Równania różniczkowe zwyczajne z laboratorium

notatki do wykładu

Piotr Kowalczyk

Wydział Matematyki, Informatyki i Mechaniki
Uniwersytetu Warszawskiego

15 maja 2025

Spis treści

Rozdział 1

Wprowadzenie

Równania różniczkowe to takie równania, w których występują pochodne funkcji szukanych. W przypadku równań różniczkowych *zwyczajnych* niewiadome są funkcjami jednej zmiennej.

Wiele zagadnień z różnych dziedzin w naturalny sposób prowadzi do równań lub układów równań różniczkowych zwyczajnych (RRZ). Stąd są one jednym z najważniejszych działów matematyki mającym praktyczne zastosowanie, a rozwijanym nieprzerwanie od czasów Newtona. Wiele spośród tych równań nie ma rozwiązań analitycznych lub trudno je uzyskać, dlatego istnieje konieczność rozwiązywania równań różniczkowych zwyczajnych przy użyciu komputerów.

Wykład rozpoczniemy od kilku przykładów zjawisk fizycznych, których model matematyczny oparty jest na równaniach różniczkowych zwyczajnych.

1.1. Przykłady

Rozpad radioaktywny. Niech $m(t)$ oznacza masę substancji radioaktywnej w chwili t. Wiadomo, że zmiana tej masy w czasie Δt jest wprost proporcjonalna (ze współczynnikiem $k > 0$) do masy w chwili t i jest równa $-km(t)\Delta t$ (następuje ubytek masy względem początkowej wartości $m(t)$). Zatem

$$m(t + \Delta t) - m(t) = -km(t)\Delta t,$$

czyli

$$\frac{m(t + \Delta t) - m(t)}{\Delta t} = -km(t).$$

Przechodząc do granicy $\Delta t \to 0^+$ otrzymujemy równanie różniczkowe zwyczajne pierwszego rzędu (tego rzędu jest najwyższa pochodna w równaniu)

$$m'(t) = -km(t).$$

Wartość pochodnej po lewej stronie równania oznacza chwilową zmianę masy w chwili t.

Uwaga. Zazwyczaj, jeśli nie prowadzi to do nieporozumień, pomija się w zapisie równania zmienną niezależną. Ponadto, przede wszystkim w zastosowaniach fizycznych, do oznaczenia pochodnej po czasie często stosuje się kropkę nad nazwą funkcji (notacja ta pochodzi od Newtona). Od tego momentu będziemy zwykle stosować te oznaczenia.

Model wzrostu populacji. Niech $p(t)$ oznacza ilość bakterii w chwili t. Wzrost populacji zależy wprost proporcjonalnie (ze współczynnikiem $k > 0$) od liczebności populacji, stąd podobnie jak poprzednio mamy

$$\dot{p} = kp.$$

W tym najprostszym modelu nie uwzględniamy ograniczeń na wzrost populacji. Dalej będzie przykład bardziej skomplikowanego modelu populacyjnego uwzględniającego pewne ograniczenia.

Dynamika punktu materialnego. Niech $x(t)$ oznacza położenie na prostej punktowej masy m w chwili t. Wtedy $v(t) := \dot{x}(t)$ to prędkość, a $a(t) := \dot{v}(t) = \ddot{x}(t)$ to przyspieszenie. II zasada dynamiki Newtona mówi, że przyspieszenie, z jakim porusza się obiekt, jest proporcjonalne do działającej na ten obiekt siły f. Jeśli ta siła zależy tylko od czasu, położenia i prędkości ciała, to dostajemy równanie

$$m\ddot{x} = f(t, x, \dot{x}),$$

które jest równaniem różniczkowym zwyczajnym drugiego rzędu w postaci rozwikłanej (normalnej) — najwyższa pochodna w równaniu (druga) zależy w jawny sposób od pozostałych pochodnych, funkcji szukanej i czasu.

Jako przykład rozważymy *spadek ciała w powietrzu*. Na spadające ciało działa siła grawitacji: $-mg$ (z minusem, bo jest skierowana w dół, a przyjmujemy układ współrzędnych skierowany w górę) oraz skierowana przeciwnie do kierunku prędkości (prędkość skierowana w dół, czyli ujemna) siła oporu powietrza, o której założymy, że jest wprost proporcjonalna (ze współczynnikiem $k > 0$) do prędkości: $-kv(t)$. Podsumowując otrzymujemy równanie

$$m\ddot{x} = \underbrace{-mg - k\dot{x}}_{=:\, f(t,x,\dot{x})}.$$

Równanie brachistochrony. Równania rzędu k-tego postaci

$$F(t, x, x', \ldots, x^{(k)}) = 0$$

nazywamy równaniami w postaci uwikłanej. Tego typu równania stosunkowo rzadko pojawiają się w praktyce. Jako przykład podamy równanie

$$[1 + (y'(x))^2]y(x) = k^2,$$

którego rozwiązaniem jest *brachistochrona*, czyli krzywa najkrótszego spadku — krzywa, po której masa punktowa pod wpływem siły ciężkości stacza się w możliwie najkrótszym czasie.

Dynamika populacji. Rozważymy teraz środowisko z dwiema populacjami: drapieżników i ofiar. Niech $x_1(t)$ oznacza liczebność populacji ofiar, a $x_2(t)$ — populacji drapieżników. Zakładamy dla populacji ofiar, że zmiana liczebności populacji jest wprost proporcjonalna do liczebności (prosty model rozważany powyżej). Współczynnik proporcjonalności k zmodyfikujemy tak, aby uwzględnić drapieżniki, $k = a - bx_2(t)$: ofiary giną pożerane przez drapieżniki, przy czym im więcej jest drapieżników, tym łatwiej giną ofiary. W przypadku drapieżników zakładamy, że ubytek populacji w wyniku zgonów jest wprost proporcjonalny do liczebności populacji (ze współczynnikiem d), a przyrost liczebności populacji zależy wprost proporcjonalnie od dostępnego pożywienia (populacji ofiar): $cx_1(t)$. O współczynnikach a, b, c, d zakładamy, że są dodatnie. Otrzymujemy zatem układ równań

$$\begin{aligned} \dot{x}_1 &= (a - bx_2)x_1, \\ \dot{x}_2 &= (cx_1 - d)x_2. \end{aligned}$$

Układ ten nosi nazwę układu (lub modelu) *Lotki-Volterry*, a został zaproponowany w 1926 r. przez Vito Volterrę jako model populacji ryb w Adriatyku.

1.2. Rozwiązanie RRZ

Co to znaczy rozwiązać równanie różniczkowe zwyczajne? Nieformalnie można powiedzieć, że znaczy to wskazać funkcję, która spełnia to równanie. Zanim poznamy formalną definicję, spróbujemy wskazać funkcje spełniające niektóre z równań, które już poznaliśmy.

Przykład (Rozpad radioaktywny)**.** Zmiana masy substancji radioaktywnej dana jest równaniem

$$m' = -km, \quad k > 0.$$

Naturalne jest założenie, że $m(t) > 0$, a wtedy możemy zapisać

$$\frac{m'}{m} = -k.$$

Całkujemy obie strony równania i dostajemy

$$L := \int \frac{m'(t)}{m(t)}\,dt = \int (-k)\,dt =: P.$$

Dla prawej strony mamy

$$P = \int (-k)\,dt = -kt + C_1,$$

a dla lewej, całkując przez podstawienie $m := m(t)$, $dm = m'(t)dt$, mamy

$$L = \int \frac{m'(t)}{m(t)}\,dt = \int \frac{1}{m}\,dm = \ln(m) + C_2 = \ln(m(t)) + C_2.$$

Przyrównując obie strony dostajemy

$$\ln(m(t)) = -kt + \widetilde{C},$$

a stąd

$$m(t) = \mathrm{e}^{-kt+\widetilde{C}} = C\mathrm{e}^{-kt}.$$

Jeśli w chwili $t = 0$ mamy $m(0) = m_0$, to możemy wyznaczyć stałą C i $m(t) = m_0\mathrm{e}^{-kt}$.

Przykład (Spadek w próżni)**.** Na podstawie wcześniej rozważanego przykładu spadku z oporem powietrza możemy napisać równanie spadku w próżni:

$$m\ddot{x} = -mg.$$

Dzieląc obie strony przez m i całkując po t mamy $\dot{x} = -gt + C_1$, czyli korzystając z definicji prędkości $v(t) = -gt + C_1$. Przyjmując, że w chwili $t = 0$ mamy $v(0) = v_0$, dostajemy $C = v_0$, zatem $v(t) = -gt + v_0$. Całkujemy otrzymane równanie $\dot{x} = -gt + v_0$ po t, co daje $x(t) = -\frac{1}{2}gt^2 + v_0t + C_2$. Przyjmując, że w chwili $t = 0$ mamy $x(0) = x_0$, dostajemy $C_2 = x_0$, co daje

$$x(t) = x_0 + v_0t - \frac{1}{2}gt^2 = x_0 + (v_0 - \frac{1}{2}gt)t,$$

gdzie w nawiasie jest prędkość wypadkowa, będąca sumą prędkości początkowej i chwilowej.

Podamy teraz formalną definicję rozwiązania RRZ rzędu pierwszego w ogólnym przypadku.

Definicja 1.1. Niech $f\colon D \to \mathbb{R}^n$, gdzie $D \subset \mathbb{R} \times \mathbb{R}^n$ jest spójnym i niepustym zbiorem. Powiemy, że $x\colon I \to \mathbb{R}^n$ jest *rozwiązaniem układu* n równań różniczkowych zwyczajnych

$$x' = f(t, x) \tag{1.1}$$

na przedziale I (otwartym lub domkniętym, ograniczonym lub nie) wtedy i tylko wtedy, gdy
a) x jest różniczkowalna w I (jeśli I jest domknięty, to x ma odpowiednie pochodne jednostronne),
b) $\forall\, t \in I\ (t, x(t)) \in D$,
c) dla każdego $t \in I$ $x(t)$ spełnia układ (1.1).

W przypadku równań wyższych rzędów mamy następującą definicję.

Definicja 1.2. Niech $g\colon D \to \mathbb{R}$, gdzie $D \subset \mathbb{R} \times \mathbb{R}^k$ jest spójnym i niepustym zbiorem. Powiemy, że $u\colon I \to \mathbb{R}$ jest *rozwiązaniem* równania różniczkowego zwyczajnego k-tego rzędu

$$u^{(k)} = g(t, u, u', \ldots, u^{(k-1)}) \tag{1.2}$$

na przedziale I (otwartym lub domkniętym, ograniczonym lub nie) wtedy i tylko wtedy, gdy

a) u jest k-krotnie różniczkowalna w I (jeśli I jest domknięty, to x ma odpowiednie pochodne jednostronne),
b) $\forall\, t \in I\ (t, u(t), u'(t), \ldots, u^{(k-1)}) \in D$,
c) dla każdego $t \in I$ $u(t)$ spełnia równanie (1.2).

Okazuje się, że równanie k-tego rzędu można sprowadzić do układu k równań rzędu 1.
Niech $x_i(t) = u^{(i-1)}(t)$ dla $i = 1, \ldots, k$. Wtedy

$$x = \begin{bmatrix} u \\ u' \\ \vdots \\ u^{(k-2)} \\ u^{(k-1)} \end{bmatrix} \quad \text{oraz} \quad x' = \begin{bmatrix} u' \\ u'' \\ \vdots \\ u^{(k-1)} \\ u^{(k)} \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_k \\ g(t,x) \end{bmatrix}.$$

Zatem (1.1) jest równoważne (1.2) dla

$$f(t,x) = \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_k \\ g(t,x) \end{bmatrix} \tag{1.3}$$

z tymi samymi zbiorami I oraz D. Rzeczywiście, jeśli u jest rozwiązaniem (1.2), to $u^{(i-1)}$ są różniczkowalne i spełniają (1.1) z f zdefiniowaną wzorem (1.3). Na odwrót, jeśli x jest rozwiązaniem (1.1) z (1.3), to ponieważ $x_i' = x_{i+1}$ i $x_k' = x_1^{(k)}$ oraz na mocy ciągłości x_i i ich różniczkowalności, x_1 spełnia równanie (1.2).

Uwaga. Analogicznie układ m równań rzędu większego niż 1 można sprowadzić do układu równań rzędu 1.

Uwaga. Możliwość sprowadzenia równania wyższego rzędu do układu równań pozwala skupić się na badaniu układów równań — wystarczy umieć rozwiązywać układy RRZ rzędu 1. Często na określenie (1.1) używa się nazwy *równanie* wymiennie z *układ równań*.

Na koniec tego podrozdziału wprowadzimy dalsze definicje.

Definicja 1.3.
a) Rzut zbioru D na $\mathbb{R}^n$ (pomijamy zmienną t) nazywamy *przestrzenią fazową*, zaś sam zbiór D nazywamy *rozszerzoną przestrzenią fazową* równania (1.1).
b) Wykres rozwiązania równania (1.1) w $\mathbb{R}^{n+1}$ nazywamy *krzywą całkową* równania (1.1). Rzut krzywych całkowych na przestrzeń fazową nazywamy *portretem fazowym* równania, zaś rzut jednej krzywej całkowej nazywamy *krzywą fazową*.

1.3. Interpretacja geometryczna

Przypadek skalarny. Rozważamy równanie skalarne $x' = f(t,x)$ z rozwiązaniem $x\colon (a,b) \to \mathbb{R}$. Wykres funkcji $x(t)$ jest krzywą na płaszczyźnie. Wektory styczne do tej krzywej w punkcie (t,x) mają postać

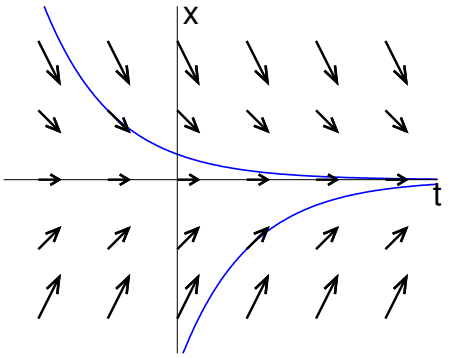
$$[1, x'(t)] = [1, f(t, x(t))]$$

(nachylenie wektorów $[1, f(t,x(t))]$ odpowiada kątowi, którego tangens wynosi $f(t,x(t))$). Zatem możemy przyjąć, że na płaszczyźnie (t,x) RRZ wyznacza *pole kierunków* $[1, f(t,x(t))]$ (pole wektorowe) zaczepionych w punktach (t,x).

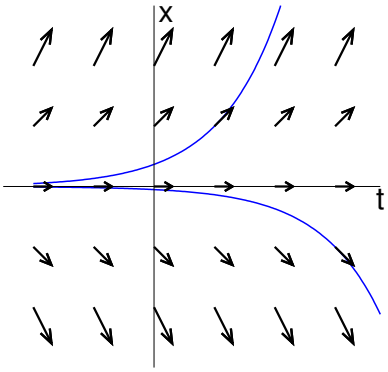
Zadanie rozwiązania RRZ odpowiada zadaniu znalezienia krzywych, będących wykresami funkcji $x(t)$, mających własność taką, że wektor styczny w (t,x) do wykresu $x(t)$ jest równy wektorowi pola kierunków.

Przypadek wektorowy. W przypadku ogólnym zastępujemy płaszczyznę (t, x) przestrzenią $\mathbb{R}^{n+1}$. Mamy wtedy analogicznie pole kierunków $[1, f_1(t, x), \ldots, f_n(t, x)] \in \mathbb{R}^{n+1}$ zaczepionych w punktach (t, x).

Przykłady:
1) $x' = -ax$, $a > 0$. Pole kierunków to $[1, -ax]$ zaczepione w (t, x) jest naszkicowane na rys. 1.1. Zauważmy, że pole to nie zależy od zmiennej t.



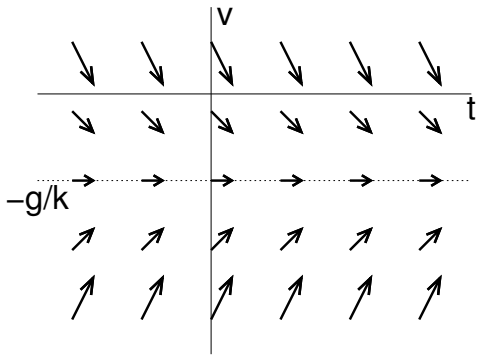
Rysunek 1.1: Pole kierunków równania $x' = -ax$.



Rysunek 1.2: Pole kierunków równania $x' = ax$.

2) $x' = ax$, $a > 0$. Pole kierunków to $[1, ax]$ zaczepione w (t, x) jest naszkicowane na rys. 1.2. Podobnie jak wyżej pole to nie zależy od zmiennej t.

3) $x'' = -g - kx'$ jest równoważne układowi $\begin{cases} x' = v, \\ v' = -g - kv. \end{cases}$ Pole kierunków to $[1, v, -g - kv]$. Nie zależy ono od zmiennej x, możemy zatem bez straty ogólności naszkicować je w płaszczyźnie (t, v), tzn. rysujemy pole $[1, -g - kv]$ (patrz rys. 1.3).



Rysunek 1.3: Pole kierunków równania $x'' = -g - kx'$ w zmiennych (t, v).

Pole kierunków jest takie jak pole z przykładu 1) przesunięte w dół. Zatem prędkość zawsze po dostatecznie długim czasie stabilizuje się na poziomie $-g/k$.

1.4. Równania autonomiczne

Definicja 1.4. Równanie różniczkowe

$$x' = f(x), \tag{1.4}$$

w którym prawa strona nie zależy bezpośrednio od t, nazywamy *równaniem autonomicznym*.

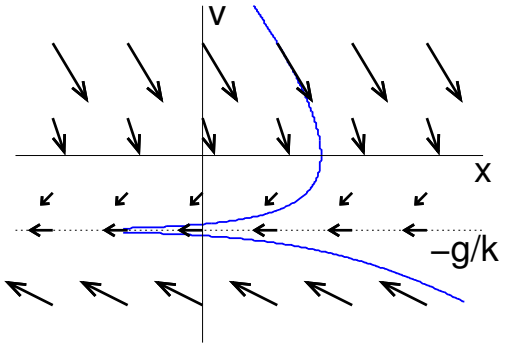
Uwaga. Każde równanie postaci $y' = g(t, y)$ można sprowadzić do równania autonomicznego. Oznaczmy $x(t) := \begin{bmatrix} t \\ y(t) \end{bmatrix}$, skąd $\frac{d}{dt}x(t) = \begin{bmatrix} 1 \\ y'(t) \end{bmatrix} = \begin{bmatrix} 1 \\ g(t,y) \end{bmatrix}$. Przyjmując dalej $f(x) := \begin{bmatrix} 1 \\ g(t,y) \end{bmatrix} = \begin{bmatrix} 1 \\ g(x) \end{bmatrix}$ dostajemy $x' = f(x)$.

Jeśli $x(t)$ jest rozwiązaniem (1.4), to wektor styczny do wykresu rozwiązania można zapisać jako

$$[x_1', x_2', \ldots, x_n'] = f(x).$$

Nie zależy on od t (tak jak i f) i możemy w przypadku równań autonomicznych ograniczyć się do rozważania przestrzeni fazowej.

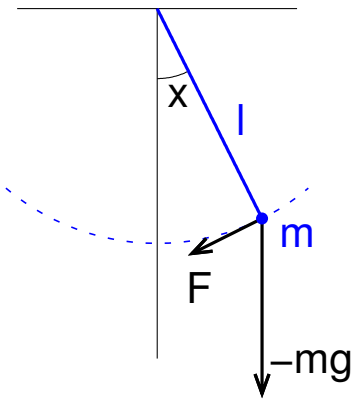
Przykład (spadek w powietrzu)**.** Portret fazowy równania spadku w powietrzu (w zmiennych (x, v)), czyli zależność prędkości od położenia naszkicowany jest na rys. 1.4. Zauważmy, że wektory pola kierunków dla $v = 0$ mają postać $[0, -g]$, zaś dla $v = -g/k$ mają postać $[-g/k, 0]$ (to jest przypadek spadku ze stałą prędkością — siła ciężkości jest równoważona siłą oporu — na tym poziomie stabilizuje się po dostatecznie długim czasie prędkość spadającego ciała). Krzywa fazowa prezentuje trajektorię rozwiązania w kolejnych chwilach czasu (trzeba podkreślić, że z portretu fazowego nie wynika, jak szybko porusza się punkt wzdłuż krzywej fazowej).



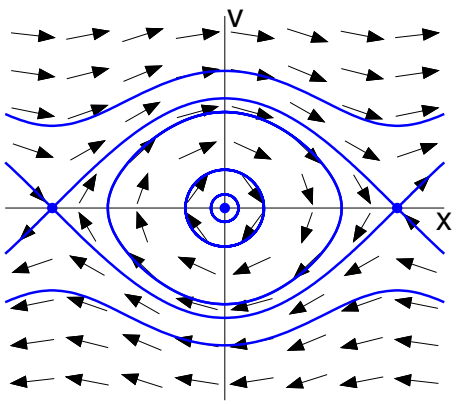
Rysunek 1.4: Portret fazowy równania spadku.

Uwaga. W przypadku układu autonomicznego dwóch równań skalarnych: $\dot{x} = F(x, v)$, $\dot{v} = G(x, v)$, możemy przejść do równania na funkcję $v = v(x(t))$ jako funkcję zmiennej x eliminując z równania t:

$$\frac{dv}{dt} = \frac{dv}{dx}\frac{dx}{dt}, \quad \text{skąd} \quad \frac{dv}{dx} = v' = \frac{G(x,v)}{F(x,v)}.$$



Rysunek 1.5: Wahadło matematyczne.



