

# 1. Błędy i arytmetyka zmiennoprzecinkowa

Liczby zmiennoprzecinkowe (binarne) to liczby w postaci

$$\pm m \cdot 2^c,$$

gdzie  $m$  jest liczbą z przedziału  $[1, 2)$ , a  $c$  liczbą całkowitą. W praktyce (np. w rzeczywistych komputerach) żądamy, aby mantysa miała określoną liczbę miejsc po przecinku (w postaci ułamka binarnego), a cecha określoną liczbę bitów. Rzeczywiście stosowane standardy są nieco bardziej skomplikowane (definiują zapis zera, nieskończoności, tzw. liczby zdenormalizowane oraz wartości wyjątkowe nie przedstawiające liczb). Najczęściej stosowany standard w informatyce nosi nazwę IEEE 754.

**Twierdzenie 1.1.** Jeżeli  $x > y$  są znormalizowanymi liczbami zmiennoprzecinkowymi, oraz  $2^{-q} \leq 1 - y/x \leq 2^{-p}$ , to przy obliczaniu  $x - y$  tracimy co najmniej  $p$  i co najwyżej  $q$  bitów precyzji.

**Twierdzenie 1.2.** Przy obliczaniu sumy  $\sum_{k=0}^n x_k$  liczb maszynowych najbardziej naturalnym sposobem popełniamy błąd względny równy co najwyżej  $(1 + \epsilon)^n - 1 \approx n\epsilon$ , gdzie  $\epsilon$  to dokładność maszynowa.

**Zad. 1.1** Ile jest dodatnich liczb zmiennoprzecinkowych o 3-cyfrowej mantysie i cesze  $c$  spełniającej  $-2 \leq c \leq 1$ . Przyjmij, że arytmetyka używa systemu dwójkowego. Wypisz kilka kolejnych takich liczb.

**Zad. 1.2** Stosowane w arytmetyce komputerowej liczby zmiennoprzecinkowe pojedynczej precyzji standardu IEEE 754 mają 23-bitową mantysę i 8-bitową cechę (tj. wraz z bitem znaku, 32 bity). Jaka jest najmniejsza i największa liczba dodatnia możliwa do zapisania w tym systemie?

**Zad. 1.3** Ile cyfr istotnych, poprawnych i znaczących mają liczby:

- a)  $0,0025624 \pm 4 \cdot 10^{-6}$ ;      b)  $0,0011995 \pm 3 \cdot 10^{-5}$ ;      c)  $0,002223 \pm 6 \cdot 10^{-5}$ ;

**Zad. 1.4** Oblicz

$$f(x) = \frac{1 - \cos x}{x^2}$$

dla  $x$  bliskich 0 według powyższego wzoru i według algorytmu:

$$w = \frac{\sin x/2}{x/2}$$
$$f(x) = \frac{ww}{2}.$$

Pokaż, że matematycznie są to identyczne metody. Która jest lepsza numerycznie?

**Zad. 1.5** Niech

$$y_n = \int_0^1 \frac{t^n}{t+5} dt.$$

Można pokazać wzór rekurencyjny  $y_n + 5y_{n-1} = \frac{1}{n}$ .

- a) Pokaż, że podany wzór rekurencyjny jest prawdziwy.  
b) Przyjmując  $y_0 \cong 0,182$  oblicz  $y_1, \dots, y_8$ .  
c) Przyjmując  $y_8 \cong 0,019$  oblicz  $y_7, \dots, y_0$ .  
d) Które wyniki są bardziej wiarygodne?

**Zad. 1.6** Oblicz przybliżenie pochodnej funkcji  $x \mapsto x^2$  w ustalonym punkcie ze wzoru

$$f'(x) \cong \frac{f(x+h) - f(x)}{h}$$

dla różnych, coraz mniejszych wartości  $h$ . Co zauważasz?

