie da się dobrać parametrów a , b_1 i b_2 tak, by rząd metody był równy 3.

Klasa metod Runge-Kutty

Ogólna r -stopniowa metoda RK może zostać zapisana w postaci:

$$x_{n+1} = x_n + h \sum_{i=1}^r b_i k_i,$$

gdzie k_i są obliczane na podstawie wartości funkcji f :

$$k_i = f\left(t_n + c_i h, x_n + h \sum_{j=1}^r a_{i,j} k_j\right), \quad i = 1, \dots, r.$$

Widzimy, że nachylenie stycznej do przybliżonego rozwiązania będzie **średnią ważoną** nachyleń w różnych punktach pośrednich (o ile mamy odpowiednie c_i).

Klasa metod Runge-Kutty

Wobec tego naturalnym warunkiem jest nałożenie ograniczenia

$$c_i = \sum_{j=1}^r a_{i,j}, \quad i = 1, \dots, r.$$

Metoda zawiera wówczas $r^2 + r$ parametrów $\{a_{i,j}, b_j\}$.

Tablica Butchera

c_1	$a_{1,1}$	$a_{1,2}$	\dots	$a_{1,r-1}$	$a_{1,r}$
c_2	$a_{2,1}$	$a_{2,2}$	\dots	$a_{2,r-1}$	$a_{2,r}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
c_r	$a_{r,1}$	$a_{r,2}$	\dots	$a_{r,r-1}$	$a_{r,r}$
	b_1	b_2	\dots	b_{r-1}	b_r

Parametry metody RK często zapisuje się w postaci tabeli (tzw. **tablica Butchera**).

Tablica Butchera

c_1	$a_{1,1}$	$a_{1,2}$	\dots	$a_{1,r-1}$	$a_{1,r}$
c_2	$a_{2,1}$	$a_{2,2}$	\dots	$a_{2,r-1}$	$a_{2,r}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
c_r	$a_{r,1}$	$a_{r,2}$	\dots	$a_{r,r-1}$	$a_{r,r}$
	b_1	b_2	\dots	b_{r-1}	b_r

Zauważmy, że w ogólności wyznaczenie wartości k_i wymaga rozwiązania r (najczęściej nieliniowych) równań:

$$k_i = f\left(t_n + c_i h, x_n + h \sum_{j=1}^r a_{i,j} k_j\right), \quad i = 1, \dots, r.$$

Metody RK są więc w ogólności metodami **niejawnymi**.

Tablica Butchera

c_1	$a_{1,1}$	$a_{1,2}$	\dots	$a_{1,r-1}$	$a_{1,r}$
c_2	$a_{2,1}$	$a_{2,2}$	\dots	$a_{2,r-1}$	$a_{2,r}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
c_r	$a_{r,1}$	$a_{r,2}$	\dots	$a_{r,r-1}$	$a_{r,r}$
	b_1	b_2	\dots	b_{r-1}	b_r

Jeśli jednak tablica Butchera jest ściśle trójkątna dolna (tzn. mamy $a_{i,j} = 0$ dla wszystkich $j \geq i$), to metoda staje się **jawna** – nie trzeba wówczas rozwiązywać żadnych równań nieliniowych.

Rząd metody RK

Definicja 7.

Mówimy, że metoda Runge-Kutty jest rzędu p , jeśli dla dostatecznie gładkich funkcji x mamy

$$|x(t_0 + h) - x_1| \leq C \cdot h^{p+1},$$

tzn. błąd lokalny jest równy $\mathcal{O}(h^{p+1})$.

Określenie warunków, by metoda miała odpowiedni rząd wiąże się zazwyczaj ze żmudnymi obliczeniami (widzieliśmy to na przykładzie metody 2-stopniowej). Ograniczymy się do metod jawnych.

Rząd metody – warunki dla metody rzędu 3

Twierdzenie 6.

Jawna metoda RK jest rzędu 3 wtedy i tylko wtedy, gdy:

$$\sum_{j=1}^r b_j = 1,$$

$$\sum_{j=2}^r \sum_{k=1}^{j-1} b_j a_{j,k} = \frac{1}{2},$$

$$\sum_{j=2}^r \sum_{k=1}^{j-1} \sum_{\ell=1}^{j-1} b_j a_{j,k} a_{j,\ell} = \frac{1}{3},$$

$$\sum_{j=3}^r \sum_{k=2}^{j-1} \sum_{\ell=1}^{k-1} b_j a_{j,k} a_{k,\ell} = \frac{1}{6}.$$

Pierwszy warunek gwarantuje rząd 1, kolejny – rząd 2 (to już widzieliśmy), a kolejne dwa gwarantują rząd 3. Musimy mieć zatem co najmniej 3-stopniową metodę.

Rząd metody – warunki dla metody rzędu 4

Twierdzenie 7.