Rysunek 1.6: Portret fazowy równania wahadła.

Przykład (wahadło matematyczne)**.** Jako kolejny przykład rozważymy ruch wahadła matematycznego, rys. 1.5. Siła działająca na punkt materialny o masie m, to $F = -mg \sin x(t)$, gdzie $x(t)$ jest

kątem wychylenia wahadła. Z drugiej strony $F = m\ddot{s}(t)$, gdzie $s(t)$ jest położeniem punktu na łuku okręgu o promieniu l, po którym porusza się punkt materialny. Ponieważ zachodzi $s(t) = lx(t)$, to ostatecznie dostajemy równanie

$$\ddot{x} = -\frac{g}{l}\sin x.$$

Zatem przechodząc do układu równań rzędu 1 i przyjmując $k := \frac{g}{l} > 0$ mamy $\begin{cases} \dot{x} = v, \\ \dot{v} = -k\sin x. \end{cases}$

Portret fazowy tego układu naszkicowany jest na rys. 1.6.

1.5. Równania skalarne

W tej części wykładu zajmiemy się podstawowymi typami skalarnych równań różniczkowych zwyczajnych. Zbadamy, kiedy istnieją rozwiązania i podamy sposoby rozwiązywania. Będziemy rozważać następujące typy równań z warunkiem początkowym $x(t_0) = x_0$:

1) $\dot{x} = f(t)$,
2) $\dot{x} = f(x)$ — autonomiczne,
3) $\dot{x} = f(t)g(x)$ — o zmiennych rozdzielonych,
4) $\dot{x} = f(\frac{x}{t})$ — jednorodne,
5) $P(x,y)dx + Q(x,y)dy = 0$ — w postaci różniczek,
6) $\dot{x} = a(t)x + b(t)$ — liniowe.

Podstawowe typy równań skalarnych:

1) Równania postaci $\dot{x} = f(t)$.

Jeśli $f : (a,b) \to \mathbb{R}$ jest ciągła, to wystarczy scałkować równanie od t_0 do t dla $t_0, t \in (a,b)$:

$$\int_{t_0}^{t} \frac{dx}{ds}\,ds = \int_{t_0}^{t} f(s)\,ds \quad \Longrightarrow \quad x(t) = \int_{t_0}^{t} f(s) + x_0.$$

To rozwiązanie jest *jednoznaczne*: jeśli $x(t)$ i $y(t)$ są dwoma rozwiązaniami, to $\dot{x} - \dot{y} = 0$, czyli $x(t) - y(t) = C$ dla każdego $t \in (a,b)$. Ponieważ $x(t_0) = x_0 = y(t_0)$, to $C = 0$ i stąd $x(t) = y(t)$.

2) Równania autonomiczne.

Twierdzenie 1.1. *Niech $f : (c,d) \to \mathbb{R}$ będzie ciągła oraz $f(x) \neq 0$ dla $x \in (c,d)$. Wtedy przez każdy punkt $(t_0, x_0) \in \mathbb{R} \times (c,d)$ przechodzi dokładnie jedna krzywa całkowa równania $\dot{x} = f(x)$ dana wzorem*

$$x(t) = F^{-1}(t - t_0 + F(x_0)),$$

gdzie $F(x)$ jest funkcją pierwotną funkcji $1/f(x)$.

Dowód. Funkcja f jest różna od zera, więc mamy $\frac{1}{f(x)}\dot{x} = 1$. Niech $F(x)$ będzie funkcją pierwotną funkcji $1/f(x)$. Wtedy $\frac{d}{dt}F(x(t)) = 1$ i całkując w przedziale (t_0, t) otrzymujemy $F(x(t)) - F(x(t_0)) = t - t_0$, czyli $F(x(t)) = t - t_0 + F(x_0)$. Funkcja F jest monotoniczna, bo f jest stałego znaku. Mamy zatem $x(t) = F^{-1}(t - t_0 + F(x_0))$. Jednoznaczność rozwiązania wynika z definicji funkcji pierwotnej. □

Uwaga.

1) Inaczej możemy zapisać równanie $\frac{dx}{dt} = f(x)$, skąd formalnie $\frac{dx}{f(x)} = dt$ (ta postać to równość 1-form różniczkowych — często będziemy korzystali z tego sposobu interpretowania równań pochodzącego od Leibniza). Przykładając całki postępujemy dalej jak wyżej.

Możemy też powyżej korzystać od razu z całek oznaczonych. Mamy wtedy $\int_{x_0}^{x} \frac{dy}{f(y)} = \int_{t_0}^{t} 1\,dt$ i dalej postępujemy jak w dowodzie twierdzenia.

2) Łatwo widać, że jeśli $f(x_0) = 0$ w pewnym punkcie x_0, to rozwiązaniem jest $x(t) = x_0$. Jeśli odrzucimy założenie o niezerowaniu się funkcji $f(x)$, to jednoznaczność rozwiązania uzyskamy przy dodatkowym założeniu, że $f(x)$ spełnia lokalny warunek Lipschitza.

Załóżmy zatem, że $f(x_0) = 0$. Mamy dla $x_1 < x_0$ (możemy tak przyjąć bez straty ogólności) równość $\int_{x_1}^{x} \frac{dy}{f(y)} = t + C$. Jeśli $x(t) = F_{x_1}^{-1}(t + C)$ (indeks dolny przy F oznacza warunek początkowy) miałoby przechodzić przez punkt x_0, to ponieważ f nie zmienia znaku na (x_1, x_0) oraz z warunku Lipschitza ze stałą $K > 0$ dla punktów y i x_0 mamy

$$|t_0 + C| = \left| \int_{x_1}^{x_0} \frac{dy}{f(y)} \right| = \int_{x_1}^{x_0} \frac{dy}{|f(y)|} \geqslant \frac{1}{K} \int_{x_1}^{x_0} \frac{dy}{x_0 - y} = +\infty,$$

a t_0 jest skończone. Mamy sprzeczność, zatem rozwiązanie jest jednoznaczne.

Przykład. Rozważmy równanie $\dot{x} = x^{2/3}$. Łatwo sprawdzić, że ma ono co najmniej dwa rozwiązania $x(t) = 0$ oraz $x(t) = (t/3)^3$, które spełniają warunek $x(0) = 0$. Zatem nie mamy jednoznaczności rozwiązania (zauważmy, że funkcja $f(x) = x^{2/3}$ nie spełnia warunku Lipschitza w 0).

3) Równania o zmiennych rozdzielonych.

Twierdzenie 1.2. *Niech $f : (a, b) \to \mathbb{R}$ będzie ciągła, a $g : (c, d) \to \mathbb{R}$ spełnia warunek Lipschitza. Wtedy rozwiązanie równania $\dot{x} = f(t)g(x)$ z warunkiem $x(t_0) = x_0$ istnieje i jest jednoznaczne.*

Dowód. Tezę pokazujemy analogicznie jak dla równań autonomicznych, uwzględniając uwagę 2) powyżej. Rozwiązujemy dzieląc obie strony przez $g(x)$ (o ile $g(x_0) \neq 0$) i całkując:

$$G(x) := \int_{x_0}^{x} \frac{ds}{g(s)} = \int_{t_0}^{t} f(s)\, ds =: F(t).$$

Stąd $x(t) = G^{-1}(F(t) + C)$. □

Uwaga. Rozwiązanie $x(t)$ istnieje dla t takich, dla których $F(t) \in G((c, d))$. Jeśli więc oznaczymy przez $I = (s_1, s_2)$ największy przedział taki, że $t_0 \in I \subset (a, b)$ oraz $F(I) \subset G((c, d))$, to wtedy istnieje rozwiązanie równania na I przechodzące przez (t_0, x_0).

Przykład. Rozważmy równanie $\dot{x} = x^2 \cos t$ z warunkiem $x(0) = \sqrt{2}$. Stosujemy notację z twierdzenia. Mamy dziedziny $(a, b) = (-\infty, \infty)$ oraz $(c, d) = (0, \infty)$ lub $(-\infty, 0)$, przy czym wybieramy $(c, d) = (0, \infty)$ ze względu na warunek początkowy (wiemy z jednoznaczności, że rozwiązanie spełniające podany warunek początkowy nie będzie przyjmowało wartości 0). Liczymy (zgodnie z warunkiem początkowym): $F(t) = \sin t$, $G(x) = -1/x + 1/\sqrt{2}$ oraz $G((0, \infty)) = (-\infty, 1/\sqrt{2})$. Stąd, ponieważ musi być $F(I) \subset G((0, \infty))$, dostajemy warunek $\sin t < 1/\sqrt{2}$ na zbiór I. Mamy zatem rozwiązanie $-\frac{1}{x} + \frac{1}{\sqrt{2}} = \sin t$, czyli $x(t) = \frac{1}{\frac{1}{\sqrt{2}} - \sin t}$, określone dla $t \in I = (-\frac{5\pi}{4}, \frac{\pi}{4})$.

4) Równania jednorodne.

Niech funkcja $u(t)$ będzie taka, że $x(t) = u(t)t$. Wówczas $\dot{x} = \dot{u}t + u$ i podstawiając do równania $\dot{x} = f(x/t)$ otrzymujemy równanie na funkcję $u(t)$ o zmiennych rozdzielonych

$$\dot{u} = \frac{f(u) - u}{t}.$$

5) Równania w postaci różniczek.

W przypadku równania

$$x' = \frac{dx}{dt} = f(t, x)$$

zmienną zależną jest x. Równanie to możemy zapisać jako równość dwóch różniczek (1-form różniczkowych)

$$dx = f(t, x)dt,$$

gdzie wybór zmiennej zależnej może być już dowolny. W szczególności możemy rozważać t jako zmienną zależną, a równanie na funkcję $t(x)$ zapisujemy w tradycyjnej postaci jako

$$t' = \frac{dt}{dx} = \frac{1}{f(t, x)}.$$

Dla zachowania symetrii między zmiennymi rozważamy następujące równanie w postaci różniczek

$$P(x,y)dx + Q(x,y)dy = 0,$$

które można zapisać w postaci tradycyjnej na dwa sposoby

$$\frac{dy}{dx} = -\frac{P(x,y)}{Q(x,y)} \quad \text{lub} \quad \frac{dx}{dy} = -\frac{Q(x,y)}{P(x,y)}.$$

Twierdzenie 1.3. *Niech $R = (a,b) \times (c,d)$, a funkcje $P : R \to \mathbb{R}$ i $Q : R \to \mathbb{R}$ oraz ich pochodne cząstkowe $\frac{\partial P}{\partial y}(x,y)$ i $\frac{\partial Q}{\partial x}(x,y)$, które będziemy oznaczać przez P_y i Q_x odpowiednio, będą ciągłe w R. Jeśli $P_y = Q_x$ oraz jedna z funkcji $P(x,y)$ lub $Q(x,y)$ jest różna od zera w zbiorze R, to przez każdy punkt $(x_0,y_0) \in R$ przechodzi dokładnie jedna krzywa całkowa równania $P(x,y)dx + Q(x,y)dy = 0$.*

Dowód. Przy założeniach twierdzenia istnieje funkcja $H(x,y)$ taka, że $P(x,y) = \frac{\partial H}{\partial x}$ i $Q(x,y) = \frac{\partial H}{\partial y}$ oraz

$$dH(x,y) = \frac{\partial H}{\partial x}dx + \frac{\partial H}{\partial y}dy = P(x,y)dx + Q(x,y)dy,$$

czyli wyrażenie $P(x,y)dx + Q(x,y)dy$ jest *różniczką zupełną* (fakt z analizy). Mamy zatem $dH(x,y) = 0$. Załóżmy bez straty ogólności, że $Q(x,y) \neq 0$. Wówczas możemy zapisać rozwiązanie $y(x)$ w postaci uwikłanej $H(x,y(x)) = C$ (zauważmy, że $\frac{d}{dx}H(x,y(x)) = \frac{\partial H}{\partial x} + \frac{\partial H}{\partial y}\frac{dy}{dx} = 0$ i $\frac{\partial H}{\partial y} \neq 0$). Stałą wyznaczamy z warunku początkowego $C = H(x_0,y_0)$. Spełnione są założenia twierdzenia o funkcji uwikłanej, zatem równanie $H(x,y(x)) = C$ można rozwikłać i istnieje jednoznaczna funkcja $y(x)$ o ciągłej pochodnej. W przypadku, gdy $P(x,y) \neq 0$ możemy analogicznie wyznaczyć jednoznaczne rozwiązanie w postaci funkcji $x(y)$. □

Uwaga.

1) Często się zdarza, że równanie wyjściowe nie jest w postaci różniczki zupełnej. Wtedy można spróbować sprowadzić równanie do tej postaci mnożąc go przez pewną funkcję $\mu(x,y)$, którą nazywamy *czynnikiem całkującym*, dla której będzie już zachodziło $\frac{\partial(\mu P)}{\partial y} = \frac{\partial(\mu Q)}{\partial x}$. Nie ma ogólnej metody pozwalającej na znalezienie czynnika całkującego. Natomiast w szczególnych przypadkach mamy:

- jeśli $(Q_x - P_y)/P = g(y)$, to $\mu = \mu(y) = \exp\left(\int g(y)\,dy\right)$,
- jeśli $(P_y - Q_x)/Q = f(x)$, to $\mu = \mu(x) = \exp\left(\int f(x)\,dx\right)$,
- jeśli istnieją takie funkcje $f(x)$ i $g(y)$, że $P_y - Q_x = Qf(x) - Pg(y)$, to

$$\mu = \mu(x,y) = \exp\Big(\int f(x)\,dx\Big)\exp\Big(\int g(y)\,dy\Big).$$

2) Równanie o zmiennych rozdzielonych $\dot{x} = f(t)g(x)$ jest równaniem w postaci różniczki zupełnej $f(t)dt - \frac{1}{g(x)}dx = 0$, przy czym $H(t,x) = F(t) - G(x)$, gdzie F i G to funkcje pierwotne funkcji $f(t)$ i $1/g(x)$ odpowiednio.

Sposoby rozwiązania.

1) Mamy $\frac{\partial H}{\partial y} = Q$, zatem możemy scałkować po y, co daje $H(x,y) = \int_{y_0}^{y} Q(x,s)\,ds + C(x)$ (funkcja $C(x)$ jest stałą względem y). Ostatnią równość różniczkujemy po x i przyrównujemy do P (bo $\frac{\partial H}{\partial x} = P$). Otrzymane równanie na C' rozwiązujemy całkując od x_0 do x. Wyznaczone $C(x)$ wstawiamy do wzoru na $H(x,y)$.

2) Wiemy z analizy, że $H(x,y) = \int_{(x_0,y_0)}^{(x,y)} Pdx + Qdy$, gdzie całka krzywoliniowa skierowana nie zależy od drogi całkowania (bo pole wektorowe jest gradientem pola skalarnego: $[P,Q] = \nabla H$). Obliczamy tę całkę parametryzując drogę po dwóch odcinkach równoległych do osi układu współrzędnych:

$$\begin{aligned} &\phi_1(t) = (t,y_0), \quad t \in (x_0,x), \qquad \phi_1'(t) = (1,0),\\ &\phi_2(s) = (x,s), \quad s \in (y_0,y), \qquad \phi_2'(s) = (0,1), \end{aligned}$$

co daje

$$H(x,y) = \int_{x_0}^{x} P(t,y_0)\,dt + \int_{y_0}^{y} Q(x,s)\,ds.$$

Ponieważ tutaj $H(x_0,y_0) = 0$, to pozostaje rozwikłać $H(x,y) = 0$ względem x lub y (o ile jest to analitycznie wykonalne).

6) Równania liniowe.

Definicja 1.5. Niech $a : (c,d) \to \mathbb{R}$ i $b : (c,d) \to \mathbb{R}$. Równanie postaci

$$\dot{x} = a(t)x + b(t) \tag{1.5}$$

nazywa się równaniem *liniowym*. Jeśli $b(t) \equiv 0$, to wtedy mówimy o równaniu *liniowym jednorodnym*.

Twierdzenie 1.4. *Jeśli $a : (c,d) \to \mathbb{R}$ i $b : (c,d) \to \mathbb{R}$ są ciągłe, to dla każdych $t_0 \in (c,d)$ i $x_0 \in \mathbb{R}$ istnieje jednoznaczne rozwiązanie $x(t)$ równania (1.5) spełniające $x(t_0) = x_0$. Maksymalnym przedziałem istnienia każdego rozwiązania jest przedział (c,d).*

Dowód. Rozważmy najpierw równanie jednorodne $\dot{x} = a(t)x$. Jest to równanie o zmiennych rozdzielonych. Ma zatem ogólne rozwiązanie postaci

$$x(t) = \bar{C} \exp\Big(\int_{t_0}^{t} a(\tau)\, d\tau \Big).$$

Rozwiązanie równania niejednorodnego znajdujemy metodą *uzmienniania stałej*. W miejsce stałej $\bar{C}$ wstawiamy nieznaną funkcję $C(t)$ i szukamy rozwiązania postaci

$$x(t) = C(t) \exp\Big(\int_{t_0}^{t} a(\tau)\, d\tau \Big).$$

Wstawiamy powyższe $x(t)$ do równania (1.5):

$$\dot{x} = \dot{C} \exp\Big(\int_{t_0}^{t} a(\tau)\, d\tau \Big) + C(t)a(t) \exp\Big(\int_{t_0}^{t} a(\tau)\, d\tau \Big) = a(t)C(t) \exp\Big(\int_{t_0}^{t} a(\tau)\, d\tau \Big) + b(t)$$

i po uproszczeniu otrzymujemy

$$\dot{C} = b(t) \exp\Big(- \int_{t_0}^{t} a(\tau)\, d\tau \Big).$$

Całkując otrzymujemy

$$C(t) = C(t_0) + \int_{t_0}^{t} b(\tau) \exp\Big(- \int_{t_0}^{\tau} a(s)\, ds \Big)\, d\tau.$$

Tak obliczoną funkcję $C(t)$ wstawiamy do funkcji $x(t)$ i wyznaczając z warunku początkowego stałą $C(t_0) = x_0$ otrzymujemy rozwiązanie:

$$x(t) = \left(x_0 + \int_{t_0}^{t} b(\tau) \exp\Big(- \int_{t_0}^{\tau} a(s)\, ds \Big)\, d\tau \right) \exp\Big(\int_{t_0}^{t} a(\tau)\, d\tau \Big).$$

Jednoznaczność rozwiązania wykazujemy zakładając, że mamy dwa rozwiązania x_1 i x_2 równania (1.5) spełniające ten sam warunek początkowy $x_i(t_0) = x_0 = x_2(t_0)$. Wówczas $z(t) = x_1(t) - x_2(t)$ spełnia równanie $\dot{z} = a(t)z$ z warunkiem $z(t_0) = 0$. Jest to równanie o zmiennych rozdzielonych, którego jednoznacznym rozwiązaniem jest $z(t) \equiv 0$, czyli mamy $x_1(t) = x_2(t)$.

Aby pokazać maksymalny przedział istnienia rozwiązania, wystarczy pokazać, że rozwiązanie jest ograniczone w każdym punkcie przedziału (c,d). Niech $x(t)$ będzie rozwiązaniem takim, że $x(t_0) = x_0$. Pokażemy, że jeśli $t_1 \in (c,d)$, to $x(t_1)$ jest ograniczone. Załóżmy bez straty ogólności, że $t_1 > t_0$. Z postaci rozwiązania mamy

$$|x(t_1)| \leqslant \Big(|x_0| + \int_{t_0}^{t_1} |b(\tau)|\, d\tau \Big) \mathrm{e}^{K(t_1 - t_0)},$$

gdzie $K = \sup_{t \in [t_0, t_1]} |a(t)|$. Z ograniczoności b na $[t_0, t_1]$ możemy oszacować

$$|x(t_1)| \leqslant M \mathrm{e}^{K(t_1 - t_0)},$$

czyli rozwiązanie jest ograniczone dla każdego $t_1 \in (c,d)$. □

Uwaga.

1) Dowód twierdzenia podaje konstruktywny sposób rozwiązywania równania liniowego.

2) Rozwiązanie ogólne równania liniowego można interpretować jako sumę ogólnego rozwiązania równania jednorodnego i szczególnego rozwiązania równania niejednorodnego.

Rozdział 2

Podstawowe twierdzenia

Często nie umiemy rozwiązać analitycznie równania, ale możemy badać na przykład, czy to rozwiązanie istnieje (jakie do tego warunki musi spełniać równanie), kiedy to rozwiązanie jest jednoznaczne, jaki jest przedział istnienia tego rozwiązania.

Zaczniemy od definicji zagadnienia początkowego i jego rozwiązania.

Definicja 2.1. Niech $f : D \to \mathbb{R}^n$, $D \subset \mathbb{R} \times \mathbb{R}^n$ niepusty i spójny oraz $(t_0, x_0) \in D$. Powiemy, że $x : I \to \mathbb{R}^n$ jest rozwiązaniem *zagadnienia początkowego* (inaczej *zagadnienia Cauchy'ego*)

$$\begin{cases} x' = f(t,x), & (2.1) \\ x(t_0) = x_0 & (2.2) \end{cases}$$

na przedziale I wtedy i tylko wtedy, gdy
a) x jest rozwiązaniem układu równań (2.1) (w sensie definicji 1.1),
b) $t_0 \in I$ oraz x spełnia (2.2).

Lemat 2.1. *Jeśli $f(t,x)$ jest ciągła, to $x(t)$ jest rozwiązaniem zagadnienia Cauchy'ego* (2.1)-(2.2) *wtedy i tylko wtedy, gdy $x(t)$ jest ciągła na I, $(t, x(t)) \in D$ dla wszystkich $t \in I$, $t_0 \in I$ oraz*

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds. \tag{2.3}$$

Dowód. Oczywisty. □

2.1. Twierdzenia o istnieniu i jednoznaczności

Pierwsze twierdzenie o lokalnym istnieniu rozwiązań zagadnienia początkowego wymaga tylko ciągłości funkcji $f(t,x)$.