**Zad. 1.7** Obliczyć  $W(x)$  dla  $x \in [0,99; 1,01]$  co 0,001, gdy

- a)  $W(x) = (x - 1)^8$ ;  
 b)  $W(x) = x^8 - 8x^7 + 28x^6 - 56x^5 + 70x^4 - 56x^3 + 28x^2 - 8x + 1$ ;  
 c)  $W(x) = ((((((x - 8)x + 28)x - 56)x + 70)x - 56)x + 28)x - 8)x + 1$ .

Który wzór jest najlepszy? Zauważ, że za każdym razem jest to ten sam wielomian.

**Zad. 1.8** Dane jest równanie:

$$x^2 + px + q = 0.$$

Dla podanych współczynników oblicz pierwiastki korzystając z algorytmu szkolnego oraz z następującego algorytmu:

Jeżeli  $p > 0$ , to  $x_1 = \frac{-p - \sqrt{p^2 - 4q}}{2}$ , w przeciwnym przypadku  $x_1 = \frac{-p + \sqrt{p^2 - 4q}}{2}$ . Następnie przyjmij  $x_2 = \frac{q}{x_1}$ .

- a)  $p = 0,87002$ ;  $q = 0,174$ ;      b)  $p = -2 \cdot 10^7$ ;  $q = 0,2$ ;  
 c)  $p = 0,999962 \cdot 10^7$ ;  $q = 0,99992$ .

**Zad. 1.9** Załóżmy, że przy obliczaniu  $\sqrt{x^2 + 1} - 1$  chcemy stracić co najwyżej dwa bity precyzji. Jakie wartości  $x$  należy wykluczyć.

**Zad. 1.10** Napisz program, który obliczy

$$f(x) = \sqrt{x^2 + 1} - 1$$

$$g(x) = \frac{x^2}{\sqrt{x^2 + 1} + 1}$$

- a) Pokaż, że  $f(x) = g(x)$ .  
 b) Napisz program Obliczający wartość  $f$  i  $g$  dla coraz mniejszych wartości  $x$ . Co zauważasz? Które wartości są bardziej wiarygodne?

**Zad. 1.11** Używając przybliżeń do sześciu cyfr dziesiętnych (z poprawnym zaokrągleniem) oblicz  $(10000 + \pi) + e$  używając:

- algorytmu naiwnego;
- algorytmu Kahana.

Porównaj wyniki. Który jest bliższy prawdziwej wartości? Zauważ, że dodawanie liczb zmiennoprzecinkowych nie jest przemienne!

**Zad. 1.12** Porównaj wyniki dla różnych algorytmów sumowania uruchamiając w Julii Kod źródłowy 1.

**Zad. 1.13** W arytmetyce zmiennoprzecinkowej znajdź oszacowanie błędu względnego wyniku obliczeń (przyjmij, że  $x$ ,  $y$  i  $z$  są liczbami rzeczywistymi, nie maszynowymi):

- a)  $x(y + z)$ ;      b)  $xyz$ ;  
 c)  $x + y + z$ .

---

## 2. Współczynnik uwarunkowania zadania

---

Współczynnik uwarunkowania zadania pokazuje jak błąd względny danych wpływa na błąd względny wyniku. Z reguły definiujemy go jako stosunek tych dwóch wartości. Poznamy współczynnik uwarunkowania zadania obliczenia wartości funkcji i współczynnik uwarunkowania zadania znalezienia rozwiązania układu równań liniowych.

```

using LinearAlgebra

function KahanSum(v::Vector{T}) where {T<:Number}
    s = zero(T)
    c = zero(T)
    for i in 1:length(v)
        y = v[i] - c
        t = s + y
        c = (t - s) - y
        s = t
    end
    return s
end

function NaiveSum(v::Vector{T}) where {T<:Number}
    s = zero(T)
    for i in 1:length(v)
        s += v[i]
    end
    return s
end

n = 200000000
v = fill(1.5f0, n)
println(NaiveSum(v)/n)
println(KahanSum(v)/n)
println(sum(v)/n)

```

## Kod źródłowy 1: Sumowanie Kahana

Przyjmijmy, że argument  $x$  dany jest z błędem względnym  $\delta$  wtedy błąd względny wartości funkcji możemy oszacować następująco:

$$\left| \frac{f(x(1+\delta)) - f(x)}{f(x)} \right| \cong \left| \frac{f'(x)\delta x}{f(x)} \right| = \left| \frac{f'(x)x}{f(x)} \right| |\delta|.$$

Wartość  $\left| \frac{f'(x)x}{f(x)} \right|$  to poszukiwany współczynnik uwarunkowania.