Jawna metoda RK jest rzędu 4 wtedy i tylko wtedy, gdy są spełnione warunki dla metody rzędu 3 oraz:

$$\sum_{j=2}^r \sum_{k=1}^{j-1} \sum_{\ell=1}^{j-1} \sum_{m=1}^{j-1} b_j a_{j,k} a_{j,\ell} a_{j,m} = \frac{1}{4}, \quad \sum_{j=3}^r \sum_{k=1}^{j-1} \sum_{\ell=2}^{j-1} \sum_{m=1}^{\ell-1} b_j a_{j,k} a_{j,\ell} a_{\ell,m} = \frac{1}{8},$$

$$\sum_{j=3}^r \sum_{k=2}^{j-1} \sum_{\ell=1}^{k-1} \sum_{m=1}^{k-1} b_j a_{j,k} a_{k,\ell} a_{k,m} = \frac{1}{12}, \quad \sum_{j=4}^r \sum_{k=3}^{j-1} \sum_{\ell=2}^{k-1} \sum_{m=1}^{\ell-1} b_j a_{j,k} a_{k,\ell} a_{\ell,m} = \frac{1}{24}.$$

Warunki robią się coraz bardziej skomplikowane.

Ile stopni powinna mieć metoda rzędu p ?

- Można przypuszczać, że metoda rzędu p potrzebuje p stopni (zgadza się dla $p = 1, 2, 3, 4$). Jednak **nie** jest to prawda dla $p > 4$.
- Znane są dokładne liczby potrzebnych stopni dla $p \leq 8$.
- Aby uzyskać rząd metody równy 9, potrzeba między 12 a 17 stopni, a parametry muszą spełniać 486 nieliniowych równań algebraicznych. :-)

Najpopularniejsza z metod RK

$$\begin{array}{c|cccc}
 0 & 0 & & & \\
 \frac{1}{2} & \frac{1}{2} & 0 & & \\
 \frac{1}{2} & 0 & \frac{1}{2} & 0 & \\
 1 & 0 & 0 & 1 & 0 \\
 \hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
 \end{array}$$

$$\begin{array}{c|cccc}
 0 & 0 & & & \\
 \frac{1}{3} & \frac{1}{3} & 0 & & \\
 \frac{2}{3} & -\frac{1}{3} & 1 & 0 & \\
 1 & 1 & -1 & 1 & 0 \\
 \hline
 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
 \end{array}$$

W 1901 roku Kutta zaproponował dwie metody 4-stopniowe, które osiągają rząd 4. Powyżej znajdują się tablice Butchera tych metod.

Pierwsza z nich jest bardziej popularna (nazywana jest często po prostu **metodą Runge-Kutty**), a druga jest bardziej dokładna.

Zbieżność metod Runge-Kutty

Zakładamy, że zagadnienie $x'(t) = f(t, x(t))$ dla $t > t_0$, $x(t_0) = x_0$ ma jednoznaczne rozwiązanie. Oznaczmy przez x_h łamaną łączącą punkty rozwiązania numerycznego z krokiem h .

Twierdzenie 8.

Przypuśćmy, że w pewnym otoczeniu rozwiązania x wyjściowego zagadnienia zachodzi $|f_x| \leq L$, a także, że błąd lokalny nie przekracza wartości $C \cdot h^{p+1}$. Wówczas dla odpowiednio małych $h > 0$ mamy

$$|x(t) - x_h(t)| \leq \frac{C}{L} \left(e^{L(t-t_0)} - 1 \right) \cdot h^p.$$

Porównanie jawnych metod stałokrokowych

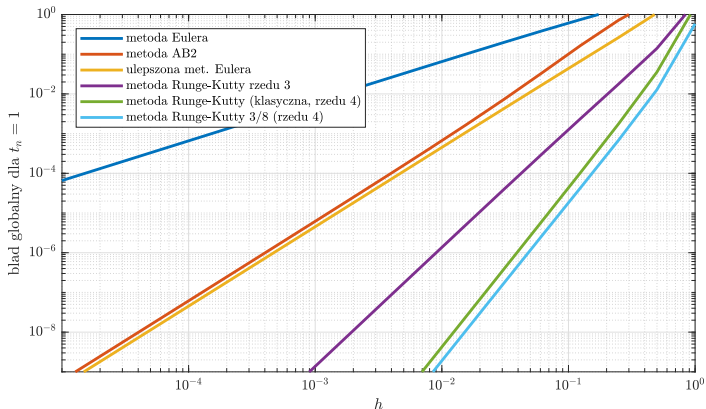
Wróćmy jeszcze do Przykładu 5 z równaniem

$$x'(t) - x(t) = -\frac{1}{2}e^{t/2} \sin(5t) + 5e^{t/2} \cos(5t), \quad t > 0,$$

z warunkiem $x(0) = 0$.

Porównajmy błąd globalny dla $t = 1$ przy różnych metodach i różnych wartościach h .

Porównanie jawnych metod stałokrokowych



Podsumowanie

- Każdą z przedstawionych metod można uogólnić na rozwiązywanie układów równań.
- Na laboratorium, poza implementacją podstawowych metod, zajmiemy się również kwestią doboru wielkości kroku siatki tak, aby osiągnąć pożądaną poziom dokładności rozwiązania (czyli zajmiemy się adaptacyjnym wyborem kroku).

KONIEC CZĘŚCI 2