Twierdzenie 2.1 (Peano — jednostronne)**.** *Niech $f : D \to \mathbb{R}^n$, $D \subset \mathbb{R} \times \mathbb{R}^n$ niepusty i spójny. Niech*

$$Q = \{(t,x) : t \in [t_0, t_0 + a], \|x - x_0\| \leqslant b\} \subset D$$

dla pewnych $t_0 \in \mathbb{R}$, $x_0 \in \mathbb{R}^n$, $a, b > 0$. Niech ponadto f będzie ciągła w Q. Wtedy zagadnienie Cauchy'ego (2.1)-(2.2) *ma przynajmniej jedno rozwiązanie na przedziale $[t_0, t_0 + \alpha]$, gdzie $\alpha = \min\{a, \frac{b}{M}\}$, $M = \sup_{(t,x)\in Q} \|f(t,x)\|$.*

Dowód. Wykorzystamy w dowodzie „proste" rozwiązania „przybliżone" (tak zwane łamane Eulera) i pokażemy, że są one zbieżne do funkcji, która spełnia nasze równanie. Dowód przeprowadzimy w kilku krokach.

1) Dzielimy odcinek $[t_0, t_0+\alpha]$ na N równych części punktami $t_i = t_0 + ih$, $h = \alpha/N$, $i = 0, 1, \ldots, N$. Określamy rozwiązanie „przybliżone" $y_N(t)$ dla dowolnego $N \geqslant 1$ tak, że w punktach t_i jest

$$\begin{aligned} y_N(t_0) &:= x_0, \\ y_N(t_{i+1}) &:= y_N(t_i) + hf(t_i, y_N(t_i)), \quad i = 0, 1, \ldots, N-1, \end{aligned}$$

a poza tymi punktami interpolujemy liniowo

$$y_N(t) := y_N(t_i) + f(t_i, y_N(t_i))(t - t_i), \quad t \in [t_i, t_{i+1}].$$

Otrzymujemy więc funkcję kawałkami liniową (łamaną).

Spodziewamy się, że zagęszczając podział, czyli zmniejszając h (tzn. zwiększając N), będziemy mieli coraz lepsze przybliżenie rozwiązania, o ile ono istnieje, funkcjami $y_N(t)$.

2) Pokażemy teraz, że funkcje $y_N(t)$ spełniają

$$\|y_N(t) - x_0\| \leqslant b \quad \text{dla } t \in [t_0, t_0 + \alpha]$$

(czyli że są dobrze określone: $(t, y_N(t)) \in Q$), co pociąga za sobą, że są wspólnie ograniczone:

$$\|y_N(t)\| \leqslant \|x_0\| + b \quad \text{dla } t \in [t_0, t_0 + \alpha],\ N = 1, 2, \ldots$$

Zauważmy, że

$$\begin{aligned} y_N(t_j) &= y_N(t_{j-1}) + hf(t_{j-1}, y_N(t_{j-1})) \\ &= y_N(t_{j-2}) + hf(t_{j-2}, y_N(t_{j-2})) + hf(t_{j-1}, y_N(t_{j-1})) = \ldots = \underbrace{y_N(t_0)}_{x_0} + h\sum_{k=0}^{j-1} f(t_k, y_N(t_k)). \end{aligned}$$

Stąd wynika indukcyjny dowód, że $(t_i, y_N(t_i)) \in Q$ dla $i = 0, 1, \ldots, N$. Mamy bowiem $(t_0, y_N(t_0)) \in Q$ z założenia twierdzenia. Zakładając teraz, że $(t_0, y_N(t_0)), \ldots, (t_{j-1}, y_N(t_{j-1})) \in Q$ (założenie indukcyjne), z powyższego wzoru dostajemy

$$\|y_N(t_j) - x_0\| \leqslant h\sum_{k=0}^{j-1} \|f(t_k, y_N(t_k))\| \leqslant h\sum_{k=0}^{j-1} M \leqslant hNM = \alpha M \leqslant b.$$

Mamy więc dla $t \in [t_i, t_{i+1}]$

$$y_N(t) = x_0 + h\sum_{k=0}^{i-1} f(t_k, y_N(t_k)) + f(t_i, y_N(t_i))(t - t_i),$$

a stąd

$$\|y_N(t) - x_0\| \leqslant h\sum_{k=0}^{i-1} \|f(t_k, y_N(t_k))\| + h\|f(t_i, y_N(t_i))\| \leqslant hMN = \alpha M \leqslant b,$$

co pokazuje, że $(t, y_N(t)) \in Q$ dla $t \in [t_0, t_0 + \alpha]$.

3) Następnie pokażemy, że rodzina funkcji $\{y_N\}$ jest jednakowo ciągła, a dokładniej pokażemy, że dla dowolnych $N = 1, 2, \ldots$ oraz $t, s \in [t_0, t_0 + \alpha]$ zachodzi $\|y_N(t) - y_N(s)\| \leqslant M|t - s|$ (rodzina jest lipschitzowska ze stałą M).

Definiujemy funkcje kawałkami stałe $T_N(t)$ i $Z_N(t)$:

$$\begin{aligned} T_N(t) &= t_i, \quad \text{dla } t \in [t_i, t_{i+1}), \\ Z_N(t) &= y_N(t_i) \quad \text{dla } t \in [t_i, t_{i+1}). \end{aligned}$$

Możemy teraz dla $t \in [t_i, t_{i+1})$ zapisać

$$y_N(t) = x_0 + \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} f\big(T_N(\tau), Z_N(\tau)\big)\, d\tau + \int_{i_i}^{t} f\big(T_N(\tau), Z_N(\tau)\big)\, d\tau = x_0 + \int_{t_0}^{t} f\big(T_N(\tau), Z_N(\tau)\big)\, d\tau,$$

bo funkcje podcałkowe są na przedziałach $[t_j, t_{j+1})$ stałe. Wtedy mamy dla $s \leqslant t$

$$\|y_N(t) - y_N(s)\| = \Big\| \int_s^t f\big(T_N(\tau), Z_N(\tau)\big)\, d\tau \Big\| \leqslant \int_s^t \|f\big(T_N(\tau), Z_N(\tau)\big)\|\, d\tau \leqslant \int_s^t M\, d\tau = M(t - s).$$

4) Mamy rodzinę $\{y_N\}$ funkcji wspólnie ograniczonych i jednakowo ciągłych określonych na przedziale zwartym. Zatem z twierdzenia Arzeli-Ascoliego ciąg $\{y_N\}$ zawiera podciąg $\{y_{N_k}\}$ zbieżny jednostajnie do pewnej funkcji y^* na zbiorze $[t_0, t_0 + \alpha]$

$$y_{N_k} \rightrightarrows y^* \quad \text{dla } k \to +\infty$$

(dla prostoty zapisu podciąg ten będziemy oznaczali dalej przez y_N). Funkcja y^* jest ciągła na $[t_0, t_0+\alpha]$ jako granica jednostajnie zbieżnego ciągu funkcji ciągłych.

5) Pokażemy teraz, że y^* jest rozwiązaniem zagadnienia Cauchy'ego (2.1)-(2.2). Oczywiście $y^*(t_0) = x_0$, bo $y_N(t_0) = x_0$ dla wszystkich N. Zatem wystarczy pokazać, że y^* jest rozwiązaniem (2.1). Z ciągłości y^* i lematu 2.1 wynika, że wystarczy w tym celu wykazać równość

$$y^*(t) = x_0 + \int_{t_0}^{t} f(s, y^*(s))\, ds.$$

Wiemy, że $y_N(t) = x_0 + \int_{t_0}^{t} f\big(T_N(s), Z_N(s)\big)\, ds$, a ponieważ $y_{N_k} \rightrightarrows y^*$, to y^* będzie rozwiązaniem, jeśli pokażemy, że

$$\forall t \in [t_0, t_0 + \alpha] \quad \int_{t_0}^{t} f\big(T_N(s), Z_N(s)\big)\, ds \xrightarrow{N\to\infty} \int_{t_0}^{t} f(s, y^*(s))\, ds.$$

Aby tak było, wystarczy jednostajna zbieżność funkcji podcałkowych $f\big(T_N(t), Z_N(t)\big) \rightrightarrows f(t, y^*(t))$ na $[t_0, t_0 + \alpha]$. Zatem wystarczy udowodnić, że

$$\forall \varepsilon > 0 \ \exists N_0 > 0 \ \forall N > N_0 \ \forall t \in [t_0, t_0 + \alpha] \quad \|f\big(T_N(t), Z_N(t)\big) - f(t, y^*(t))\| < \varepsilon.$$

Funkcja f jest jednostajnie ciągła, bo jest ciągła na zwartym zbiorze Q. Zatem dla ustalonego $\varepsilon > 0$ istnieje $\delta > 0$ taka, że jeśli $\max\{|t - s|, \|x - y\|\} < \delta$, to $\|f(t, x) - f(s, y)\| < \varepsilon$. Wystarczy więc, że pokażemy, że można dobrać takie $N_0 > 0$, że dla $N > N_0$ zachodzi

$$|T_N(t) - t| < \delta \quad \text{oraz} \quad \|Z_N(t) - y^*(t)\| < \delta \quad \forall t \in [t_0, t_0 + \alpha].$$

Rzeczywiście, gdy $t \in [t_i, t_{i+1})$, to $|T_N(t) - t| = |t_i - t| \leqslant h = \frac{\alpha}{N}$, skąd dla dostatecznie dużych N mamy $\frac{\alpha}{N} < \delta$. Ponadto

$$\|Z_N(t) - y^*(t)\| \leqslant \|Z_N(t) - y_N(t)\| + \|y_N(t) - y^*(t)\| =: A_1 + A_2$$

Mamy znów dla $t \in [t_i, t_{i+1})$ z lipschitzowskości rodziny $\{y_N\}$ dla dostatecznie dużych N

$$A_1 = \|Z_N(t) - y_N(t)\| = \|y_N(t_i) - y_N(t)\| \leqslant M|t_i - t| \leqslant \frac{M\alpha}{N} < \frac{\delta}{2}.$$

Dalej, z jednostajnej zbieżności y_N wynika, że dla dostatecznie dużych N

$$A_2 = \|y_N(t) - y^*(t)\| < \frac{\delta}{2} \quad \forall t \in [t_0, t_0 + \alpha].$$

Ostatecznie pokazaliśmy, że y^* jest rozwiązaniem. □

Uwaga. Twierdzenie Peano nie wyklucza oczywiście istnienia rozwiązania na odcinku dłuższym niż $[t_0, t_0 + \alpha]$.

Z twierdzenia Peano wynika istnienie rozwiązania na przedziale jednostronnym $[t_0, t_0+\alpha]$. W prosty sposób możemy rozszerzyć ten wynik na przedział obustronny $[t_0 - \alpha, t_0 + \alpha]$.

Wniosek 2.1 (Obustronne twierdzenie Peano)**.** *Jeśli zachodzą założenia twierdzenia Peano dla zbioru*

$$Q := \{(t, x) \in \mathbb{R} \times \mathbb{R}^n : |t - t_0| \leqslant a, \quad \|x - x_0\| \leqslant b\},$$

to rozwiązanie zagadnienia Cauchy'ego istnieje na zbiorze $[t_0 - \alpha, t_0 + \alpha]$, gdzie α jest określone jak w tw. Peano.

Szkic dowodu. Funkcja f jest ciągła dla $t \in [t_0 - a, t_0 + a]$, więc z tw. Peano jednostronnego istnieje rozwiązanie y_F („w przód") na $[t_0, t_0 + \alpha]$. W celu sprawdzenia istnienia rozwiązania y_B („w tył") na $[t_0 - \alpha, t_0]$ „odwrócimy" kierunek czasu.

Na mocy tw. Peano istnieje rozwiązanie $z : [t_0, t_0 + \alpha] \to \mathbb{R}^n$ zagadnienia

$$\begin{cases} z'(t) = g(t, z(t)) := -f(2t_0 - t, z(t)), \\ z(t_0) = x_0. \end{cases}$$

Przyjmijmy $y_B(t) = z(2t_0 - t)$ dla $t \in [t_0 - \alpha, t_0]$. Wówczas $y_B'(t) = f(t, y_B(t))$ i $y_B(t_0) = x_0$. Definiujemy rozwiązanie na $[t_0 - \alpha, t_0 + \alpha]$

$$y(t) := \begin{cases} y_F(t), & t \geqslant t_0, \\ y_B(t), & t \leqslant t_0. \end{cases}$$

Łatwo sprawdzić, że jest ono różniczkowalne w t_0. □

Pokażemy teraz, że można udowodnić jednoznaczność lokalnego rozwiązania, ale przy wzmocnionych założeniach. W dowodzie twierdzenia o jednoznaczności wykorzystamy ważny lemat, zwany również nierównością Gronwalla.

Lemat 2.2 (Gronwalla)**.** *Niech na przedziale $[0, T]$ dane będą funkcje rzeczywiste ciągłe $a(t)$, $b(t)$, $u(t)$, przy czym funkcja $b(t)$ jest nieujemna. Niech $u(t)$ spełnia nierówność*

$$u(t) \leqslant a(t) + \int_0^t b(s)u(s)\,ds, \quad t \in [0, T].$$

Wtedy zachodzi oszacowanie

$$u(t) \leqslant a(t) + \int_0^t a(s)b(s) \exp\left(\int_s^t b(\tau)\,d\tau\right) ds.$$

Jeśli dodatkowo funkcja $a(t)$ jest niemalejąca, to mamy

$$u(t) \leqslant a(t) \exp\left(\int_0^t b(s)\,ds\right).$$

Dowód. Niech $\Phi(t) = \exp\big(-\int_0^t b(s)\,ds\big)$. Mamy wówczas

$$\begin{aligned} \frac{d}{dt}\left[\Phi(t)\int_0^t b(s)u(s)\,ds\right] &= \Phi'(t)\int_0^t b(s)u(s)\,ds + \Phi(t)\frac{d}{dt}\int_0^t b(s)u(s)\,ds \\ &= -b(t)\Phi(t)\int_0^t b(s)u(s)\,ds + \Phi(t)b(t)u(t) \\ &= \Phi(t)b(t)\left[u(t) - \int_0^t b(s)u(s)\,ds\right] \leqslant \Phi(t)a(t)b(t), \end{aligned}$$

gdzie w nierówności skorzystaliśmy z założenia, że $b(t)$ jest nieujemne. Całkujemy powyższą nierówność i otrzymujemy

$$\Phi(t)\int_0^t b(s)u(s)\,ds \leqslant \int_0^t \Phi(s)a(s)b(s)\,ds. \tag{2.4}$$

Dzielimy (2.4) obustronnie przez $\Phi(t)$ i do obu stron dodajemy $a(t)$ otrzymując nierówność

$$a(t) + \int_0^t b(s)u(s)\,ds \leqslant a(t) + \int_0^t a(s)b(s)\exp\left(\int_s^t b(\tau)\,d\tau\right) ds,$$

która daje tezę pierwszej części lematu.

W dowodzie drugiej części tezy wykorzystujemy monotoniczność $a(t)$ szacując prawą stronę (2.4) po podzieleniu przez $\Phi(t)$:

$$\begin{aligned} \int_0^t a(s)b(s)\frac{\Phi(s)}{\Phi(t)}\,ds &\leqslant a(t)\int_0^t b(s)\exp\left(\int_s^t b(\tau)\,d\tau\right) ds \\ &= -a(t)\int_0^t \frac{d}{ds}\exp\left(\int_s^t b(\tau)\,d\tau\right) ds = -a(t) + a(t)\exp\left(\int_0^t b(s)\,ds\right). \end{aligned}$$

Powyższa nierówność wraz z (2.4) dają drugą część tezy lematu. □

Przechodzimy teraz do twierdzenia o lokalnym istnieniu i jednoznaczności.

Twierdzenie 2.2 (Picarda-Lindelöfa)**.** *Niech będą spełnione założenia twierdzenia Peano. Ponadto załóżmy. że funkcja f spełnia jednostajny warunek Lipschitza po drugiej zmiennej.*

$$\exists L > 0 \quad \forall (t,x),(t,y) \in Q \quad \|f(t,x) - f(t,y)\| \leqslant L\|x - y\|.$$

Wtedy istnieje dokładnie jedno rozwiązanie zagadnienia Cauchy'ego (2.1)-(2.2)*, określone na przedziale $[t_0, t_0 + \alpha]$, gdzie α jest określone jak w twierdzeniu Peano.*

Dowód. Wystarczy pokazać jednoznaczność rozwiązania.

Niech $x(t)$ i $y(t)$ będą rozwiązaniami zagadnienia początkowego, czyli

$$x(t) = x_0 + \int_{t_0}^{t} f(\tau, x(\tau))\, d\tau, \quad y(t) = x_0 + \int_{t_0}^{t} f(\tau, y(\tau))\, d\tau.$$

Zatem

$$x(t) - y(t) = \int_{t_0}^{t} [f(\tau, x(\tau)) - f(\tau, y(\tau))]\, d\tau,$$

skąd dla $u(t) := \|x(t) - y(t)\|$ wynika

$$u(t) \leqslant \int_{t_0}^{t} \|f(\tau, x(\tau)) - f(\tau, y(\tau))\|\, d\tau \leqslant L \int_{t_0}^{t} \|x(\tau) - y(\tau)\|\, d\tau = L \int_{t_0}^{t} u(\tau)\, d\tau.$$

Zatem z nierówności Gronwalla (przy $a(t) = 0$ i $b(t) = L$) mamy $u(t) \leqslant 0$, czyli $x \equiv y$. □

Wniosek 2.2. *Prawdziwe jest obustronne (dla $t \in [t_0 - \alpha, t_0 + \alpha]$) twierdzenie Picarda-Lindelöfa.*

Ważnym wnioskiem z przedstawionych twierdzeń jest stabilność ze względu na warunek początkowy.

Wniosek 2.3. *Niech będą spełnione założenia tw. Picarda-Lindelöfa. Niech ponadto $x(t)$ i $y(t)$ będą rozwiązaniami zagadnień początkowych dla warunków początkowych x_0 i y_0 odpowiednio. Wtedy*

$$\|x(t) - y(t)\| \leqslant \|x_0 - y_0\|\, e^{L(t - t_0)}.$$

Dowód. Teza wynika wprost z dowodu twierdzenia Picarda-Lindelöfa oraz lematu Gronwalla. □

Komentarze i przykłady

1. Można podać przykład funkcji $f(t,x)$ ciągłej na $\mathbb{R} \times \mathbb{R}$ (spełniającej założenia tw. Peano) takiej, że dla dowolnego (t_0, x_0) na dowolnym przedziale $[t_0 - \varepsilon, t_0 + \varepsilon]$, $\varepsilon > 0$, istnieje więcej niż jedno rozwiązanie zagadnienia początkowego z warunkiem $x(t_0) = x_0$ (patrz: Hartman, *Ordinary Differential Equations*, SIAM 2002).

2. Rozwiązania mogą nie być określone na całej prostej nawet, jeśli f jest ciągła na $\mathbb{R}$. Rozważmy zagadnienie początkowe $x' = x^2$, $x(0) = x_0 > 0$. Rozwiązaniem jest funkcja $x(t) = x_0/(1 - tx_0)$, która ma asymptotę pionową w $t = 1/x_0$. Co wynika z twierdzeń Peano i Picarda-Lindelöfa? Funkcja $f(t,x) = x^2$ jest ciągła na $\mathbb{R} \times \mathbb{R}$ i lipschitzowska na dowolnym przedziale ograniczonym. Ustalmy zatem liczbę b jak w założeniach o zbiorze Q. Wtedy $M = (b + x_0)^2$ i mamy istnienie rozwiązania na $[t_0, t_0 + \alpha]$, gdzie $\alpha = b/(b + x_0)^2$. Zauważmy, że $b/(b + x_0)^2 \leqslant 1/(4x_0)$, co daje ograniczenie $\alpha < 1/x_0$ niezależne od wyboru b.

3. Bez założenia warunku Lipschitza może nie być jednoznaczności. Rozważmy zagadnienie początkowe $x' = 2\sqrt{x}$, $x(0) = 0$. Oczywistym rozwiązaniem jest funkcja $x(t) = 0$. Zakładając, że $x(t) > 0$ i rozwiązując w standardowy sposób równanie autonomiczne dostajemy kolejne rozwiązanie $x(t) = t^2$, $t \geqslant 0$.

4. Bez założenia ciągłości $f(t,x)$ nie ma różniczkowalności rozwiązań.
   Równanie $x' = \begin{cases} -1 & t < 0 \\ 1 & t \geqslant 0 \end{cases}$ ma rozwiązanie ogólne $x = \begin{cases} -t + C_1 & t < 0 \\ t + C_2 & t \geqslant 0 \end{cases}$ i jedyne rozwiązanie ciągłe jest dla $C_1 = C_2$. Nie jest ono różniczkowalne w $t = 0$.

2.2. Przedłużanie rozwiązań

Twierdzenia Peano i Picarda-Lindelöfa gwarantują tylko lokalne istnienie rozwiązań. Zbadamy teraz, czy można takie rozwiązanie w jakiś sposób „przedłużyć".

Jeśli $x(t)$ jest rozwiązaniem lokalnym na $[t_0, t_0+\alpha]$, to przyjmując $t_1 = t_0+\alpha$ i $x(t_1)$ za nowy warunek początkowy, można rozwiązać równanie na $[t_1, t_1+\alpha_1]$ itd. Analogicznie można konstruować przedłużenia w lewo od t_0. Jak daleko można to kontynuować? Jaki może być maksymalny przedział istnienia rozwiązania?