Dla funkcji wielu zmiennych wzór jest analogiczny ( $x$  jest wektorem  $n$  zmiennych):

$$\left| \frac{F(x(1+\delta)) - F(x)}{F(x)} \right| \leq \sum_{k=1}^n \left| \frac{\frac{\partial F}{\partial x_k}(x)x_k}{F(x)} \right| |\delta_{x_k}|.$$

Współczynnik uwarunkowania zadania znalezienia rozwiązania układu równań  $Ax = b$  możemy oszacować:

$$\left\| \frac{A^{-1}(x + \varepsilon) - A^{-1}x}{A^{-1}x} \right\| = \left\| \frac{A^{-1}\varepsilon}{A^{-1}x} \right\| \leq \left\| \frac{x}{A^{-1}x} \right\| \left\| \frac{A^{-1}x}{x} \right\| \left\| \frac{\varepsilon}{x} \right\| \leq \|A\| \|A^{-1}\| \left\| \frac{\varepsilon}{x} \right\|,$$

zatem współczynnik uwarunkowania jest oszacowany z góry przez  $\|A\| \|A^{-1}\|$  (bierzemy dowolną normę na przestrzeni macierzy, mimo że powyższe rozumowanie używa norm operatorowych – wszystkie normy są równoważne dopóki rozważamy macierze tego samego wymiaru).

Zadanie dla którego współczynnik uwarunkowania jest mały nazywamy dobrze postawionym, takie dla którego jest duży, zadaniem źle uwarunkowanym. Jeżeli współczynnik jest nieskończony, to mówimy o zadaniu źle postawionym.

- Zad. 2.1** Oblicz współczynnik uwarunkowania zadania dla zadania obliczenia iloczynu skalarnego wektorów  $[1, 2, 3]$  oraz  $[4, 5, -5]$  (traktować iloczyn skalarny jako funkcję o 6 zmiennych).
- Zad. 2.2** Oblicz współczynnik uwarunkowania zadania dla zadania obliczenia wartości funkcji

- a)  $f(x) = x^2 - 1$ ;                      b)  $f(x) = \sqrt{x}$ ;                      c)  $f(x, y) = \frac{\sqrt{x}}{y}$ ;  
d)  $f(x, y) = 2x^2 - y$ ;                      e)  $f(x, y) = y\sqrt{x}$ ;                      f)  $f(x, y) = y^2 + x$ ;  
g)  $f(x, y) = (x + y)x$ ;                      h)  $f(x, y) = \frac{2x}{\sqrt{y}}$ .

Dla jakich wartości argumentu zadanie jest dobrze (źle) uwarunkowane? Dla jakich jest źle postawione?

**Zad. 2.3** Dla coraz mniejszych wartości  $\varepsilon$  oblicz z użyciem języka Julia współczynnik uwarunkowania macierzy  $A + \varepsilon I$ , gdzie

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix},$$

a  $I$  jest identycznością.

**Rozwiązanie.**

```
using LinearAlgebra
```

```
A=[1 2; 1 2]
cond(A+1e-1*I)
cond(A+1e-2*I)
cond(A+1e-3*I)
```

**Zad. 2.4** Używając języka Julia oblicz dla kilku wartości  $n$  współczynnik uwarunkowania  $n$ -tej macierzy Hilberta  $H_n = \left[ \frac{1}{i+j-1} \right]_{i,j=1}^n$ . Spróbuj numerycznego rozwiązania układu równań, którego macierzą jest macierz Hilberta odpowiednio dużego wymiaru (macierz ta jest zawsze nieosobliwa). Co zauważasz?