Definicja 2.2. Rozwiązanie $x(t)$ określone na przedziale $J \subset \mathbb{R}$ nazywa się *rozwiązaniem wysyconym*, jeśli nie istnieje przedłużenie tego rozwiązania na przedział J_1 taki, że J jest jego podzbiorem właściwym. Przedział J nazywa się wtedy *maksymalnym przedziałem istnienia* rozwiązania $x(t)$.

Naszym celem jest zbadanie zachowania rozwiązania wysyconego na brzegu maksymalnego przedziału istnienia rozwiązania, co da nam pewną charakterystykę tego przedziału. Ale najpierw zobaczymy zachowanie rozwiązania na końcach dowolnego przedziału istnienia.

Lemat 2.3. *Niech $x(t)$ będzie rozwiązaniem równania $\dot{x} = f(t,x)$ w przedziale ograniczonym (a_1, a_2) takim, że $(t, x(t)) \in Q$ dla każdego $t \in (a_1, a_2)$, gdzie Q jest zbiorem otwartym w $\mathbb{R}^{n+1}$. Jeśli funkcja $f(t,x)$ jest ciągła i ograniczona na Q, to istnieją granice $x(a_1^+) := \lim_{t\to a_1^+} x(t)$ i $x(a_2^-) := \lim_{t\to a_2^-} x(t)$. Jeśli funkcja $f(t,x)$ jest ciągła w punkcie $(a_1, x(a_1^+))$ lub $(a_2, x(a_2^-))$, to rozwiązanie $x(t)$ może być przedłużone na przedział $[a_1, a_2)$ lub $(a_1, a_2]$.*

Dowód. Funkcja $x(t)$ spełnia dla $a_1 < t_0 \leqslant t < a_2$

$$x(t) = x(t_0) + \int_{t_0}^{t} f(s, x(s))\, ds.$$

Ponieważ $f(t,x)$ jest ograniczona na Q, to dla $M = \sup\limits_{(t,x)\in Q} \|f(t,x)\|$ mamy

$$\|x(t_2) - x(t_1)\| \leqslant \int_{t_1}^{t_2} \|f(s,x(s))\|\, ds \leqslant M(t_2 - t_1), \text{ gdzie } t_1 \leqslant t_2,\ t_1, t_2 \in (a_1, a_2).$$

Zatem jeśli $t_1, t_2 \to a_1^+$, to $x(t_2) - x(t_1) \to 0$, a stąd wynika istnienie granicy $x(a_1^+)$. Podobnie pokazujemy istnienie granicy $x(a_2^-)$.

Jeśli $f(t,x)$ jest ciągła aż do $(a_2, x(a_2^-))$, to $x(a_2)$ jest zdefiniowana wzorem

$$x(a_2) = x(t_0) + \int_{t_0}^{a_2} f(s, x(s))\, ds.$$

Analogicznie dowodzimy przedłużalności na $[a_1, a_2)$. □

Twierdzenie 2.3 (O przedłużaniu)**.** *Niech $f(t,x)$ będzie funkcją ciągłą w zbiorze otwartym $Q \subset \mathbb{R}^{n+1}$ i niech $x(t)$ będzie rozwiązaniem zagadnienia Cauchy'ego* (2.1)-(2.2) *w pewnym przedziale $[t_0, t_0+\alpha]$ takim, że $(t, x(t)) \in Q$ dla każdego $t \in [t_0, t_0+\alpha]$. Wtedy funkcja $x(t)$ może być przedłużona (jako rozwiązanie równania) do rozwiązania* wysyconego *z maksymalnym przedziałem istnienia rozwiązania (ω_1, ω_2). Jeśli ciąg $\{t_n\}$ jest zbieżny do jednego z końców przedziału (ω_1, ω_2), to ciąg $\{(t_n, x(t_n))\}$ jest zbieżny do brzegu Q, o ile Q jest ograniczony. Jeśli zbiór Q jest nieograniczony, to ciąg punktów $(t_n, x(t_n))$ jest nieograniczony dla $t_n \to \omega_1^+$ lub $t_n \to \omega_2^-$.*

Szkic dowodu. 1) Niech $U \subset V \subset \overline{V} \subset Q$, gdzie U zwarty, V otwarty i ograniczony, $\overline{V}$ jest domknięciem V. Jeśli $(t_0, x_0) \in U$, to rozwiązanie $x(t)$ zaczynające się w (t_0, x_0) można przedłużyć na przedział $[t_0, t_1]$ taki, że $(t_1, x(t_1)) \notin U$. Wynika to ze zwartości U oraz iterowania tw. Peano na kolejne przedziały aż do wyjścia z U.
2) Konstruujemy pokrycie zbioru Q ciągiem wstępującym zbiorów Q_j otwartych i ograniczonych takich, że $\overline{Q_j} \subset Q_{j+1}$. Istnieją ciągi $\{t_i\}$ i $\{j_i\}$ takie, że $(t_i, x(t_i)) \in Q_{j_i}$ oraz $(t_i, x(t_i)) \notin Q_{j_{i-1}}$. Ciąg $\{t_i\}$ jest zatem monotoniczny, więc ma granicę. Jeśli jest nieskończona, to koniec dowodu.

3) Załóżmy, że $\{t_i\}$ jest ograniczony z góry i ma granicę $\omega_2 = \lim_{i\to+\infty} t_i$. Jeśli ciąg $(t_i, x(t_i))$ jest nieograniczony, to koniec dowodu. W przeciwnym przypadku z lematu 2.3 $x(t)$ ma granicę $x(\omega_2^-)$, gdzie punkt $(\omega_2, x(\omega_2^-))$ leży na brzegu zbioru Q. Gdyby to był punkt wewnętrzny, to z lematu 2.3 $(\omega_2, x(\omega_2)) \in Q_k$ i można przedłużyć $x(t)$ w prawo na większy przedział — sprzeczność. Podobnie pokazujemy przedłużenie w lewo. □

Uwaga. Z twierdzenia wynika, że rozwiązanie wysycone może istnieć tylko na przedziale otwartym.

Przykład. Rozważmy zagadnienie początkowe $\dot{x} = x^2$, $x(0) = 1$. Rozwiązanie $x(t) = 1/(1-t)$ jest określone (jako funkcja) na $(-\infty, 1)$. Czy jest wysycone? Zbadamy zachowanie rozwiązania przy $t \to 1^-$. Mamy $x(t) \to \infty$, czyli ciąg $(t, x(t))$ jest nieograniczony (zbiór Q ciągłości f jest całym $\mathbb{R}^2$). Zatem rozwiązanie jest wysycone.

Na koniec udowodnimy, że z ograniczoności w $\mathbb{R}^{n+1}$ prawej strony równania wynika istnienie rozwiązania na całej prostej.

Wniosek 2.4. *Niech będą spełnione założenia twierdzenia Peano. Dodatkowo niech $f(t,x)$ ograniczona w $\mathbb{R}^{n+1}$: $\exists K > 0 \quad \|f(t,x)\| \leqslant K$ dla $(t,x) \in \mathbb{R}^{n+1}$. Wtedy istnieje rozwiązanie zagadnienia Cauchy'ego określone dla $t \in \mathbb{R}$.*

Dowód. Z twierdzenia Peano dla pewnych a i b takich, że $|t - t_0| \leqslant a$ i $\|x - x_0\| \leqslant b$ mamy $\alpha = \min(a, b/M)$, gdzie $M = \sup_{(t,x)\in Q} \|f(t,x)\|$. Oczywiście $M \leqslant K$. Zatem możemy wybrać $\overline{\alpha} = \min(a, b/K)$ takie, że $x(t)$ istnieje dla $t \in [t_0 - \overline{\alpha}, t_0 + \overline{\alpha}]$. Ponieważ K nie zależy od żadnego z parametrów a, b, t_0, x_0, możemy mieć dowolnie duże $\overline{\alpha}$ przyjmując odpowiednio duże a i b. Uzyskane rozwiązanie można przedłużyć, startując z punktu $(t_0 + \overline{\alpha}, x(t_0 + \overline{\alpha}))$, aż do $t_0 + 2\overline{\alpha}$. Powtarzając tę procedurę (w obie strony od t_0) dostajemy rozwiązanie na $\mathbb{R}$. □

Rozdział 3

Schematy różnicowe

Wiele razy stajemy przed problemem rozwiązania równania różniczkowego, ale nie umiemy znaleźć analitycznego wzoru na rozwiązanie, albo taki wzór po prostu nie istnieje. Wtedy pozostaje odwołać się do numerycznego przybliżenia rozwiązania.

3.1. Wprowadzenie

Szukamy przybliżonego rozwiązania zagadnienie początkowego (2.1)-(2.2) na $[t_0, T]$ (zakładamy, że rozwiązanie istnieje na tym przedziale). W tym celu będziemy rozważać *schematy różnicowe*, którymi przybliżamy rozwiązanie w skończonej liczbie punktów z przedziału $[t_0, T]$. Przedział ten dzielimy na N równych podprzedziałów o długości $h = \frac{T-t_0}{N}$, za pomocą punktów $t_k = t_0 + kh$, $k = 0, 1, \ldots, N$. Będziemy szukać przybliżenia x_k rozwiązania w chwili t_k, czyli $x_k \approx x(t_k)$. Pierwszy schemat różnicowy pojawił się już w dowodzie tw. Peano.

Metoda Eulera

Najprostszy, a zarazem podstawowy, schemat różnicowy to schemat *Eulera otwarty*, który dla $f_k := f(t_k, x_k)$ przyjmuje postać

$$\begin{cases} x_0 = x(t_0), \\ x_{k+1} = x_k + hf_k, \qquad k = 0, 1, \ldots, N-1. \end{cases}$$

Z dowodu tw. Peano wiemy, że metoda Eulera daje ciąg łamanych Eulera, którego podciąg jest zbieżny do rozwiązania. Przy jakich założeniach mamy zbieżność całego ciągu? Jaka jest szybkość zbieżności? To są kluczowe pytania przy numerycznym rozwiązywaniu równań różniczkowych. Odpowiedzi na nie znajdziemy w dalszej części wykładu. Wkrótce przekonamy się też, że gdy f jest dostatecznie regularna, to metoda Eulera ma błąd rzędu $O(h)$ i zobaczymy, że można konstruować metody z błędem wyższego rzędu.

Uwaga. Zauważmy, że jeśli w metodzie Eulera podstawimy dokładne rozwiązanie $x(t)$, to nie dostaniemy równości, ale otrzymamy resztę $T_{k,h}$, którą nazywamy *błędem obcięcia*:

$$x(t_{k+1}) - x(t_k) - hf(t_k, x(t_k)) = T_{k,h}.$$

Z definicji rozwiązania $f(t_k, x(t_k)) = \dot{x}(t_k)$, zatem ze wzoru Taylora dla $x(t_k + h)$ w punkcie t_k mamy

$$T_{k,h} = x(t_k + h) - x(t_k) - h\dot{x}(t_k) = \frac{1}{2}\ddot{x}(\xi)h^2 = O(h^2), \;\; \xi \in (t_k, t_k + h),$$

o ile x jest dostatecznie regularne. Zwykle będziemy pomijać indeks k w oznaczeniu $T_{k,h}$.

Metody Taylora

Jeśli $x(t)$ jest odpowiednio regularne, to można wziąć więcej wyrazów w rozwinięciu Taylora dla $x(t+h)$. Pochodne wyliczamy korzystając z rozwiązywanego równania. Na przykład pierwszą pochodną mamy z definicji równania $\dot{x}(t) = f(t, x(t))$. Dla drugiej pochodnej mamy

$$\ddot{x} = \frac{d}{dt} f(t,x) = \frac{\partial f}{\partial t}(t,x) + f(t,x)\frac{\partial f}{\partial x}(t,x).$$

Dostajemy zatem rodzinę schematów Taylora

$$\begin{cases} x_0 = x(t_0), \\ x_{k+1} = x_k + hf_k + \dfrac{h^2}{2}\left(\dfrac{\partial f}{\partial t}(t_k,x_k) + f_k\dfrac{\partial f}{\partial x}(t_k,x_k)\right) + \ldots \text{(ewentualnie kolejne wyrazy)}. \end{cases}$$

Schemat Taylora ma błąd obcięcia $T_h = O(h^{p+1})$ (czyli rząd wyrazu, który pomijamy we wzorze Taylora).

Metody wielokrokowe

Schemat Eulera może być podstawą konstrukcji schematów innego typu. Korzystając ze wzoru Taylora „w przód" i „w tył" mamy

$$\begin{aligned} x(t+h) &= x(t) + hf(t,x(t)) + \frac{h^2}{2}\ddot{x}(t) + O(h^3), \\ x(t-h) &= x(t) - hf(t,x(t)) + \frac{h^2}{2}\ddot{x}(t) + O(h^3). \end{aligned}$$

Odejmując stronami otrzymujemy

$$x(t+h) - x(t-h) = 2hf(t,x(t)) + O(h^3).$$

Stąd dostajemy metodę dwukrokową (schemat *midpoint*):

$$\begin{cases} x_0 = x(t_0), \\ x_{k+1} = x_{k-1} + 2hf_k, \qquad k = 1,2,\ldots,N-1 \end{cases}$$

z błędem obcięcia $T_h = O(h^3)$. Jak widać metoda ta potrzebuje do działania wartości x_1, którą można wyliczyć np. metodą Eulera czy Taylora.

Metody Rungego-Kutty

Można konstruować schematy jeszcze innego typu. Rozwiązanie $x(t)$ spełnia równanie całkowe, mamy więc

$$x(t_1) = x_0 + \int_{t_0}^{t_1} f(\tau, x(\tau))\, d\tau.$$

Całkę możemy przybliżyć biorąc wartość f w punkcie $(t_0+t_1)/2$ i mnożąc przez $t_1 - t_0$ (czyli stosując kwadraturę prostokątów — suma Riemanna). Wtedy

$$x(t_1) \approx x_0 + hf(t_0 + \frac{h}{2}, x(t_0 + \frac{h}{2})).$$

Wartości $x(t_0 + h/2)$ nie znamy, dlatego przybliżamy ją metodą Eulera $x(t_0 + h/2) \approx x_0 + \frac{h}{2} f(t_0, x_0)$. Dostajemy w ten sposób jedną z metod Rungego-Kutty (*zmodyfikowany schemat Eulera*):

$$\begin{cases} x_0 = x(t_0), \\ x_{k+1} = x_k + hf\big(t_k + \dfrac{h}{2}, x_k + \dfrac{h}{2} f_k\big), \qquad k = 0,1,\ldots,N-1, \end{cases}$$

z błędem obcięcia

$$T_h = x(t+h) - x(t) - hf\Big(t + \frac{h}{2}, x(t) + \frac{h}{2}f(t,x(t))\Big) = O(h^3),$$

który wynika z zależności

$$x(t+h) = x(t) + h\dot{x} + \frac{h}{2}\ddot{x} + O(h^3)$$

oraz

$$f\Big(t + \frac{h}{2}, x + \frac{h}{2}f(t,x)\Big) = \underbrace{f(t,x)}_{\dot{x}} + \frac{h}{2}\Big(\underbrace{\frac{\partial f}{\partial t} + f(t,x)\frac{\partial f}{\partial x}}_{\frac{d}{dt}f = \ddot{x}}\Big) + O(h^3).$$

Podstawowe definicje

- Schemat *jednokrokowy* korzysta wyłącznie z dwóch przybliżonych wartości rozwiązania w sąsiednich chwilach czasu. Można go zapisać w ogólnej postaci $x_{k+1} = x_k + h\Phi(t_k, x_k, x_{k+1}, h)$. Schematami jednokrokowymi są np. metody Taylora oraz Rungego-Kutty.

- Schemat *wielokrokowy* jest postaci $x_{k+1} = \Psi(t_k, x_{k+1}, x_k, \ldots, x_{k-q+1}, h)$. O takim schemacie będziemy mówić, że jest q-krokowy. Na tym wykładzie będziemy się zajmowali wyłącznie schematami wielokrokowymi *liniowymi*, które zdefiniujemy później. Przykładem jest schemat midpoint. Zauważmy, że schemat jednokrokowy jest szczególnym przypadkiem schematu wielokrokowego, przy czym $\Psi(t_k, x_{k+1}, x_k, h) = x_k + h\Phi(t_k, x_k, x_{k+1}, h)$.

- Schemat jest *samostartujący*, jeśli każde x_k jest określone jednoznacznie przez Φ oraz $x_0 = x(t_0)$. Zatem wszystkie schematy jednokrokowe są samostartujące.

- Schemat jest *jawny* (inaczej *otwarty*), jeśli wyznaczenie x_{k+1} sprowadza się do wyliczenia jawnego wzoru zawierającego wcześniejsze wartości przybliżenia rozwiązania. Podane wyżej przykładowe schematy są jawne.

- Schemat jest *niejawny* (inaczej *zamknięty*), jeśli wyznaczenie x_{k+1} wymaga rozwiązania pewnego równania (zwykle nieliniowego). Na przykład schemat *niejawny Eulera*: $x_{k+1} = x_k + hf_{k+1}$.

- *Błąd obcięcia schematu* (inaczej *lokalny błąd schematu*) oznaczamy symbolem T_h i definiujemy dla dokładnego rozwiązania $x(t)$:

  $$T_h := x(t+h) - \Psi(t, x(t+h), x(t), \ldots, x(t-(q-1)h), h).$$

  Możemy patrzeć na tę wielkość, jak na błąd popełniany w jednym kroku czasowym. W przypadku schematu jednokrokowego definiujemy $T_h := x(t+h) - x(t) - h\Phi(t, x(t), x(t+h), h)$.

- Schemat jest *rzędu co najmniej* p, jeśli dla każdego rozwiązania klasy $C^{p+1}([t_0, T])$ zachodzi oszacowanie

  $$\exists\ c > 0, h_0 > 0 \quad \forall\ h \leqslant h_0 \quad \|T_h\| \leqslant ch^{p+1}.$$

- Schemat jest *rzędu* p, jeśli jest rzędu co najmniej p i nie jest rzędu co najmniej $p+1$.

  Schemat Eulera (jawny i niejawny) jest rzędu $p = 1$, zaś schematy midpoint i zmodyfikowany schemat Eulera są rzędu $p = 2$.

- Schemat jest *zbieżny*, jeśli dla każdego ustalonego punktu $t = t_k = t_0 + kh \in [t_0, T]$ dla $h \to 0$ zachodzi

  $$\|x_k - x(t)\| \to 0$$

  (i) dla każdego rozwiązania $x(\cdot)$ zagadnienia początkowego,
  (ii) dla każdego rozwiązania $\{x_k\}$ schematu różnicowego spełniającego

  $$x_0 = \eta_0(h), \ldots, x_{q-1} = \eta_{q-1}(h), \quad \text{gdzie } \eta_i(h) \underset{h\to 0}{\longrightarrow} x_0.$$

Uwaga. Chodzi tu o zbieżność $x_k \to x(t)$ przy ustalonym $t = t_0 + kh$, gdy $h \to 0$, skąd $k = \frac{t-t_0}{h}$ i $k \to \infty$. Wielkość $\|x_k - x(t)\|$ nazywamy *błędem aproksymacji* schematu. W punkcie (ii) uwzględniamy fakt, że warunek początkowy może nie być dokładnie wyznaczony, a ponadto w przypadku schematów wielokrokowych wartości początkowe trzeba wyznaczyć przy pomocy innych schematów.

- Schemat ma *rząd zbieżności co najmniej p*, jeśli jest zbieżny oraz $\|x_k - x(t)\| = O(h^p)$ dla wszystkich rozwiązań klasy $C^{p+1}([t_0, T])$ oraz wszystkich warunków początkowych takich, że $\eta_i(h) = O(h^p)$, $i = 0, \ldots, q-1$.
- Schemat ma *rząd zbieżności p*, jeśli ma rząd zbieżności co najmniej p i nie ma rzędu zbieżności co najmniej $p+1$.

3.2. Zbieżność schematów jednokrokowych

Zbadamy teraz zbieżność schematów jednokrokowych. Intuicyjnie wydaje się, że lokalne błędy schematu sumują się, gdy przenoszone są do punktu t. Zatem wartość błędu globalnego powinna zależeć od wielkości błędów lokalnych (czyli również od rzędu schematu) oraz wpływu tych błędów na końcową wartość w punkcie t (to zależy w szczególności od stabilności schematu bądź równania, czy zależności od danych początkowych).

W dowodzie twierdzenia o zbieżności schematów jednokrokowych otwartych wykorzystamy następującą dyskretną wersję lematu Gronwalla.

Lemat 3.1 (Dyskretna nierówność Gronwalla)**.** *Niech $a, b > 0$. Jeśli ciąg liczbowy $\{u_k\}$ spełnia*

$$|u_{k+1}| \leqslant a|u_k| + b, \quad \text{dla } k = 0, 1, \ldots,$$

to dla $k = 1, 2, \ldots$ zachodzi

$$|u_k| \leqslant a^k |u_0| + \begin{cases} \dfrac{a^k - 1}{a - 1} b, & \text{gdy } a \neq 1, \\ kb, & \text{gdy } a = 1. \end{cases}$$

Dowód. Oczywisty z zasady indukcji. □

Twierdzenie 3.1 (O zbieżności schematów otwartych)**.** *Załóżmy, że $f : [t_0, T] \times U \to \mathbb{R}^n$, $U \subset \mathbb{R}^n$, jest taka, że $x(t)$ jest jedynym rozwiązaniem na $[t_0, T]$ zagadnienia Cauchy'ego* (2.1)-(2.2)*. Niech $\Phi : [t_0, T] \times U \times (0, h_0] \to \mathbb{R}^n$ dla pewnego h_0. Załóżmy, że schemat otwarty postaci*

$$x_{k+1} = x_k + h\Phi(t_k, x_k, h)$$

jest rzędu p, czyli dla $x \in C^{p+1}([t_0, T])$ $\exists\ c > 0$ $\forall\ h \in (0, h_0]$ $\|T_h\| \leqslant ch^{p+1}$, oraz Φ spełnia warunek Lipschitza

$$\exists\ L_\Phi > 0 \quad \forall\ t \in [t_0, T] \quad \forall\ x, z \in U \quad \forall\ h \in (0, h_0] \quad \|\Phi(t, x, h) - \Phi(t, z, h)\| \leqslant L_\Phi \|x - z\|.$$

Ponadto niech $\|x(t_0) - x_0\| \leqslant c_0 h^p$.