**Rozwiązanie.**

```
using GenericSVD
using LinearAlgebra
```

```
Hilbert(n::T) where {T<:Integer} = [T(1)/T(x+y) for x = 1:n, y = 1:n]
c(n::T) where {T<:Integer} = cond(Hilbert(n))
```

### 3. Schemat Hornera

Schemat Hornera pozwala w szybki sposób obliczyć wartość zadanego wielomianu w punkcie, wielomian i resztę z dzielenia wielomianu przez dwumian oraz pochodne znormalizowane w zadanym punkcie. Metoda ta opiera się na poniższej równości:

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = (\dots ((a_n x + a_{n-1})x + a_{n-2})x + a_{n-3}) \dots + a_1)x + a_0$$

#### Wartość wielomianu

Aby obliczyć wartość wielomianu  $P(x) = \sum_{k=0}^n a_k x^k$  w punkcie  $a$  korzystamy z procedury rekurencyjnej:

- $b_n = a_n$
- $b_k = ab_{k+1} + a_k, k = n-1, \dots, 0$

Tak obliczona wartość  $b_0$  jest wartością  $P$  w punkcie  $a$ . Obliczenia z reguły organizujemy w tabelkę:

$$\begin{array}{c|c|c|c|c|c} & a_n & a_{n-1} & \dots & a_1 & a_0 \\ \hline a & b_n & b_{n-1} & \dots & b_1 & b_0 \end{array}$$

Dla przykładu, obliczając wartość wielomianu  $P(x) = x^3 - 6x^2 + 11x - 6$  w punkcie 4, utworzymy tabelkę:

	1	-6	11	-6
4	1	-2	3	6

Zatem  $P(4) = 6$ .

Julia zawiera wbudowane makro `@evalpoly`, które tworzy kod obliczający wartość wielomianu w zadanym punkcie z użyciem schematu Hornera:

```
@evalpoly(z, c0, c1, c2, ...)
```

Przez kompilator wygenerowany zostanie kod, który obliczy wartość wielomianu  $c_0 + c_1z + c_2z^2 + \dots$  w punkcie  $z$ , dla przykładu:

```
julia> @evalpoly(2, 1, 2, 3)
17

julia> f(x, a, b, c) = @evalpoly(x, a, b, c)
f (generic function with 1 method)

julia> @code_native f(1,1,1,1)
.text
;   ┌ @ REPL[1]:1 within `f'
;   └─┬─ @ math.jl:150 within `macro expansion'
;     └─┬─ @ promotion.jl:397 within `muladd'
;       └─┬─ @ REPL[1]:1 within `*'
;         └─ imulq    %rdi, %rcx
;       └─┬─ LL
;         └─┬─ @ int.jl:53 within `muladd'
;           └─ leaq    (%rcx,%rdx), %rax
;         └─┬─ L
;           └─┬─ @ promotion.jl:397 within `muladd'
;             └─┬─ @ int.jl:54 within `*'
;               └─ imulq    %rdi, %rax
;             └─┬─ L
;               └─┬─ @ int.jl:53 within `+'
;                 └─ addq    %rsi, %rax
;             └─┬─ LLL
;               └─ retq
;             └─ L

julia> @code_native f(1,1,1,1.0)
.text
;   ┌ @ REPL[8]:1 within `f'
;   └─┬─ @ math.jl:150 within `macro expansion'
;     └─┬─ @ promotion.jl:348 within `muladd'
;       └─┬─ @ promotion.jl:290 within `promote'
;         └─┬─ @ promotion.jl:268 within `_promote'
;           └─┬─ @ number.jl:7 within `convert'
;             └─┬─ @ REPL[8]:1 within `Type'
;               └─ vcvtsi2sdq    %rdi, %xmm1, %xmm2
;                 vcvtsi2sdq    %rdx, %xmm1, %xmm1
;               └─┬─ LLLLLL
;                 └─┬─ @ float.jl:404 within `muladd'
;                   └─ vfmadd231sd    %xmm0, %xmm2, %xmm1
;                 └─┬─ L
;                   └─┬─ @ promotion.jl:348 within `muladd'
```

```

; | | | | @ promotion.jl:290 within `promote'
; | | | | @ promotion.jl:268 within `_promote'
; | | | | @ number.jl:7 within `convert'
; | | | | @ float.jl:60 within `Type'
; | | | | vcvtsi2sdq %rsi, %xmm3, %xmm0
; | | | LLLL
; | | | @ promotion.jl:348 within `muladd' @ float.jl:404
; | | | vfmadd213sd %xmm0, %xmm2, %xmm1
; | LL
; | vmovapd %xmm1, %xmm0
; | retq
; | nop
; | L

```

**Zad. 3.1** Oblicz wartość wielomianu  $P$  w punkcie  $a$ :

- a)  $P(x) = 5x^4 - 4x^3 + 7x^2 - 8x, a = 2$       b)  $P(x) = x^5 + 3x^4 + x^3 + x^2 - 1, a = -2$   
c)  $P(x) = x^5 + 2x^4 + 3x^3 + 4x^2 + 5x^1 + 6, a = -2$       d)  $P(x) = x^5 - 2x^4 + 3x^3 - 4x^2 + 5x^1 - 6, a = 3$   
e)  $P(x) = x^6 - 5x^3 - 6, a = -2$

## Dzielenie przez dwumian

Obliczanie wartości wielomianu w punkcie daje także od razu iloraz wielomianu przez dwumian. Mamy bowiem:

**Fakt 3.1.** Niech  $P(x) = \sum_{k=0}^n a_k x^k$  i niech liczby  $b_n$  spełniają zależność podaną powyżej. Wtedy

$$P(x) = \left( \sum_{k=0}^{n-1} b_{k+1} x^k \right) (x - a) + b_0$$

Z powyżej obliczonego przykładu wynika, że  $x^3 - 6x^2 + 11x - 6 = (x^2 - 2x + 3)(x - 4) + 6$ .

**Zad. 3.2** Oblicz wielomian będący ilorazem wielomianu  $P(x)$  przez zadany dwumian  $D(x)$ :

- a)  $P(x) = 2x^3 + 4x^2 - 3x + 4, D(x) = x - 3;$       b)  $P(x) = 5x^4 - 2x^3 + x^2 + 7, D(x) = x + 1;$   
c)  $P(x) = x^4 - 4x^2 - 5x + 1, D(x) = x + 7;$

## Pochodne znormalizowane

**Definicja 3.2.** Pochodną znormalizowaną ( $n$ -tą) funkcji  $f$  w punkcie  $x$  nazywamy wyrażenie  $\frac{f^{(n)}(x)}{n!}$ .

*Uwaga 3.3.* Pochodne znormalizowane to współczynniki szeregu Taylora funkcji  $f$  w punkcie  $x$ .

**Fakt 3.4.** Pochodne znormalizowane można obliczyć korzystając wielokrotnie ze schematu Hornera (patrz przykład poniżej).

Dla przykładu znajdziemy wielomian Taylora dla wielomianu  $P(x) = x^3 - 6x^2 + 11x - 6$  w punkcie 4. Tworzymy tabelkę:

	1	-6	11	-6
4	1	-2	3	⑥
4	1	2	⑪	
4	1	⑥		
4	①			

Stąd

$$x^3 - 6x^2 + 11x - 6 = (x - 4)^3 + 6(x - 4)^2 + 11(x - 4) + 6.$$

**Zad. 3.3** Udowodnij Fakt 3.4.

**Zad. 3.4** Rozwiń wielomian  $P(x)$  w szereg Taylora wokół zadanego punktu  $x_0$ :

$$\text{a) } P(x) = 7x^3 - 5x^2 - 2x + 1, \quad x_0 = 3;$$

$$\text{b) } P(x) = x^4 - x^3 + 3x - 2, \quad x_0 = -2;$$

$$\text{c) } P(x) = 4x^4 + 5x^3 - 5x^2 - 4, \quad x_0 = 1;$$

## 4. Interpolacja wielomianowa

### Zagadnienie interpolacyjne Lagrange'a

Dany jest ciąg **różnych** punktów  $x_0, \dots, x_n$  i odpowiadający im ciąg wartości  $y_0, \dots, y_n$ . Zagadnienie interpolacyjne Lagrange'a to znalezienie wielomianu  $L(x)$  stopnia co najwyżej  $n$  takiego, że  $L(x_i) = y_i$ .