Wtedy, o ile $x_i \in U$, $i = 0, 1, \ldots, k$, dla każdego ustalonego $t = t_k = t_0 + kh \in [t_0, T]$ zachodzi

$$\|x_k - x(t)\| \leqslant h^p \left[\frac{c}{L_\Phi} \left(e^{L_\Phi (t - t_0)} - 1 \right) + c_0 e^{L_\Phi (t - t_0)} \right].$$

Jeśli $L_\Phi = 0$, to $\|x_k - x(t)\| \leqslant h^p \left[c_0 + c(t - t_0) \right]$.

Dowód. Oznaczmy $e_k := x(t_k) - x_k$. Odejmując stronami równości

$$x(t_{k+1}) = x(t_k) + x(t_{k+1}) - x(t_k),$$
$$x_{k+1} = x_k + h\Phi(t_k, x_k, h)$$

otrzymujemy

$$\begin{aligned} e_{k+1} &= e_k + x(t_{k+1}) - x(t_k) - h\Phi(t_k, x_k, h) \\ &= e_k + \underbrace{x(t_{k+1}) - x(t_k) - h\Phi(t_k, x(t_k), h)}_{\text{błąd obcięcia}} + h\left[\Phi(t_k, x(t_k), h) - \Phi(t_k, x_k, h)\right] \\ &= e_k + T_h + h\left[\Phi(t_k, x(t_k), h) - \Phi(t_k, x_k, h)\right]. \end{aligned}$$

Stąd

$$\|e_{k+1}\| \leqslant \|e_k\| + ch^{p+1} + hL_\Phi \| \underbrace{x(t_k) - x_k}_{e_k} \| = (1 + hL_\Phi)\|e_k\| + ch^{p+1}.$$

Na mocy dyskretnego lematu Gronwalla dla $a = (1 + hL_\Phi)$ i $b = ch^{p+1}$ mamy

$$\|e_k\| \leqslant \mathrm{e}^{khL_\Phi}\|e_0\| + \begin{cases} \dfrac{\mathrm{e}^{khL_\Phi} - 1}{hL_\Phi} ch^{p+1}, & L_\Phi > 0, \\ ckh^{p+1}, & L_\Phi = 0, \end{cases}$$

gdzie skorzystaliśmy z nierówności $1 + hL_\Phi \leqslant \mathrm{e}^{hL_\Phi}$. Podstawiając powyżej $kh = t_k - t_0 = t - t_0$ oraz $\|e_0\| \leqslant c_0 h^p$, otrzymujemy tezę. □

Wniosek 3.1. *Przy założeniach twierdzenia mamy:*

$$\max_{k=0,\ldots,N} \|x_k - x(t_k)\| \leqslant h^p \left[\frac{c}{L_\Phi} \left(e^{L_\Phi (T - t_0)} - 1 \right) + c_0 e^{L_\Phi (T - t_0)} \right] = Ch^p.$$

Uwaga. Zauważmy, że stała C we wniosku zależy od $T - t_0$ wykładniczo. Zatem oszacowanie błędu szybko rośnie wraz z długością odcinka czasu, na którym szukamy rozwiązania (nie oznacza to, że *zawsze* rzeczywisty błąd też rośnie).

Uwaga. Nie warto stosować schematów wysokiego rzędu, gdy rozwiązanie analityczne nie jest dostatecznie gładkie (potwierdza to praktyka).

Uwaga. Schemat zamknięty $x_{k+1} = x_k + h\Phi_z(t_k, x_k, x_{k+1}, h)$, aby był stosowalny, musi dać się rozwikłać przynajmniej lokalnie. To znaczy, że możemy zastąpić go (lokalnie) równoważnym schematem jawnym $x_{k+1} = x_k + h\Phi_o(t_k, x_k, h)$, do którego, jeśli spełnia założenia, stosujemy twierdzenie.

Jawna metoda Eulera

$$x_{k+1} = x_k + hf(t_k, x_k)$$

Lokalny błąd schematu policzyliśmy już wcześniej: $\|T_h\| \leqslant ch^2$, o ile $x \in C^2([t_0, T])$, czyli schemat jest rzędu 1. Ponadto, jeśli f spełnia warunek Lipschitza ze stałą L ze względu na drugą zmienną, to Φ również spełnia ten warunek i wtedy

$$\|x_k - x(t)\| \leqslant h\frac{c}{L}\left(\mathrm{e}^{L(t-t_0)} - 1\right) = \mathrm{O}(h).$$

Niejawna metoda Eulera

$$x_{k+1} = x_k + hf(t_{k+1}, x_{k+1})$$

Podobnie jak wyżej pokazujemy, że $\|T_h\| \leqslant ch^2$, o ile $x \in C^2([t_0, T])$ (tym razem rozwijamy $x(t)$ w punkcie $t + h$). Chcemy „rozwikłać" ten schemat (przynajmniej lokalnie) i zapisać w postaci jawnej $x_{k+1} = x_k + h\Phi_o(t_k, x_k, h)$ dla pewnej funkcji Φ_o. Określamy zatem

$$\Phi_o(t_k, x_k, h) = f(t_k + h, x_{k+1}) = f(t_k + h, x_k + h\Phi_o(t_k, x_k, h)),$$

czyli mamy niejawną definicję funkcji Φ_o, o której musimy pokazać, że spełnia warunek Lipschitza względem drugiej zmiennej. Zakładamy, że f spełnia warunek Lipschitza ze stałą L ze względu na drugą zmienną. Mamy wtedy

$$\Phi(t, x, h) = f(t + h, x + h\Phi(t, x, h))$$

oraz

$$\begin{aligned}\|\Phi(t,x,h)-\Phi(t,y,h)\| &= \|f(t+h,x+h\Phi(t,x,h))-f(t+h,y+h\Phi(t,y,h))\| \\ &\leqslant L\|x-y+h\left[\Phi(t,x,h)-\Phi(t,y,h)\right]\| \\ &\leqslant L\|x-y\|+hL\|\Phi(t,x,h)-\Phi(t,y,h)\|.\end{aligned}$$

Stąd

$$\|\Phi(t,x,h)-\Phi(t,y,h)\| \leqslant \frac{L}{1-hL}\|x-y\|, \quad \text{o ile } h<\frac{1}{L}.$$

Zatem z twierdzenia o zbieżności mamy $\|x_k - x(t)\| = \mathrm{O}(h)$.

Uwaga. Zwróćmy uwagę, że otrzymane ograniczenie na krok h wynika z zastosowania twierdzenia o punkcie stałym (tak tutaj lokalnie rozwikłaliśmy schemat zamknięty). Jeśli stała Lipschitza funkcji f nie jest zbyt wielka, to ten warunek nie jest bardzo ograniczający. W przypadku dużych wartości L lepiej stosować do rozwikłania metodę Newtona lub podobną (mimo większego kosztu obliczeniowego).

Uwaga. Mimo że schematy niejawne są droższe w zastosowaniu: w każdym kroku trzeba rozwiązać nieliniowe (zwykle) równanie, mają istotne zastosowanie w praktyce. Schematy zamknięte pozwalają często na użycie dłuższego kroku h niż schematy jawne przy tym samym poziomie błędu, zwłaszcza w przypadku rozwiązywania problemów *sztywnych* (o tym będzie pod koniec wykładu), często pojawiających się w praktyce.

3.3. Metody typu predyktor-korektor

Problemem schematów niejawnych jest duży koszt obliczeniowy związany z dokładnym rozwiązaniem równania nieliniowego. Można tego uniknąć szukając rozwiązania przybliżonego. W ten sposób konstruuje się metody typu *predyktor-korektor* (ang. *predictor-corrector*), w których, jako pierwsze przybliżenie x_{k+1} w schemacie zamkniętym, stosujemy wartość uzyskaną ze schematu otwartego tego samego (na ogół) rzędu. Zwróćmy uwagę, że te schematy są jawne.

Mamy zatem predyktor:

$$\text{P:} \qquad x_{k+1} = x_k + h\Phi_o(t_k, x_k, h)$$

i to przybliżenie wstawiamy do prawej strony wzoru korektora (schematu zamkniętego):

$$\text{C:} \qquad x_{k+1} = x_k + h\Phi_z(t_k, x_k, x_{k+1}, h).$$

Schemat korektora C można kilka razy iterować dla zwiększenia dokładności, ale często się z tego rezygnuje.

Przykład. Użyjemy pary schematów Eulera: otwartego i zamkniętego:

$$\begin{aligned}&\text{P:} \qquad x_{k+1} = x_k + hf(t_k, x_k),\\ &\text{C:} \qquad x_{k+1} = x_k + hf(t_{k+1}, x_{k+1}).\end{aligned}$$

W ten sposób otrzymujemy schemat:

$$x_{k+1} = x_k + hf(t_{k+1}, x_k + hf(t_k, x_k)).$$

Przykład. Dla pary schematów: otwartego Eulera i schematu *trapezów*:

$$\begin{aligned}&\text{P:} \qquad x_{k+1} = x_k + hf(t_k, x_k),\\ &\text{C:} \qquad x_{k+1} = x_k + \frac{h}{2}\left[f(t_k, x_k) + f(t_{k+1}, x_{k+1})\right],\end{aligned}$$

otrzymujemy schemat *Heuna*:

$$x_{k+1} = x_k + \frac{h}{2}\left[f(t_k, x_k) + f(t_{k+1}, x_k + hf(t_k, x_k))\right].$$

3.4. Metody Rungego-Kutty

Metody Rungego-Kutty tworzą wielką rodzinę metod jednokrokowych.

Definicja 3.1. Schemat postaci

$$\begin{aligned} K_1 &= f(t_n, x_n), \\ K_2 &= f(t_n + c_2 h, x_n + h a_{21} K_1), \\ K_3 &= f(t_n + c_3 h, x_n + h(a_{31} K_1 + a_{32} K_2)), \\ &\vdots \\ K_r &= f(t_n + c_r h, x_n + h \sum_{j=1}^{r-1} a_{rj} K_j), \\ x_{n+1} &= x_n + h \sum_{j=1}^{r} b_j K_j \end{aligned}$$

dla $r \geqslant 1$ oraz $a_{ij} \in \mathbb{R}$, $i = 2, \ldots, r$, $j = 1, \ldots, r-1$, $b_i \in \mathbb{R}$, $i = 1, \ldots, r$, $c_i \in \mathbb{R}$, $i = 2, \ldots, r$, nazywa się *r-poziomowym schematem jawnym Rungego-Kutty.*

Definicja 3.2. Schemat postaci

$$\begin{aligned} K_i &= f(t_n + c_i h, x_n + h \sum_{j=1}^{r} a_{ij} K_j), \quad i = 1, \ldots, r, \\ x_{n+1} &= x_n + h \sum_{j=1}^{r} b_j K_j \end{aligned}$$

dla $r \geqslant 1$ oraz $a_{ij} \in \mathbb{R}$, $i, j = 1, \ldots, r$, $b_i, c_i \in \mathbb{R}$, $i = 1, \ldots, r$, nazywa się *r-poziomowym schematem niejawnym Rungego-Kutty.*

Zwykle metody Rungego-Kutty reprezentuje się za pomocą *tabelki Butchera* (z lewej dla schematu jawnego, a z prawej dla niejawnego), w której wypisujemy wszystkie współczynniki schematu:

$$\begin{array}{c|ccccc} 0 & & & & & \\ c_2 & a_{21} & & & & \\ c_3 & a_{31} & a_{32} & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ c_r & a_{r1} & a_{r2} & \cdots & a_{r,r-1} & \\ \hline & b_1 & b_2 & \cdots & b_{r-1} & b_r \end{array} \qquad \begin{array}{c|ccc} c_1 & a_{11} & \cdots & a_{1r} \\ c_2 & a_{21} & \cdots & a_{2r} \\ c_3 & a_{31} & \cdots & a_{3r} \\ \vdots & \vdots & & \vdots \\ c_r & a_{r1} & \cdots & a_{rr} \\ \hline & b_1 & \cdots & b_r \end{array}$$

Lemat 3.2. *Schematy Rungego-Kutty spełniają warunek Lipschitza, o ile funkcja $f(t, x)$ spełnia warunek Lipschitza.*

Dowód. Dowód przeprowadzimy dla schematów otwartych (dla zamkniętych dowodzimy podobnie). Ogólną metodę jawną Rungego-Kutty możemy zapisać w postaci

$$x_{n+1} = x_n + h\Phi(t_n, x_n, h),$$

gdzie $\Phi(t_n, x_n, h) = \sum_{j=1}^{r} b_j K_j$. Możemy również napisać $\Phi(t_n, y_n, h) = \sum_{j=1}^{r} b_j M_j$, dla $M_j = f(t_n + c_j h, y_n + h \sum_{l=1}^{j-1} a_{jl} M_l)$. Mamy wtedy

$$\|\Phi(t_n, x_n, h) - \Phi(t_n, y_n, h)\| \leqslant \sum_{j=1}^{r} |b_j| \|K_j - M_j\|.$$

Ponadto, dla $t = t_n + c_j h$ i stałej Lipschitza L funkcji f

$$\begin{aligned}\|K_j - M_j\| &= \Big\| f(t, x_n + h\sum_{l=1}^{j-1} a_{jl}K_l) - f(t, y_n + h\sum_{l=1}^{j-1} a_{jl}M_l)\Big\| \\ &\leqslant L\|x_n - y_n\| + hL\sum_{l=1}^{j-1}|a_{jl}|\|K_l - M_l\|.\end{aligned}$$

Ponieważ $\|K_1 - M_1\| \leqslant L\|x_n - y_n\|$, to indukcyjnie pokazujemy, że istnieje stała Λ_j taka, że $\|K_j - M_j\| \leqslant \Lambda_j\|x_n - y_n\|$. Stąd

$$\|\Phi(t_n, x_n, h) - \Phi(t_n, y_n, h)\| \leqslant \underbrace{\Big(\sum_{j=1}^{r}|b_j|\Lambda_j\Big)}_{L_\Phi}\|x_n - y_n\|. \qquad \square$$

Lemat 3.3. *Metoda Rungego-Kutty jest rzędu co najmniej* 1 *wtedy i tylko wtedy, gdy* $\sum_{i=1}^{r} b_i = 1$.

Dowód. Ze wzoru Taylora możemy napisać $K_i = f(t + c_h, x + h\sum_j a_{ij}K_j) = f(t,x) + \mathrm{O}(h)$. Zatem

$$T_h = x(t+h) - x(t) - h\sum_i b_i K_i = \dot{x}(t)h - \underbrace{f(t,x)}_{\dot{x}(t)}h\Big(\sum_i b_i\Big) + \mathrm{O}(h^2). \qquad \square$$

Uwaga. W metodach Rungego-Kutty w naturalny sposób otrzymujemy warunek na współczynniki c_i. Równanie $\dot{x} = f(t,x)$ można zapisać w postaci $\dot{u} = F(u)$, gdy przyjmiemy $u = [t,x]^\mathsf{T}$ oraz $F(u) = [1, f(t,x)]^\mathsf{T}$. Spodziewamy się, że zastosowanie metody Rungego-Kutty do każdego z tych równań da takie samo rozwiązanie numeryczne. To daje nam warunek $c_i = \sum_{j=1}^{r} a_{ij}$ (lub $c_i = \sum_{j=1}^{i-1} a_{ij}$ w przypadku schematu otwartego), który będziemy zakładać.

Przykłady metod jawnych Rungego-Kutty

- Najprostszą metodą jawną Rungego-Kutty jest metoda Eulera, dla której mamy tabelkę:

0	
	1

- Zmodyfikowana metoda Eulera rzędu 2 (dwupoziomowa metoda RK):

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

- Metoda Heuna rzędu 3:

0			
$\frac{1}{3}$	$\frac{1}{3}$		
$\frac{2}{3}$	0	$\frac{2}{3}$	
	$\frac{1}{4}$	0	$\frac{3}{4}$

- Klasyczna metoda 4-poziomowa rzędu 4:

0				
$\frac{1}{2}$	$\frac{1}{2}$			
$\frac{1}{2}$	0	$\frac{1}{2}$		
1	0	0	1	
	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$

- „Schemat 3/8" rzędu 4:

0				
$\frac{1}{3}$	$\frac{1}{3}$			
$\frac{2}{3}$	$-\frac{1}{3}$	1		
1	1	-1	1	
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Zwykle chcemy korzystać z metod o jak najwyższym rzędzie (odpowiadającym gładkości rozwiązania). Powyższe przykłady metod wskazują, że zwiększanie rzędu wiąże się z większą złożonością i kosztem obliczeniowym. Przedstawimy teraz bez dowodów twierdzenia podające najważniejsze zależności rzędu metod jawnych Rungego-Kutty od liczby ich poziomów.

Twierdzenie 3.2 (I bariera Butchera)**.** *Dla $r \geqslant 5$ nie istnieje jawna r-poziomowa metoda Rungego-Kutty rzędu r.*

Twierdzenie 3.3 (II bariera Butchera)**.** *Dla $r \geqslant 7$ nie istnieje jawna $(r+1)$-poziomowa metoda Rungego-Kutty rzędu r.*

Twierdzenie 3.4 (III bariera Butchera)**.** *Dla $r \geqslant 8$ nie istnieje jawna $(r+2)$-poziomowa metoda Rungego-Kutty rzędu r.*

Przykłady metod niejawnych Rungego-Kutty

- Najprostsza niejawna metoda Rungego-Kutty to metoda Eulera z tabelką: $\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}$
- Metoda *midpoint* rzędu 2: $\begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline & 1 \end{array}$
- Metoda Hammera-Hollingswortha rzędu 3: $\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array}$
- Metoda Hammera-Hollingswortha rzędu 4: $\begin{array}{c|cc} \frac{1}{2}-\frac{\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4}-\frac{\sqrt{3}}{6} \\ \frac{1}{2}+\frac{\sqrt{3}}{6} & \frac{1}{4}+\frac{\sqrt{3}}{6} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$

Uwaga. Istotną wadą niejawnej r-poziomowej metody Rungego-Kutty jest konieczność wyznaczenia na każdym kroku współczynników K_i jako rozwiązania układu równań nieliniowych wymiaru rd, gdzie d jest wymiarem zadania.

Uwaga. Dla dowolnej liczby poziomów r istnieje r-poziomowa niejawna metoda Rungego-Kutty rzędu $2r$. Jest to duża zaleta metod niejawnych, ale problemem jest, że często wyższy rząd (a tym samym możliwość używania dłuższego kroku) nie przeważa nad wysokim kosztem iteracji.

Uwaga. Ogólną wadą metod Rungego-Kutty jest konieczność r-krotnego obliczenia wartości funkcji f w jednym kroku metody, co czasem może być bardzo kosztowne.

3.5. Schematy liniowe wielokrokowe

Definicja 3.3. Schemat *liniowy q-krokowy* ma postać

$$\sum_{j=0}^{q} \alpha_j x_{n+j} = h \sum_{j=0}^{q} \beta_j f_{n+j}, \tag{3.1}$$

gdzie $q \geqslant 1$, $\alpha_j, \beta_j \in \mathbb{R}$, $\alpha_q \neq 0$, $|\alpha_0| + |\beta_0| > 0$ oraz $x_{n+j} \approx x(t_n + jh)$ i $f_{n+j} = f(t_{n+j}, x_{n+j})$. Jeśli $\beta_q = 0$, to schemat jest *jawny*, w przeciwnym przypadku jest *niejawny.*

Analogicznie jak w przypadku schematów jednokrokowych określamy lokalny błąd aproksymacji. Podstawiamy rozwiązanie dokładne $x(t)$ do schematu (3.1) i rozwijając ze wzoru Taylora $x(t_{n+j})$ w punkcie t_n otrzymujemy

$$\sum_{j=0}^{q} \alpha_j x(t_n + jh) - h \sum_{j=0}^{q} \beta_j x'(t_n + jh) = \sum_{j=0}^{p} c_j h^j x^{(j)}(t_n) + \mathrm{O}(h^{p+1}), \tag{3.2}$$

a stad można łatwo wyznaczać rząd schematu.

Wniosek 3.2. *Schemat* (3.1) *jest rzędu* $p \geqslant 0$ *wtedy i tylko wtedy, gdy we wzorze* (3.2) *zachodzi* $c_0 = c_1 = \ldots = c_p = 0$ *i* $c_{p+1} \neq 0$*, gdzie*

$$\begin{aligned} c_0 &= \sum_{j=0}^{q} \alpha_j, \\ c_1 &= \sum_{j=0}^{q} j\alpha_j - \sum_{j=0}^{q} \beta_j, \\ &\vdots \\ c_p &= \frac{1}{p!}\left(\sum_{j=0}^{q} j^p \alpha_j - p \sum_{j=0}^{q} j^{p-1} \beta_j \right). \end{aligned}$$

Można więc powiedzieć, że uzyskaliśmy „maszynkę" do sprawdzania rzędu schematu.

Definicja 3.4. Schemat (3.1) jest *zgodny* (inaczej: *konsystentny*), jeśli jest rzędu co najmniej 1.

Wniosek 3.3. *W schemacie zgodnym mamy* $c_0 = c_1 = 0$.

Definicja 3.5. *Wielomian charakterystyczny* równania (3.1) to $\rho(\lambda) := \sum_{j=0}^{q} \alpha_j \lambda^j$, natomiast *wielomian tworzący* to $\sigma(\lambda) := \sum_{j=0}^{q} \beta_j \lambda^j$.

Wniosek 3.4. *W schemacie zgodnym mamy* $\rho(1) = 0$ *oraz* $\rho'(1) = \sigma(1)$.

Definicja 3.6. Schemat (3.1) jest *stabilny*, jeśli wszystkie pierwiastki wielomianu ρ leżą w kole jednostkowym $\{z \in \mathbb{C} : |z| \leqslant 1\}$, a te o module równym 1, są jednokrotne.

Definicja 3.7. Schemat (3.1) jest *silnie stabilny*, jeśli jest stabilny i jedynym pierwiastkiem $\rho(\lambda)$ o module 1 jest $\lambda = 1$.

Uwaga. Własność stabilności, jak się niedługo okaże, jest konieczna do zbieżności, natomiast silna stabilność eliminuje możliwe oscylacje rozwiązania.

Przykład. Schematy *Adamsa* to schematy q-krokowe postaci

$$x_{n+q} - x_{n+q-1} = h \sum_{j=0}^{q} \beta_j f_{n+j}.$$

Są one silnie stabilne, bo $\rho(\lambda) = \lambda^{q-1}(\lambda - 1)$. Przykładowe schematy w tej grupie:

q	β_0	β_1	β_2	β_3
1	1			
2	$-\frac{1}{2}$	$\frac{3}{2}$		
3	$\frac{5}{12}$	$-\frac{16}{12}$	$\frac{23}{12}$	
4	$-\frac{9}{24}$	$\frac{37}{24}$	$-\frac{59}{24}$	$\frac{55}{24}$

jawne
(Adamsa-Bashfortha)

q	β_0	β_1	β_2	β_3	β_4
1	0	1			
1	$\frac{1}{2}$	$\frac{1}{2}$			
2	$-\frac{1}{12}$	$\frac{8}{12}$	$\frac{5}{12}$		
3	$\frac{1}{24}$	$-\frac{5}{24}$	$\frac{19}{24}$	$\frac{9}{24}$	
4	$-\frac{19}{720}$	$\frac{106}{720}$	$-\frac{264}{720}$	$\frac{646}{720}$	$\frac{251}{720}$

niejawne
(Adamsa-Moultona)

Zwróćmy uwagę, że jawny schemat Adamsa jednokrokowy ($q = 1$) to schemat Eulera, natomiast niejawne schematy Adamsa jednokrokowe to kolejno schematy niejawny Eulera i trapezów.

Uwaga. Schematy Adamsa powstają przez interpolację funkcji f w punktach $\{(t_i, x_i)\}_{i=n,\ldots,n+q-1,n+q}$, a następnie wykorzystanie wzoru $x_{n+q} = x_{n+q-1} + \int_{t_{n+q-1}}^{t_{n+q}} w(t)\,dt$, gdzie w jest wielomianem interpolacyjnym.

Przykład. Inny sposób konstrukcji schematów wielokrokowych, to metody bazujące na różniczkowaniu. Interpolujemy rozwiązanie x w punktach $\{(t_i, x_i)\}_{i=n,\ldots,n+q}$ i żądamy, aby wielomian interpolacyjny w spełniał $w'(t_{n+q}) = f(t_{n+q}, x_{n+q})$. W ten sposób dostajemy metody zwane metodami *różniczkowania wstecz* (z ang. BDF). Jako przykłady można podać metodę niejawną Eulera (którą określa się jako BDF1) oraz schemat BDF2:

$$x_{n+2} - \frac{4}{3}x_{n+1} + \frac{1}{3}x_n = \frac{2}{3}hf_{n+2}.$$

Metody te są używane przede wszystkim do rozwiązywania zadań sztywnych (wrócimy do tego pod koniec wykładu).

Przykład. Grupa schematów *Milnego* pochodzi z interpolacji f i całkowania na przedziale $[t_{n-1}, t_{n+1}]$. Jednym z przykładów jest przedstawiony wcześniej schemat midpoint $x_{n+1} = x_{n-1} + 2hf_n$ (nie jest silnie stabilny) oraz schemat *Milne-Simpsona*:

$$x_{n+1} = x_{n-1} + \frac{h}{3}(f_{n-1} + 4f_n + f_{n+1}).$$

Poniższe twierdzenie podaje możliwy maksymalny rząd metod wielokrokowych.

Twierdzenie 3.5 (I bariera Dahlquista)**.** *Rząd p metody q-krokowej, która jest stabilna, spełnia nierówności:*

$$p \leqslant \begin{cases} q+2 & \text{jeśli } q \text{ parzyste,} \\ q+1 & \text{jeśli } q \text{ nieparzyste,} \\ q & \text{jeśli } \frac{\beta_q}{\alpha_q} \leqslant 0 \text{ (w szczególności dla metod jawnych).} \end{cases}$$

Lemat 3.4. *1) Dla q-krokowego schematu jawnego Adamsa jest $p = q$, zaś dla niejawnego $p = q + 1$.*
2) Dla schematów różniczkowania wstecz mamy $p = q$.
3) Dla schematu midpoint mamy $p = 2$, zaś dla Milne-Simpsona $p = 4$.

Zbieżność metody wielokrokowej zależy nie tylko od zgodności schematu, ale i od jego stabilności (zwróćmy uwagę, że metody jednokrokowe są zawsze stabilne).

Twierdzenie 3.6. *Jeśli schemat* (3.1) *jest zbieżny, to jest stabilny i zgodny.*

Dowód. Stosujemy schemat (3.1) do problemu $x'(t) = 0$, $x(0) = 0$ i otrzymujemy równanie różnicowe $\sum_{j=0}^{q} \alpha_j x_{n+j} = 0$, którego rozwiązanie jest postaci $x_n = c_1\lambda_1^n + \ldots + c_q\lambda_q^n$, o ile wielomian charakterystyczny nie ma pierwiastków wielokrotnych (o sposobie rozwiązywania tego typu równań będzie na ćwiczeniach). Zauważmy, że jeśli istnieje pierwiastek λ taki, że $|\lambda| > 1$, to wtedy $x_n \to \pm\infty$, gdy $h \to 0$ (czyli $n \to \infty$), co przeczy zbieżności. Oznacza to, że wszystkie pierwiastki muszą spełniać $|\lambda| \leqslant 1$. Co więcej, jeśli istnieje pierwiastek o module 1 i jest wielokrotny, to pojawi się w rozwiązaniu składnik postaci $n\lambda_j^n$ i znowu $x_n \to \pm\infty$, gdy $n \to \infty$. Zatem schemat musi być stabilny, aby był zbieżny.

Rozważmy teraz problem $x'(t) = 0$, $x(0) = 1$, którego dokładnym rozwiązaniem jest $x(t) = 1$. Zastosowanie schematu (3.1) ponownie daje równanie $\sum_{j=0}^{q} \alpha_j x_{n+j} = 0$. Gdy $h \to 0$, zbieżność implikuje $x_{n+j} \to 1$, a stąd $\rho(1) = 0$. Dalej rozważmy zadanie $x'(t) = 1$, $x(0) = 0$ z rozwiązaniem $x(t) = t$. Wiemy już, że $\rho(1) = 0$, i łatwo sprawdzamy, że $x_n = nhK = tK$ jest rozwiązaniem numerycznym, w którym $K = \sigma(1)/\rho'(1)$. Z założenia o zbieżności schematu dostajemy $K = 1$, czyli schemat musi być zgodny. □

Uwaga. Pojęcie *silnej stabilności* jest związane z jakością rozwiązania i powstawaniem oscylacji. Jeśli istnieje pierwiastek $\lambda \neq 1$ taki, że $|\lambda| = 1$, to w rozwiązaniu w praktyce pojawi się składowa oscylująca. Jeśli zaś tylko $\lambda = 1$ ma moduł 1, to wszystkie składowe zanikają poza składową stałą przy $t \to \infty$.

Prawdziwe jest też twierdzenie odwrotne, które podajemy w wersji z zadanym rzędem schematu $p \geqslant 1$ (oczywiście schemat jest wtedy zgodny).

Twierdzenie 3.7. *Niech $f(t,x)$ będzie ciągła i lipschitzowska względem drugiej zmiennej na zbiorze $[t_0,t_0+a]\times U$ i niech $x(t)$ będzie rozwiązaniem klasy $C^{p+1}([t_0,t_0+a])$ zagadnienia Cauchy'ego (2.1)-(2.2) na $[t_0,t_0+a]$, a $\{x_n\}$ będzie rozwiązaniem schematu (3.1) takim, że $x_n\in U$ dla wszystkich n. Jeśli schemat (3.1) ma rząd $p\geqslant 1$, jest stabilny oraz punkty startowe spełniają*

$$\|x_j - x(t_j)\| \leqslant ch^p, \quad j=0,1,\dots,q-1,$$

to dla każdego $t=t_N=t_0+Nh\in[t_0,t_0+a]$, $h=\frac{t-t_0}{N}$ zachodzi: $\|x_N-x(t)\|\leqslant ch^p$.

Szkic dowodu. Dowód polega na odpowiednim sprowadzeniu schematu (3.1) do metody jednokrokowej, ale wyższego wymiaru.

Pokażemy teraz zarys dowodu w przypadku równania skalarnego i schematu jawnego.

Załóżmy bez straty ogólności, że $\alpha_q=1$. Mamy wtedy

$$x_{n+q} = -\sum_{j=0}^{q-1}\alpha_j x_{n+j} + h\sum_{j=0}^{q-1}\beta_j f_{n+j}. \tag{3.3}$$

Krok 1. Niech

$$Y_n = [x_{n+q-1},\dots,x_n]^{\mathsf{T}}.$$

Wówczas schemat (3.3) możemy zapisać w postaci

$$Y_{n+1} = AY_n + h\Phi(t_n,Y_n,h), \tag{3.4}$$

gdzie

$$A := \begin{bmatrix} -\alpha_{q-1} & -\alpha_{q-2} & \dots & -\alpha_0 \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{bmatrix} \quad \text{oraz} \quad \Phi(t_n,Y_n,h) := \begin{bmatrix} \sum_{j=0}^{q-1}\beta_j f_{n+j} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Zauważmy, że jeśli schemat (3.3) jest rzędu p, to schemat (3.4) też jest rzędu p, bo dla dokładnego rozwiązania $X(t) := [x(t+(q-1)h),\dots,x(t)]^{\mathsf{T}}$ mamy

$$\begin{aligned} \|T_h\| &= \|X(t+h) - AX(t) - h\Phi(t,X(t),h)\| \\ &= \left\| \begin{bmatrix} x(t+qh) + \sum_{j=0}^{q-1}\alpha_j x(t+jh) - h\sum_{j=0}^{q-1}\beta_j f(t+jh,x(t+jh)) \\ x(t+(q-1)h) - x(t+(q-1)h) \\ \vdots \\ x(t) - x(t) \end{bmatrix} \right\| \\ &= \Big\| x(t+qh) + \sum_{j=0}^{q-1}\alpha_j x(t+jh) - h\sum_{j=0}^{q-1}\beta_j f(t+jh,x(t+jh)) \Big\|, \end{aligned}$$

czyli lokalny błąd obcięcia schematu (3.3).

Krok 2. Dalej postępujemy tak, jak w dowodzie zbieżności dla schematów jednokrokowych (przy czym musimy pamiętać, że tym razem mamy jeszcze mnożenie przez A). Mamy wyrażenie na błąd:

$$\begin{aligned} X(t+h) - Y_{n+1} &= AX(t) + \big(X(t+h) - AX(t)\big) - AY_n - h\Phi(t,Y_n,h) \\ &= A\big(X(t) - Y_n\big) + \big(X(t+h) - AX(t) - h\Phi(t,X(t),h)\big) \\ &\quad + h\big(\Phi(t,X(t),h) - \Phi(t,Y_n,h)\big). \end{aligned}$$

Szacując teraz normę błędu dostajemy:

$$\|X(t+h) - Y_{n+1}\| \leqslant \|A\|\|X(t) - Y_n\| + \|T_h\| + hL_\Phi\|X(t) - Y_n\|,$$

bo lipschitzowskość f pociąga spełnianie warunku Lipschitza przez Φ z pewną stałą L_Φ.

Dla $\|A\| \leqslant 1$ daje to oszacowanie takie jak dla schematów jednokrokowych:

$$\|X(t+h) - Y_{n+1}\| \leqslant (1 + hL_\Phi)\|X(t) - Y_n\| + Ch^{p+1}.$$

Krok 3. Pozostaje pokazać, że $\|A\| \leqslant 1$.

Łatwo sprawdzić, że wartości własne A są pierwiastkami wielomianu charakterystycznego $p(\lambda)$ schematu (3.3). Zatem w postaci Jordana mamy:

$$J = T^{-1}AT = \begin{bmatrix} \lambda_1 & & & & & & \\ & \ddots & & & & & \\ & & \lambda_k & & & & \\ & & & \lambda_{k+1} & \varepsilon_{k+1} & & \\ & & & & \ddots & \varepsilon_{n-1} & \\ & & & & & \lambda_n & \end{bmatrix},$$

gdzie ze stabilności schematu $|\lambda_1| \leqslant 1, \ldots, |\lambda_k| \leqslant 1$, a $|\lambda_{k+1}| < 1, \ldots, |\lambda_n| < 1$, ale ich krotność może być większa niż 1, więc ε_j jest równe 0 lub 1. Pokazuje się (dobierając odpowiednie skalowanie T), że $|\varepsilon_j| < 1 - |\lambda_j|$, a stąd $\|J\|_\infty \leqslant 1$. Definiujemy normę $|\!|\!| x |\!|\!| := \|T^{-1}x\|_\infty$ i wtedy

$$|\!|\!| Ax |\!|\!| = |\!|\!| TJT^{-1}x |\!|\!| = \|JT^{-1}x\|_\infty \leqslant \|J\|_\infty \|T^{-1}x\|_\infty \leqslant |\!|\!| x |\!|\!|,$$

czyli $|\!|\!| A |\!|\!| \leqslant 1$. Korzystając z równoważności norm wektorowych dostajemy żądaną tezę. □

Rozdział 4

Zależność rozwiązań od parametrów i warunku początkowego

Zbadamy teraz zależność rozwiązania równania różniczkowego od parametrów i warunku początkowego. Będziemy w tym celu przybliżać rozwiązania innymi funkcjami i potrzebne będzie nam oszacowanie błędu tego przybliżenia. Wykorzystamy do tego pewne nierówności różniczkowe, które mogą stanowić same w sobie narzędzie do badania rozwiązań równań różniczkowych. Zaczniemy od pokazania ciągłej zależności od warunku początkowego.

4.1. Ciągła zależność od warunku początkowego

Zastanowimy się teraz nad zależnością rozwiązania zagadnienia Cauchy'ego od warunku początkowego. Wystarczy badać zależność od x_0, co wynika z następującego rozumowania. Podstawiając nową zmienną s poprzez zależność $t = s + t_0$ dostajemy zagadnienie

$$\begin{cases} \dfrac{d\widetilde{x}}{ds} = g(s, \widetilde{x}), \\ \widetilde{x}(0) = x_0, \end{cases}$$

gdzie $\widetilde{x}(s) = x(t)$, zaś $g(s, \widetilde{x}) := f(s + t_0, \widetilde{x})$. Zauważmy, że w warunku początkowym nie ma teraz zależności od t_0, a poprzez podstawienie $t = s + t_0$ zależność x od t_0 jest taka jak od t.

Będziemy używali oznaczenia $x(t, x_0)$, aby podkreślić, że rozwiązanie x jest również funkcją warunku początkowego. Następujące twierdzenie podaje warunki, przy jakich mamy ciągłą zależność rozwiązania od x_0.

Twierdzenie 4.1. *Niech $f(t, x)$ będzie ograniczona i ciągła dla $(t, x) \in J \times A \subset \mathbb{R} \times \mathbb{R}^n$, gdzie $J \times A$ jest zbiorem otwartym i ograniczonym. Załóżmy, że $(t_0, x_0) \in J \times A$ oraz f spełnia po zmiennej x warunek Lipschitza ze stałą L niezależną od t. Wtedy rozwiązanie $x(t, x_0)$ zależy w sposób ciągły od obu zmiennych (tzn. jest ciągłą funkcją argumentu $(t, x_0) \in J \times A$).*

Dowód. Niech będą dane rozwiązania $x(t, x_1)$ i $x(t, x_2)$ z warunkami początkowymi $x(t_0) = x_1$ i $x(t_0) = x_2$ odpowiednio. Z lematu Gronwalla i dowodu twierdzenia Picarda-Lindelöfa wynika oszacowanie

$$\|x(t, x_1) - x(t, x_2)\| \leqslant \|x_1 - x_2\| \mathrm{e}^{L(t-t_0)}.$$

Przedział J jest ograniczony, zatem $K := \sup_{t \in J} \mathrm{e}^{L(t-t_0)} < +\infty$. Mamy wtedy

$$\|x(t, x_1) - x(t, x_2)\| \leqslant K \|x_1 - x_2\|,$$

co oznacza, że rozwiązanie $x(t, y)$ spełnia warunek Lipschitza po warunku początkowym y ze stałą K niezależną od t. Jest zatem ciągłą funkcją warunku początkowego jednostajnie po $t \in J$. Ponadto $x(t, y)$ jest ciągłą funkcją t przy ustalonym y. Zatem (na mocy twierdzenia z analizy) jest ciągłą funkcją argumentu $(t, y) \in J \times A$. □

Powyższe twierdzenie można udowodnić przy nieco słabszych założeniach.

Twierdzenie 4.2. *Jeśli $f(t,x)$ jest ograniczona i ciągła w pewnym zbiorze otwartym $Q \subset \mathbb{R} \times \mathbb{R}^n$, a przez każdy punkt $(t_0, x_0) \in Q$ przechodzi dokładnie jedna krzywa całkowa $x(t, x_0)$, to x zależy w sposób ciągły od x_0.*

4.2. Nierówności różniczkowe

Zaczniemy od definicji rozwiązania przybliżonego.

Definicja 4.1. Funkcję $v : [t_0, t_0 + \alpha)$ nazwiemy *przybliżonym rozwiązaniem* zagadnienia Cauchy'ego (2.1)-(2.2) (jak zwykle $f : D \to \mathbb{R}^n$, $D \subset \mathbb{R} \times \mathbb{R}^n$, $(t_0, x_0) \in D$), jeśli spełnia następujące warunki:

i) v jest funkcją ciągłą, prawostronnie różniczkowalną w $[t_0, t_0 + \alpha)$, tzn.

$$\forall t \in [t_0, t_0 + \alpha) \quad \lim_{h \to 0^+} \frac{v(t+h) - v(t)}{h} \quad \text{istnieje}$$

(będziemy tę prawostronna pochodną oznaczać przez $v'_+(t)$),

ii) $\forall t \in [t_0, t_0 + \alpha) \quad (t, v(t)) \in D$.

Uwaga. 1) Funkcja v może więc być bardzo złym przybliżeniem. 2) Można wziąć jako v łamaną Eulera (ma prawostronną pochodną).

Naszym celem jest oszacowanie błędu przybliżenia prawdziwego rozwiązania $x(t)$ przez $v(t)$

$$\|v(t) - x(t)\| \leqslant \ ?$$

w zależności od błędu przybliżenia warunku początkowego $\|v(t_0) - x_0\|$ oraz błędu zgodności (inaczej residuum) $\|v'_+(t) - f(t, v(t))\|$. Jak wprowadzić pochodne do tego oszacowania? Intuicyjnie wydaje się, że można różniczkować normę, ale nie musi być ona różniczkowalna, a do tego v nie musi być różniczkowalna. Dlatego wprowadzimy pewne uogólnienie pochodnych: pochodne Diniego.

Definicja 4.2. *Dolna prawostronna pochodna Diniego* (w skrócie *pochodna Diniego*) rzeczywistej funkcji m określonej w pewnym prawostronnym otoczeniu punktu $t \in \mathbb{R}$, to

$$D_+ m(t) := \liminf_{h \to 0^+} \frac{m(t+h) - m(t)}{h}.$$

Przykład. 1) Jeśli m ma prawostronną pochodną w t, to

$$D_+ m(t) = \lim_{h \to 0^+} \frac{m(t+h) - m(t)}{h} = m'_+(t).$$

2) Niech m będzie funkcją Dirichleta, czyli funkcją charakterystyczną zbioru liczb wymiernych. Jest ona wszędzie nieciągła, mimo to łatwo zauważyć, że jej dolna pochodna Diniego istnieje:

$$\begin{aligned} \forall t \in \mathbb{Q} \quad & D_+ m(t) = -\infty, \\ \forall t \in \mathbb{R} \setminus \mathbb{Q} \quad & D_+ m(t) = 0. \end{aligned}$$

Lemat 4.1. *Niech m będzie określona w prawostronnym otoczeniu $t \in \mathbb{R}$ o wartościach w $\mathbb{R}^n$ i niech istnieje $m'_+(t)$. Wtedy*

$$D_+ \|m(t)\| \leqslant \|m'_+(t)\|.$$

Dowód. Ponieważ granica dolna zachowuje słabe nierówności, to z nierówności trójkąta mamy dla $h > 0$

$$\begin{aligned} D_+ \|m(t)\| &= \liminf_{h \to 0^+} \frac{\|m(t+h)\| - \|m(t)\|}{h} \leqslant \liminf_{h \to 0^+} \left\| \frac{m(t+h) - m(t)}{h} \right\| \\ &\overset{(1)}{=} \lim_{h \to 0^+} \left\| \frac{m(t+h) - m(t)}{h} \right\| \overset{(2)}{=} \left\| \lim_{h \to 0^+} \frac{m(t+h) - m(t)}{h} \right\| = \|m'_+(t)\|, \end{aligned}$$

gdzie równość (1) zachodzi, jeśli granica po prawej istnieje, a to jest spełnione, zaś równość (2) wynika z ciągłości normy. □

Twierdzenie 4.3 (o nierówności)**.** *Niech $m(t)$ i $u(t)$ będą funkcjami ciągłymi, rzeczywistymi, określonymi na $[t_0, t_0+\alpha]$ i niech $g : \mathbb{R}^2 \to \mathbb{R}$ będzie dowolną funkcją. Jeśli dla wszystkich $t \in [t_0, t_0+\alpha)$ zachodzi*

(i) $D_+m(t) \leqslant g(t, m(t))$,
(ii) $D_+u(t) > g(t, u(t))$,
(iii) $m(t_0) \leqslant u(t_0)$,

to dla wszystkich $t \in [t_0, t_0+\alpha]$ zachodzi $m(t) \leqslant u(t)$.