*Uwaga 4.1.* W praktyce często  $y_i = f(x_i)$  del pewnej funkcji  $f$ , którą łatwo policzyć w  $x_i$ , ale nie tak łatwo dla dowolnego argumentu.

**Fakt 4.2.** Przy powyższych założeniach istnieje co najwyżej jeden taki wielomian  $L(x)$ .

*Dowód.* Niech  $L$  i  $L'$  rozwiązuja zagadnienie Lagrange'a. Wtedy  $L - L'$  jest wielomianem stopnia co najwyżej  $n$  takim, że  $(L - L')(x_i) = 0$ . Zatem  $L - L'$  ma  $n + 1$  pierwiastków, więc musi być wielomianem zerowym. Stąd  $L = L'$ .  $\square$

**Fakt 4.3.** Zagadnienie interpolacyjne Lagrange'a ma rozwiązanie.

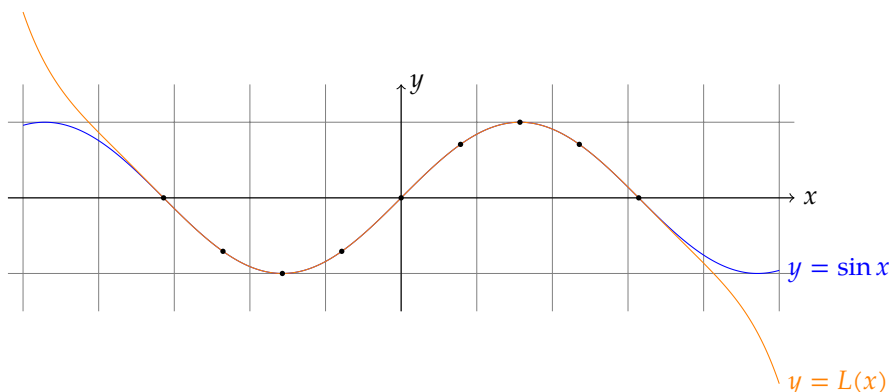
*Dowód.* Niech  $q_i(x) = (x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)$ . Wtedy

$$q_i(x_i) \neq 0 \text{ oraz } q_i(x_j) = 0 \text{ dla } i \neq j.$$

Zatem wielomian  $l_i(x) = \frac{q_i(x)}{q_i(x_i)}$  spełnia

$$l_i(x_i) = 1 \text{ oraz } l_i(x_j) = 0 \text{ dla } i \neq j.$$

Zatem  $L(x) = \sum_{i=0}^n y_i l_i(x)$  spełnia stawiane przed  $L$  wymagania.  $\square$

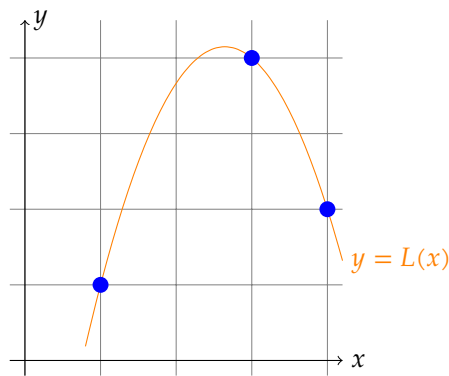


**Przykład.** Będziemy poszukiwać wartości nieznanej funkcji wyznaczając wielomian interpolacyjny Lagrange'a mając dane trzy punkty na wykresie:  $(1, 1)$ ,  $(3, 4)$ ,  $(4, 2)$ .