Dowód. Przypuśćmy, że tak nie jest i istnieje $t_1 \in [t_0, t_0+\alpha)$ takie, że $m(t_1) > u(t_1)$.

Różnica $m(t) - u(t)$ jest funkcją ciągłą, która przyjmuje w t_0 wartość niedodatnią z założenia (iii), a w t_1 dodatnią. Zatem w pewnym punkcie między t_0 a t_1 musi przyjmować wartość zero (być może w t_0). Niech t^* będzie pierwszym takim punktem na lewo od t_1, tzn. $m(t^*) = u(t^*)$ i $m(t) > u(t)$ dla $t \in (t^*, t_1]$. Mamy wtedy $m(t^*+h) > u(t^*+h)$ dla h takich, że $t^*+h \leqslant t_1$. Stąd

$$\frac{m(t^*+h) - m(t^*)}{h} > \frac{u(t^*+h) - u(t^*)}{h}.$$

Przechodząc do granicy dostajemy $D_+m(t^*) \geqslant D_+u(t^*)$.

Z drugiej strony mamy

$$D_+m(t^*) \leqslant g(t^*, m(t^*)) = g(t^*, u(t^*)) < D_+u(t^*).$$

Dostaliśmy sprzeczność, zatem $m(t) \leqslant u(t)$ dla $t \in [t_0, t_0+\alpha)$. Nierówność $m(t_0+\alpha) \leqslant u(t_0+\alpha)$ wynika z ciągłości $m(t)$ i $u(t)$. □

Uwaga. Twierdzenie pozostaje prawdziwe (z analogicznym dowodem) przy zamianie nierówności („$\leqslant$" na „$\geqslant$" oraz „$>$" na „$<$") w założeniach i tezie.

Twierdzenie 4.4 (o aproksymacji)**.** *Niech $x : [t_0, t_0+\alpha) \to \mathbb{R}^n$ będzie rozwiązaniem zagadnienia Cauchy'ego* (2.1)-(2.2) *przy czym $f : D \to \mathbb{R}^n$, $D = [t_0, t_0+\alpha) \times U$, $U \subset \mathbb{R}^n$ otwarty. Niech $v : [t_0, t_0+\alpha) \to \mathbb{R}^n$ będzie rozwiązaniem przybliżonym. Jeśli dla pewnych $\eta, \delta > 0$*

a) $\|v(t_0) - x_0\| \leqslant \eta$,
b) $\forall t \in [t_0, t_0+\alpha) \quad \|v'_+(t) - f(t, v(t))\| \leqslant \delta$,
c) $\exists L \geqslant 0 \quad \forall (t,x),(t,y) \in D \quad \|f(t,x) - f(t,y)\| \leqslant L\|x-y\|$,

to

$$\text{dla } t \in [t_0, t_0+\alpha) \qquad \|x(t) - v(t)\| \leqslant \begin{cases} \eta e^{L(t-t_0)} + \dfrac{\delta}{L}\Big(e^{L(t-t_0)} - 1\Big), & L > 0, \\ \eta + \delta(t-t_0), & L = 0. \end{cases}$$

Dowód. Idea dowodu polega na oszacowaniu błędu rozwiązania przybliżonego i skorzystaniu z twierdzenia o nierówności. Mamy

$$\begin{aligned} D_+\|x(t) - v(t)\| \leqslant \|x'(t) - v'_+\| = \|f(t,x) - v'_+\| &= \|f(t,x) - f(t,v) + f(t,v) - v'_+(t)\| \\ &\leqslant L\|x(t) - v(t)\| + \underbrace{\|v'_+(t) - f(t,v)\|}_{\text{błąd zgodności } \delta(t)} \end{aligned} \tag{4.1}$$

Niech teraz $m(t) := \|x(t) - v(t)\|$. Funkcja m jest ciągła, więc możemy skorzystać z twierdzenia o nierówności dla $t \in [t_0, t_0+\beta]$, gdzie $0 \leqslant \beta < \alpha$. Ze wzoru (4.1) mamy dla dowolnego $\varepsilon > 0$

$$D_+m(t) \leqslant \underbrace{Lm(t) + \delta}_{=:\, g(t, m(t))} < g(t, m(t)) + \varepsilon.$$

Zatem

(i) $D_+m(t) \leqslant g(t, m(t))$,
(ii) dla $u(t)$ będącego rozwiązaniem zagadnienia początkowego

$$\begin{cases} u' = Lu(t) + \delta + \varepsilon, \\ u(t_0) = \eta, \end{cases}$$

zachodzi $D_+u(t) = u'(t) = g(t, u(t)) + \varepsilon > g(t, u(t))$,

(iii) $m(t_0) = u(t_0)$.
Spełnione są więc założenia twierdzenia o nierówności. Wobec tego

$$m(t) \leqslant u(t) = \eta \mathrm{e}^{L(t-t_0)} + \frac{\delta + \varepsilon}{L}\Big(\mathrm{e}^{L(t-t_0)} - 1\Big).$$

Z dowolności ε pokazaliśmy tezę dla $t \in [t_0, t_0 + \beta]$, $\beta < \alpha$. Podobnie z dowolności β mamy tezę dla $t \in [t_0, t_0 + \alpha)$. Przypadek $L = 0$ pokazuje się analogicznie. □

Uwaga. Siła tego twierdzenia polega na tym, że v nie musi być rozwiązaniem jakiegokolwiek równania różniczkowego i wystarczy tylko prawostronna różniczkowalność v.

Podobnie można udowodnić uogólnienie tego twierdzenia na przypadek, gdy δ i L są funkcjami t.

Twierdzenie 4.5 (Uogólnione o aproksymacji)**.** *Niech będą spełnione założenia poprzedniego twierdzenia z wyjątkiem tego, że warunki b) i c) zastąpimy przez:*
b') $\forall t \in [t_0, t_0 + \alpha) \quad \|v'_+(t) - f(t, v(t))\| \leqslant \delta(t)$,
c') $\forall t \in [t_0, t_0 + \alpha) \quad \exists L(t) \geqslant 0 \quad \forall (t, x), (t, y) \in D \quad \|f(t, x) - t(t, y)\| \leqslant L(t)\|x - y\|$,
gdzie $\delta, L : [t_0, t_0 + \alpha] \to \mathbb{R}$ *ciągłe. Wtedy dla* $t \in [t_0, t_0 + \alpha)$ *oraz* $A(t) = \int_{t_0}^{t} L(s)\,ds$ *zachodzi*

$$\|x(t) - v(t)\| \leqslant e^{A(t)} \left(\eta + \int_{t_0}^{t} e^{-A(\tau)}\delta(\tau)\,d\tau\right).$$

Uwaga. Można pokazać, że oba powyższe twierdzenia zachodzą dla t ze zbioru domkniętego $[t_0, t_0 + \alpha]$ przy tych samych założeniach.

4.3. Pochodna względem parametru

Rozważamy rodzinę równań różniczkowych zależnych od parametru $p \in \mathbb{R}$:

$$\begin{cases} x'(t) = f(t, x(t), p), \\ x(t_0) = x_0. \end{cases} \tag{4.2}$$

Oczywiście rozwiązania zależą od p, więc możemy napisać $x(t) = x(t, p)$. Chcemy zbadać *wrażliwość* rozwiązań *na zmianę parametru*. Dobrą miarą tej wrażliwości jest pochodna $\frac{\partial x}{\partial p}(t, p)$, o ile istnieje, a idea jej wyznaczenia jest następująca. Różniczkujemy formalnie (4.2) względem p:

$$\begin{aligned} &\frac{\partial}{\partial p}\frac{\partial}{\partial t}x(t,p) = \frac{\mathrm{d}}{\mathrm{d}p}f(t, x(t,p), p) \\ &\qquad \| \ ? \qquad\qquad \| \\ &\frac{\partial}{\partial t}\frac{\partial}{\partial p}x(t,p) \stackrel{?}{=} \frac{\partial f}{\partial x}(t, x(t,p), p)\frac{\partial x}{\partial p}(t,p) + \frac{\partial f}{\partial p}(t, x(t,p), p) \end{aligned}$$

Nawet gdyby można było formalnie zróżniczkować prawą stronę równania, to nie wiemy, czy zachodzą równości ze znakiem zapytania. Niemniej przy odpowiednich założeniach pochodna $\psi(t) := \frac{\partial x}{\partial p}(t, p)$ spełniałaby równanie liniowe

$$\begin{cases} \psi'(t) = \dfrac{\partial f}{\partial x}(t, x(t,p), p)\psi(t) + \dfrac{\partial f}{\partial p}(t, x(t,p), p), \\ \psi(t_0) = 0. \qquad \leftarrow \text{bo w } t_0 \text{ wartość } x \text{ nie zależy od } p \end{cases}$$

Teraz to uściślimy.

Twierdzenie 4.6 (o pochodnej względem parametru)**.** *Niech* $t_0 \in \mathbb{R}$, $x_0 \in \mathbb{R}^n$, $p_0 \in \mathbb{R}$. *Niech* $f : D \to \mathbb{R}^n$, $D = J \times U \times P \subset \mathbb{R}^{n+2}$, $J \subset \mathbb{R}$, $U \subset \mathbb{R}^n$, $P \subset \mathbb{R}$ *niepuste, otwarte. Oznaczmy*

$$Q := \{(t, x, p) \in \mathbb{R}^{n+2} : \ t_0 \leqslant t \leqslant t_0 + a, \ \|x - x_0\| \leqslant b, \ |p - p_0| \leqslant c\}$$

i niech $a, b, c > 0$ będą dobrane tak, aby $Q \subset D$. Załóżmy, że f jest ciągła w D i oznaczmy $M := \sup_{(t,x,p)\in Q} \|f(t,x,p)\|$. Niech $0 < \alpha \leqslant a$ będzie takie, że rozwiązania (4.2) istnieją i są jednoznacznie określone na przedziale $[t_0, t_0+\alpha]$ dla każdego p takiego, że $|p-p_0| \leqslant c$. Załóżmy ponadto, że pochodne

$$\frac{\partial f}{\partial x} = \left[\frac{\partial f_i}{\partial x_j}\right]_{i,j=1,\dots,n} \qquad \textit{oraz} \qquad \frac{\partial f}{\partial p} = \left[\frac{\partial f_i}{\partial p}\right]_{i=1,\dots,n}$$

istnieją i są ciągłe w D. Niech $x : [t_0, t_0+\alpha] \to \mathbb{R}^n$ będzie rozwiązaniem zagadnienia (4.2) Wtedy pochodna $\psi(t) := \frac{\partial x}{\partial p}(t,p)$ istnieje w $p = p_0$ dla $t \in [t_0, t_0+\alpha]$ i spełnia równanie

$$\begin{cases} \psi'(t) = \dfrac{\partial f}{\partial x}(t, x(t,p_0), p_0)\psi(t) + \dfrac{\partial f}{\partial p}(t, x(t,p_0), p_0), \\ \psi(t_0) = 0. \end{cases} \tag{4.3}$$

Dowód. Krok 0. Zauważmy, że przy założeniach twierdzenia f spełnia warunek Lipschitza względem x w Q ze stałą L oraz względem p w Q ze stałą A, gdzie

$$L := \sup_{(t,x,p)\in Q} \left\|\frac{\partial f}{\partial x}(t,x,p)\right\|, \qquad A := \sup_{(t,x,p)\in Q} \left\|\frac{\partial f}{\partial p}(t,x,p)\right\|.$$

(Uwaga: pierwsza norma powyżej to norma macierzowa indukowana, zaś druga to norma wektorowa.)

Na mocy lipschitzowskości f mamy dla dowolnego p takiego, że $|p-p_0| \leqslant c$ i dopuszczalnych (t,x) i (t,y):

$$\|f(t,x,p) - f(t,y,p)\| \leqslant L\|x-y\|$$

i stąd na mocy twierdzenia Picarda-Lindelöfa dla każdego takiego p istnieje na $[t_0, t_0+\alpha]$ jednoznaczne rozwiązanie (4.2), gdzie $\alpha = \min\{a, \frac{b}{M}\}$. Zatem mamy zagwarantowane istnienie α postulowanego w założeniu twierdzenia. Ponadto rozwiązanie (4.3) istnieje, jako rozwiązanie układu liniowego (w przypadku skalarnym było, a w przypadku wektorowym będzie pokazane). Będziemy dalej przybliżać pochodną $\frac{\partial x}{\partial p}(t,p)$ ilorazami różnicowymi.

Krok 1. Rozważmy rozwiązanie (4.2) dla parametru $p_0+\Delta$, gdzie $|\Delta| \leqslant c$, i $z(t) = x(t, p_0+\Delta)$:

$$\begin{cases} z'(t) = f(t, z(t), p_0+\Delta), \\ z(t_0) = x_0. \end{cases}$$

Pokażemy oszacowanie na $\|z(t) - x(t,p_0)\|$. Korzystamy z twierdzenia 4.5 biorąc $z(t)$ jako rozwiązanie przybliżone. Sprawdzamy założenia:

a) $\|z(t_0) - x_0\| = 0 =: \eta$,

b) $\|z'_+(t) - f(t, z(t), p_0)\| = \|f(t, z(t), p_0+\Delta) - f(t, z(t), p_0)\| \leqslant A|\Delta| =: \delta$, (z war. Lip. f po p)

c) $\|f(t,x,p_0) - f(t,y,p_0)\| \leqslant L\|x-y\|$.

Zatem zachodzi teza twierdzenia o aproksymacji:

$$\|z(t) - x(t,p_0)\| \leqslant \frac{A}{L}(\mathrm{e}^{L(t-t_0)} - 1)|\Delta| \leqslant \frac{A}{L}(\mathrm{e}^{L\alpha} - 1)|\Delta| =: B|\Delta|$$

(Uwaga: zauważmy, że ten wynik daje ciągłą zależność od parametru.)

Krok 2. Rozważmy teraz iloraz różnicowy

$$\frac{z(t) - x(t,p_0)}{\Delta} = \frac{x(t,p_0+\Delta) - x(t,p_0)}{\Delta}.$$

Pokażemy, że granica przy $\Delta \to 0$ istnieje i jest równa $\psi(t)$. W tym celu ponownie zastosujemy twierdzenie 4.5 dla $\psi(t)$ i $\frac{1}{\Delta}(z(t) - x(t))$ jako przybliżenia (rozważamy równanie (4.3)). Sprawdzamy założenia:

a) $\|\frac{1}{\Delta}(z(t_0) - x(t_0)) - \psi(t_0)\| = 0 =: \eta$,

c) $\|\frac{\partial f}{\partial x}(t,x,p_0)v + \frac{\partial f}{\partial p}(t,x,p_0) - (\frac{\partial f}{\partial x}(t,x,p_0)u + \frac{\partial f}{\partial p}(t,x,p_0))\| = \|\frac{\partial f}{\partial x}(t,x,p_0)(v-u)\| \leqslant L\|u-v\|$

b) Niech $T := \|\frac{d}{dt}(\frac{1}{\Delta}(z(t)-x(t,p_0))) - \frac{\partial f}{\partial x}(t,x,p_0)\frac{1}{\Delta}(z(t)-x(t)) - \frac{\partial f}{\partial p}(t,x,p_0)\|$. Z definicji z i x mamy ze wzoru Taylora w punkcie (x,p_0):

$$\begin{aligned} z'(t) - x'(t,p_0) &= f(t,z(t),p_0+\Delta) - f(t,x(t,p_0),p_0) \\ &= \frac{\partial f}{\partial x}(t,x,p_0)(z(t)-x(t,p_0)) + \frac{\partial f}{\partial p}(t,x,p_0)\Delta + R(t,x,p_0,z-x,\Delta), \end{aligned}$$

gdzie reszta R spełnia

$$\forall \varepsilon > 0\ \exists\, \delta > 0 \quad \left\| \begin{bmatrix} z-x \\ \Delta \end{bmatrix} \right\| \leqslant \delta \implies \frac{\|R(t,x,p_0,z-x,\Delta)\|}{\left\| \begin{bmatrix} z-x \\ \Delta \end{bmatrix} \right\|} \leqslant \varepsilon.$$

Dla normy drugiej, której tutaj używamy, zachodzi $\left\| \begin{bmatrix} 0 \\ \Delta \end{bmatrix} \right\| = |\Delta|$ oraz $\left\| \begin{bmatrix} z-x \\ 0 \end{bmatrix} \right\| = \|z-x\|$, więc z nierówności trójkąta i kroku 1 mamy

$$\left\| \begin{bmatrix} z-x \\ \Delta \end{bmatrix} \right\| \leqslant \|z-x\| + |\Delta| \leqslant (B+1)|\Delta|.$$

Stąd

$$T = \frac{\|R(t,x,p_0,z-x,\Delta)\|}{|\Delta|} = \frac{\|R(t,x,p_0,z-x,\Delta)\|}{\left\| \begin{bmatrix} z-x \\ \Delta \end{bmatrix} \right\|} \frac{\left\| \begin{bmatrix} z-x \\ \Delta \end{bmatrix} \right\|}{|\Delta|} \leqslant \frac{\|R(t,x,p_0,z-x,\Delta)\|}{\left\| \begin{bmatrix} z-x \\ \Delta \end{bmatrix} \right\|}(B+1).$$

Niech więc dla dowolnego $\varepsilon_1 > 0$ będzie $\varepsilon := \frac{\varepsilon_1}{B+1}$ z dobranym odpowiednim δ. Do tego δ dobieramy $|\Delta| \leqslant \frac{\delta}{B+1}$ tak, aby $\left\| \begin{bmatrix} z-x \\ \Delta \end{bmatrix} \right\| \leqslant \delta$. Ostatecznie mamy $T \leqslant \frac{\varepsilon_1}{B+1}(B+1) = \varepsilon_1$.

Założenia twierdzenia o aproksymacji są spełnione, prawdziwe więc jest oszacowanie

$$\left\| \frac{1}{\Delta}(z(t)-x(t,p_0)) - \psi(t) \right\| \leqslant \frac{\varepsilon_1}{L}(e^{L(t-t_0)}-1) \leqslant \frac{\varepsilon_1}{L}(e^{L\alpha}-1),$$

co z dowolności ε_1 oznacza, że mamy punktową zbieżność

$$\frac{1}{\Delta}(x(t,p_0+\Delta) - x(t,p_0)) \underset{\Delta\to 0}{\longrightarrow} \psi(t). \qquad \square$$

Wniosek 4.1 (Pochodne cząstkowe względem k-parametrów)**.** *Twierdzenie zachodzi również dla $p \in \mathbb{R}^k$ przy odpowiednio zmodyfikowanych założeniach: $D = J\times U\times P \subset \mathbb{R}^{n+1+k}$, $J \subset \mathbb{R}$, $U \subset \mathbb{R}^n$, $P \subset \mathbb{R}^k$, $Q := \{(t,x,p) \in \mathbb{R}^{n+1+k} :\ t_0 \leqslant t \leqslant t_0 + a,\ \|x-x_0\| \leqslant b,\ \|p-p_0\|_\infty \leqslant c\}$ oraz $\frac{\partial f}{\partial p} = \left[\frac{\partial f_i}{\partial p_j}\right]_{i=1,\dots,n}^{j=1,\dots,k}$. Teza pozostaje prawdziwa bez zmian, przy czym $\psi(t)$ jest macierzą wymiaru $n\times k$ i* (4.3) *rozumiemy jako k układów o n niewiadomych.*

4.4. Pochodna względem warunku początkowego

Rozwiązania zagadnienia

$$\begin{cases} x' = f(t,x), \\ x(t_0) = x_0 =: p \end{cases}$$

możemy traktować jako rodzinę zależną od parametru p: $x(t) = x(t,x_0) = x(t,p)$. Zauważmy, że to równanie jest równoważne zadaniu z parametrem i zerowym warunkiem początkowym dla $z(t) = x(t,p) - p$:

$$\begin{cases} z'(t) = f(t,x(t)) = f(t,z(t)+p) =: F(t,z(t),p)), \\ z(t_0) = 0. \end{cases}$$

Zatem $z(t) = z(t,p)$ i mamy zagadnienie z parametrem będącym wektorem. Interesuje nas pochodna $\frac{\partial}{\partial x_0}x(t,x_0)$ (jest to macierz $n\times n$), skąd $\psi(t) := \frac{\partial}{\partial x_0}x(t,x_0) = \frac{\partial z}{\partial p}(t,p)+\frac{\partial}{\partial p}p = \frac{\partial z}{\partial p}(t,x)+I_n$, gdzie I_n jest macierzą jednostkową. Ponieważ kolumny $\psi_j = \frac{\partial x}{\partial p_j}(t,x_0)$, to możemy skorzystać n razy z twierdzenia o pochodnej względem parametru albo z wniosku przy $k = n$. Teraz uściślimy tę ideę.

Twierdzenie 4.7 (o pochodnej względem warunku początkowego)**.** *Niech $f : D \to \mathbb{R}^n$, $D = J \times U$, $J \subset \mathbb{R}$, $U \subset \mathbb{R}^n$ otwarte, niepuste. Niech $(t_0, x_0) \in D$ i oznaczmy $Q := \{(t,x) \in \mathbb{R}^{n+1} : \ t_0 \leqslant t \leqslant t_0 + a, \ \|x - x_0\| \leqslant b\}$ dla $a > 0$ i b takich, że $Q \subset D$. Niech f będzie ciągła w D oraz $M := \sup_{(t,x)\in Q} \|f(t,x)\|$. Załóżmy, że $0 < \alpha \leqslant a$ jest takie, że rozwiązania $x(t) = x(t,p)$ zagadnienia*

$$\begin{cases} x' = f(t,x), \\ x(t_0) = p \end{cases} \tag{4.4}$$

istnieją i są jednoznacznie określone na przedziale $[t_0, t_0 + \alpha]$ dla każdego p takiego, że $\|p - x_0\| \leqslant b$. Niech ponadto pochodne $\frac{\partial f}{\partial x} = \left[\frac{\partial f_i}{\partial x_j}\right]_{i,j=1,\dots,n}$ istnieją i są ciągłe w D. Załóżmy, że $x(t,x_0)$ jest rozwiązaniem (4.4) dla $p = x_0$. Wtedy $\psi(t) := \frac{\partial x}{\partial x_0}(t,x_0)$ istnieje w $p = x_0$ dla $t \in [t_0, t_0 + \alpha]$ oraz spełnia równanie

$$\begin{cases} \psi'(t) = \dfrac{\partial f}{\partial x}(t, x(t,x_0))\psi(t), \\ \psi(t_0) = I_n. \end{cases} \tag{4.5}$$

Dowód. Dowód polega na sprowadzeniu problemu do znalezienia pochodnej po parametrach. Skorzystamy z wniosku z twierdzenia o pochodnej po parametrze dla rozwiązań równania:

$$\begin{cases} z'(t) = F(t,z,p), \quad (= f(t, z(t) + p)) \\ z(t_0) = 0. \end{cases} \tag{4.6}$$

Niech $p_0 := x_0$ i $z_0 := 0$ oraz

$$\widetilde{Q} := \{(t,z,p) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n : \ t_0 \leqslant t \leqslant t_0 + a, \ \|z - z_0\| \leqslant \frac{b}{2}, \ \|p - p_0\| \leqslant \|p - p_0\|_\infty \leqslant \frac{b}{2}\}.$$

Niech ponadto $\mathcal{P} := \{p \in \mathbb{R}^n : \|p - p_0\| \leqslant \frac{b}{2}\}$ oraz $\mathcal{U} := \{z \in \mathbb{R}^n : \|z - z_0\| \leqslant \frac{b}{2}\}$. Zauważmy, że jeśli $p \in \mathcal{P}$ i $z \in \mathcal{U}$, to $(t, z + p) \in Q$. Mamy zatem

$$\max_{(t,z,p)\in\widetilde{Q}} \|F(t,z,p)\| = \max_{(t,z,p)\in\widetilde{Q}} \|f(t,z+p)\| \leqslant M.$$

Rozwiązania (4.6) istnieją i są jednoznaczne na $[t_0, t_0 + \alpha]$, bo (4.6) jest równoważne (4.4). Pozostaje sprawdzić, czy $\frac{\partial F}{\partial z}$ i $\frac{\partial F}{\partial p}$ istnieją i są ciągłe. Mamy $\frac{\partial F}{\partial z} = \frac{\partial}{\partial z}f(t,z+p) = \frac{\partial f}{\partial x}(t,z+p)$ oraz $\frac{\partial F}{\partial p} = \frac{\partial}{\partial p}f(t,z+p) = \frac{\partial f}{\partial x}(t,z+p)$, czyli istnieją i są ciągłe z założenia o f.

Zatem zachodzi teza twierdzenia i pochodna $\phi(t) := \frac{\partial z}{\partial p}(t,p_0)$ rozwiązania $z(t) = z(t,p)$ spełnia

$$\begin{cases} \phi'(t) = \dfrac{\partial f}{\partial x}(t, z+p_0)\phi(t) + \dfrac{\partial f}{\partial x}(t,z+p_0) = \dfrac{\partial f}{\partial x}(t,z+p_0)(\phi + I_n), \\ \phi(t_0) = 0. \end{cases}$$

Ale $z(t,p_0) = z(t,x_0) = x(t,x_0) - x_0$, więc $\frac{\partial z}{\partial x_0}(t,x_0) = \frac{\partial x}{\partial x_0} - I_n$. Stąd $\psi(t) = \frac{\partial x}{\partial x_0}(t,x_0)$ spełnia $\psi(t) = \phi(t) + I_n$ oraz $\psi'(t) = \phi'(t)$. Zatem $\psi(t)$ spełnia (4.5). □

Rozdział 5

Układy liniowych równań różniczkowych zwyczajnych

5.1. Równania różniczkowe liniowe drugiego rzędu

Zaczniemy ten rozdział od równań różniczkowych liniowych drugiego rzędu postaci

$$\ddot{x} + p(t)\dot{x} + q(t)x = r(t), \qquad t \in (a,b) \tag{5.1}$$

z warunkami początkowymi

$$x(t_0) = x_0^{(1)}, \qquad \dot{x}(t_0) = x_0^{(2)}.$$

Możemy zamienić to równanie na układ równań liniowych pierwszego rzędu

$$\begin{cases} \dot{y}_1 = y_2, \\ \dot{y}_2 = -p(t)y_2 - q(t)y_1 + r(t), \end{cases} \quad \text{z warunkiem} \quad \begin{cases} y_1(t_0) = x_0^{(1)}, \\ y_2(t_0) = x_0^{(2)}. \end{cases}$$

Jeśli $p(t)$, $q(t)$ i $r(t)$ są ciągłe na (a,b), to rozwiązanie równania drugiego rzędu istnieje na (a,b) i jest jednoznaczne na mocy odpowiedniego twierdzenia dla układów liniowych (co wkrótce zobaczymy).

Z uwagi na możliwość tej zamiany, równania liniowe drugiego rzędu można wpisać w ramy teorii układów liniowych. Jednakże, ze względu na ich specyfikę i znaczenie praktyczne, zbadamy je osobno.

Obok równania niejednorodnego (5.1) będziemy rozważać równanie jednorodne

$$\ddot{x} + p(t)\dot{x} + q(t)x = 0. \tag{5.2}$$

Twierdzenie 5.1. *Niech $x_1(t)$ i $x_2(t)$ będą rozwiązaniami równania jednorodnego* (5.2) *takimi, że*

$$x_1(t)\dot{x}_2(t) - \dot{x}_1(t)x_2(t) \neq 0, \qquad \textit{dla każdego} \quad t \in (a,b).$$

Wtedy każde rozwiązanie $x(t)$ równania jednorodnego można zapisać dla pewnych $c_1, c_2 \in \mathbb{R}$ jako

$$x(t) = c_1 x_1(t) + c_2 x_2(t).$$

Dowód. Niech $x_0^{(1)} := x(t_0)$ i $x_0^{(2)} = \dot{x}(t_0)$. Przy założeniach twierdzenia układ

$$\begin{cases} c_1 x_1(t_0) + c_2 x_2(t_0) = x_0^{(1)}, \\ c_1 \dot{x}_1(t_0) + c_2 \dot{x}_2(t_0) = x_0^{(2)} \end{cases}$$

ma jednoznaczne rozwiązanie c_1, c_2. Funkcja $\phi(t) = c_1 x_1(t) + c_2 x_2(t)$ oczywiście spełnia (5.2) i warunki początkowe dla $x(t)$, zatem z jednoznaczności rozwiązania jest tożsamościowo równa funkcji $x(t)$. □

Definicja 5.1. Niech $x_1(t)$ i $x_2(t)$ będą różniczkowalne na (a,b). Wyrażenie

$$W(x_1,x_2)(t) := \det \begin{bmatrix} x_1(t) & x_2(t) \\ \dot{x}_1(t) & \dot{x}_2(t) \end{bmatrix}$$

nazywamy *wyznacznikiem Wrońskiego* (*wrońskianem*) układu x_1, x_2. Jeżeli $W(x_1,x_2)(t) \neq 0$ dla każdego $t \in (a,b)$, to układ $x_1(t)$, $x_2(t)$ nazywamy *liniowo niezależnym*.

Zachodzi następujący oczywisty fakt.

Lemat 5.1. *Jeśli $x_1(t)$, $x_2(t)$ są liniowo niezależnymi rozwiązaniami równania jednorodnego, zaś $x_c(t)$ jest jakimkolwiek rozwiązaniem równania niejednorodnego, to ogólne rozwiązanie równania niejednorodnego jest dla $c_1, c_2 \in \mathbb{R}$ postaci $x(t) = x_c(t) + c_1x_1(t) + c_2x_2(t)$.*

Nie znamy ogólnej metody znajdowania rozwiązań równania jednorodnego o dowolnych współczynnikach (ani tym bardziej niejednorodnego). Dlatego ograniczymy się teraz do przypadku równania jednorodnego o stałych współczynnikach

$$a\ddot{x} + b\dot{x} + cx = 0.$$

Rozwiązań stanowiących układ liniowo niezależny będziemy szukali w postaci funkcji wykładniczej $\mathrm{e}^{\lambda t}$. Po wstawieniu w miejsce funkcji $x(t)$ do równania otrzymujemy

$$(a\lambda^2 + b\lambda + c)\mathrm{e}^{\lambda t} = 0.$$

Równanie jest spełnione tożsamościowo dla t, jeśli λ jest rozwiązaniem *równania charakterystycznego*

$$a\lambda^2 + b\lambda + c = 0.$$

Równanie charakterystyczne ma pierwiastki $\lambda_1, \lambda_2 \in \mathbb{C}$, które zależą od wyróżnika równania kwadratowego $\Delta = b^2 - 4ac$. Mamy więc przypadki:

- $\Delta > 0$: $\lambda_1, \lambda_2 \in \mathbb{R}$, $\lambda_1 \neq \lambda_2$. Funkcje $x_1(t) = \mathrm{e}^{\lambda_1 t}$ i $x_2(t) = \mathrm{e}^{\lambda_2 t}$ są niezależnymi rozwiązaniami równania jednorodnego, bo jak łatwo sprawdzić: $W(\mathrm{e}^{\lambda_1 t}, \mathrm{e}^{\lambda_2 t}) = (\lambda_2 - \lambda_1)\mathrm{e}^{(\lambda_1+\lambda_2)t} \neq 0$ dla $t \in \mathbb{R}$.
- $\Delta < 0$: $\lambda_1, \lambda_2 \in \mathbb{C} \setminus \mathbb{R}$, $\lambda_1 \neq \lambda_2$. Niech $\lambda_1 = \alpha + i\beta$, $\lambda_2 = \alpha - i\beta$. Wtedy $\mathrm{e}^{\lambda_1 t} = \mathrm{e}^{\alpha t}(\cos\beta t + i\sin\beta t)$, $\mathrm{e}^{\lambda_2 t} = \mathrm{e}^{\alpha t}(\cos\beta t - i\sin\beta t)$. Stąd funkcje $x_1(t) = \mathrm{e}^{\alpha t}\cos\beta t$ i $x_2(t) = \mathrm{e}^{\alpha t}\sin\beta t$ są liniowo niezależnymi rozwiązaniami: $W(x_1, x_2)(t) = \beta\mathrm{e}^{2\alpha t} \neq 0$ dla $t \in \mathbb{R}$.
- $\Delta = 0$: $\lambda_1 = \lambda_2 = \lambda \in \mathbb{R}$. Mamy jedno rozwiązanie $x_1(t) = \mathrm{e}^{\lambda t}$. Drugiego szukamy w postaci $x_2(t) = x_1(t)u(t)$. Podstawiając x_2 do równania jednorodnego, dostajemy równanie $\ddot{u} = 0$, skąd przyjmujemy $x_2(t) = t\mathrm{e}^{\lambda t}$. Mamy $W(x_1, x_2)(t) = \mathrm{e}^{2\lambda t} \neq 0$ dla $t \in \mathbb{R}$.

Uwaga. W przypadku równań o zmiennych współczynnikach, gdy znamy jedno dowolne rozwiązanie $x_1(t)$ (np. zgadliśmy), drugiego rozwiązania niezależnego możemy próbować szukać w postaci $x_2(t) = x_1(t)u(t)$.

Uwaga. W analogiczny sposób (znajdując pierwiastki wielomianu charakterystycznego) szukamy rozwiązań równań wyższych rzędów.

Na koniec przedstawimy wskazówki, jak można rozwiązywać równania niejednorodne.

Mając ogólne rozwiązanie równania jednorodnego $x^{(j)}(t) = c_1x_1(t) + c_2x_2(t)$, możemy szukać szczególnego rozwiązania równania niejednorodnego w postaci

$$x^{(s)}(t) = c_1(t)x_1(t) + c_2(t)x_2(t),$$

czyli *metodą uzmienniania stałej* (można to robić również w przypadku równania o zmiennych współczynnikach).

Mamy wyznaczyć dwie funkcje, a jest tylko jedno równanie. Jednakże mamy dowolność przy wyborze c_1 i c_2 (szukamy *jakiegokolwiek* rozwiązania równania niejednorodnego). Możemy więc przyjąć te funkcje tak, aby $\dot{c}_1x_1 + \dot{c}_2x_2 = 0$. Liczymy przy tym założeniu pochodną $\dot{x}^{(s)}(t) = c_1\dot{x}_1 + c_2\dot{x}_2$. Dla drugiej pochodnej otrzymamy $\ddot{x}^{(s)}(t) = c_1\ddot{x}_1 + c_2\ddot{x}_2 + \dot{c}_1\dot{x}_1 + \dot{c}_2\dot{x}_2$ (odpowiednie założenie o c_1 i c_2 pozwoliło uniknąć drugich pochodnych tych funkcji). Podstawiamy wzory do równania niejednorodnego i dostajemy równanie na $\dot{c}_1$, $\dot{c}_2$, które, razem z założonym wcześniej równaniem, daje układ równań

$$\begin{cases} \dot{c}_1x_1 + \dot{c}_2x_2 = 0, \\ \dot{c}_1\dot{x}_1 + \dot{c}_2\dot{x}_2 = r(t). \end{cases}$$

Możemy z niego, z uwagi na liniową niezależność x_1 i x_2, wyznaczyć $c_1(t)$ i $c_2(t)$ (o ile będziemy umieli policzyć odpowiednie całki).

Wadą powyższej metody jest liczenie często skomplikowanych całek. W wielu przypadkach, jeśli funkcja $r(t)$ jest odpowiedniej postaci, można *zgadnąć* rozwiązanie szczególne. W przypadku równań o stałych współczynnikach $a\ddot{x} + b\dot{x} + cx = r(t)$, dla kilku często spotykanych postaci funkcji $r(t)$, znamy sposoby postępowania.

Przypadek 1. $r(t) = d_0 + d_1 t + \ldots + d_n t^n$.
Poszukujemy rozwiązań postaci

$$x^{(s)}(t) = \begin{cases} \alpha_0 + \alpha_1 t + \ldots + \alpha_n t^n, & \text{jeśli } c \neq 0, \\ t(\alpha_0 + \alpha_1 t + \ldots + \alpha_n t^n), & \text{jeśli } c = 0. \end{cases}$$

Wstawiamy do równania i porównując współczynniki przy tych samych potęgach t wyznaczamy α_i. Jeśli $c = 0$ i $b = 0$, to mamy równanie $a\ddot{x} = r(t)$, które po prostu dwukrotnie całkujemy.

Przypadek 2. $r(t) = \mathrm{e}^{At}(d_0 + d_1 t + \ldots + d_n t^n)$.
Poszukujemy rozwiązań postaci $x^{(s)}(t) = \mathrm{e}^{At}u(t)$. Podstawiając do równania dostajemy Przypadek 1 dla równania na $u(t)$.

Przypadek 3. $r(t) = \mathrm{e}^{\alpha t}(g_n(t)\cos\beta t + h_m(t)\sin\beta t)$, gdzie g_n, h_m są wielomianami stopnia n i m.
Poszukujemy rozwiązań postaci

$$x^{(s)}(t) = t^s \mathrm{e}^{\alpha t}(u_K(t)\cos\beta t + v_K(t)\sin\beta t),$$

gdzie s jest krotnością $\alpha + i\beta$ jako pierwiastka równania charakterystycznego (jeśli nie jest pierwiastkiem, to $s = 0$), a u_K, v_K są wielomianami stopnia $K = \max\{n, m\}$. Współczynniki wielomianów u_K i v_K znajdujemy po podstawieniu $x^{(s)}(t)$ do równania i przyrównaniu współczynników przy odpowiednich potęgach t.

5.2. Istnienie rozwiązań układów liniowych

Definicja 5.2. Równanie różniczkowe postaci $\dot{x} = A(t)x + b(t)$, gdzie $A(t) = [a_{ij}(t)]_{i,j=1,\ldots,n}$, $b(t) = [b_j(t)]_{j=1,\ldots,n}$, a_{ij}, b_j są funkcjami skalarnymi, nazywamy *układem równań różniczkowych liniowych* (w skrócie: *układem liniowym* lub *równaniem liniowym*).

Rozważać będziemy zagadnienie początkowe postaci

$$\begin{cases} \dot{x} = A(t)x + b(t), \\ x(t_0) = x_0. \end{cases} \tag{5.3}$$

Zaczniemy od twierdzenia, z którego skorzystamy w dowodzie istnienia globalnego rozwiązania układu liniowego.

Twierdzenie 5.2. *Niech $I \subset \mathbb{R}$ będzie otwartym przedziałem. Załóżmy, że funkcja $f : I \times \mathbb{R}^n \to \mathbb{R}^n$ jest ciągła oraz spełnia nierówność*

$$\|f(t,x)\| \leqslant a(t)\|x\| + d(t)$$

dla wszystkich $t \in I$ i $x \in \mathbb{R}^n$, gdzie $a(t)$ i $d(t)$ są ciągłymi i nieujemnymi funkcjami rzeczywistymi. Wtedy dla wszystkich $t_0 \in I$ i $x_0 \in \mathbb{R}^n$ zagadnienie początkowe

$$\begin{cases} \dot{x} = f(t,x), \\ x(t_0) = x_0 \end{cases}$$

ma rozwiązanie określone na I.

Dowód. Niech $Q = I \times \mathbb{R}^n$ i niech $x(t)$ będzie rozwiązaniem wysyconym określonym na przedziale $J = (\alpha, \beta) \subset I$. Pokażemy, że $J = I$.

Przypuśćmy, że nie jest to prawdą. Wtedy jeden z punktów α, β należy do I; dla ustalenia uwagi niech $\beta \in I$. Z twierdzenia o przedłużaniu ciąg $(t, x(t))$ dąży do brzegu zbioru Q lub jest nieograniczony, gdy $t \to \beta^-$. Ponieważ β jest skończone, a Q nie jest ograniczony po zmiennej x, to musi być $\|x(t)\| \to \infty$, gdy $t \to \beta^-$.

Z drugiej strony rozwiązanie możemy zapisać w postaci całkowej

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds,$$

skąd, korzystając z założenia o f, dla dowolnego $t \in [t_0, \beta]$ mamy

$$\begin{aligned} \|x(t)\| &\leqslant \|x_0\| + \int_{t_0}^{t} \|f(s, x(s))\|\, ds \\ &\leqslant \|x_0\| + \int_{t_0}^{t} a(s)\|x(s)\| + d(s)\, ds \\ &\leqslant C + A \int_{t_0}^{t} \|x(s)\|\, ds, \end{aligned}$$

gdzie $A = \sup\limits_{s \in [t_0, \beta]} a(s)$ i $C = \|x_0\| + \int_{t_0}^{\beta} d(s)\, ds$. Ponieważ $[t_0, \beta] \subset I$, a funkcje $a(t)$ i $d(t)$ są ciągłe na tym przedziale, to wartości A i C są skończone. Zatem z lematu Gronwalla dostajemy

$$\|x(t)\| \leqslant C\mathrm{e}^{A(t-t_0)} \leqslant C\mathrm{e}^{A(\beta - t_0)}.$$

Prawa strona w powyższym oszacowaniu nie zależy od t, zatem funkcja $x(t)$ jest ograniczona gdy $t \to \beta$. Dostaliśmy sprzeczność z wynikiem pierwszej części dowodu, zatem $J = I$. □

Rozwiązania układu liniowego istnieją globalnie tam, gdzie są ciągłe współczynniki układu. Mówi o tym następujące twierdzenie, które łatwo wynika z twierdzenia (5.2).

Twierdzenie 5.3. *Jeśli $A(t)$ i $b(t)$ są ciągłe na przedziale otwartym $I \subset \mathbb{R}$, to przez każdy punkt (t_0, x_0) zbioru $Q = I \times \mathbb{R}^n$ przechodzi dokładnie jedna krzywa całkowa zagadnienia* (5.3) *z maksymalnym przedziałem istnienia rozwiązania równym I.*