Wprost ze wzoru mamy:

$$\begin{aligned} l_0(x) &= \frac{(x-3)(x-4)}{(1-3)(1-4)} = \frac{(x-3)(x-4)}{6}; \\ l_1(x) &= \frac{(x-1)(x-4)}{(3-1)(3-4)} = \frac{(x-1)(x-4)}{-2}; \\ l_2(x) &= \frac{(x-1)(x-3)}{(4-1)(4-3)} = \frac{(x-1)(x-3)}{3}. \end{aligned}$$

$$\text{Zatem } L(x) = l_0(x) + 4l_1(x) + 2l_2(x) = -\frac{7x^2}{6} + \frac{37x}{6} - 4.$$



Niestety taki sposób znajdowania wielomianu interpolacyjnego Lagrange'a ma poważną wadę. Proszę sobie wyobrazić, że udało nam się uzyskać jeszcze jeden punkt na wykresie: (2, 5). Wtedy wszystkie obliczenia trzeba będzie powtórzyć. Wady tej nie ma sposób prezentowany poniżej. Co prawda ilość obliczeń jest troszkę większa, to dołożenie kolejnego punktu nie powoduje, że obliczenia trzeba powtarzać.

**Twierdzenie 4.4.** Wielomian interpolacyjny Lagrange'a  $L(x)$  możemy zapisać w tzw. postaci Newtona:

$$L(x) = \sum_{i=0}^n b_i p_i(x),$$

gdzie

$$\begin{aligned} p_0(x) &= 1, \\ p_1(x) &= (x - x_0), \\ p_2(x) &= (x - x_0)(x - x_1), \\ &\dots \\ p_n(x) &= (x - x_0)(x - x_1) \dots (x - x_{n-1}) \\ b_i &= f[x_0, \dots, x_i], \end{aligned}$$

gdzie z kolei  $f[x_0, \dots, x_i]$  to tzw. ilorazy różnicowe dane rekurencyjną zależnością:

$$\begin{aligned} f[x_i] &= f(x_i) = y_i, \\ f[x_l, \dots, x_{l+k}] &= \frac{f[x_{l+1}, \dots, x_{l+k}] - f[x_l, \dots, x_{l+k-1}]}{x_{l+k} - x_l}. \end{aligned}$$

Wartości ilorazów różnicowych z reguły zapisujemy w postaci wygodnej do obliczeń trójkątnej tablicy, np:

$x_0$	$f(x_0)$			
		$f[x_0, x_1]$		
$x_1$	$f(x_1)$		$f[x_0, x_1, x_2]$	
		$f[x_1, x_2]$		$f[x_0, x_1, x_2, x_3]$
$x_2$	$f(x_2)$		$f[x_1, x_2, x_3]$	$f[x_0, x_1, x_2, x_3, x_4]$
		$f[x_2, x_3]$		$f[x_1, x_2, x_3, x_4]$
$x_3$	$f(x_3)$		$f[x_2, x_3, x_4]$	
		$f[x_3, x_4]$		
$x_4$	$f(x_4)$			

Każdy iloraz różnicowy w powyższej tabeli powstaje przez podzielenie różnicy dwóch sąsiadujących z nim z lewej strony przez różnicę odpowiednich wartości  $x_i$ , np:

$$f[x_1, x_2, x_3] = \frac{f[x_2, x_3] - f[x_1, x_2]}{x_3 - x_1}.$$

Interesujące nas ilorazy – te które występują w postaci Newtona wielomianu interpolacyjnego, zaznaczono w tabeli pogrubioną czcionką. Zauważmy, że dołożenie kolejnej pary wartości  $(x_5, f(x_5))$  zmusi nas do obliczenia jedynie wartości wzdłuż dolnej krawędzi nowej tablicy.

Bez dowodu podamy:



**Twierdzenie 4.5.** Niech  $f$  będzie funkcją klasy  $C^{n+1}$  na przedziale  $[a, b]$  i niech  $L$  będzie wielomianem interpolacyjnym Lagrange'a dla  $f$  opartym na  $n + 1$  różnych węzłach  $(x_0, \dots, x_n)$  leżących w przedziale  $[a, b]$  (tzn.  $L(x_i) = f(x_i)$ ). Wtedy dla  $x \in [a, b]$ :

$$|f(x) - L(x)| \leq \frac{M_{n+1}}{(n+1)!} |(x - x_0)(x - x_1) \dots (x - x_n)|,$$

gdzie  $M_{n+1} = \max_{x \in [a, b]} |f^{(n+1)}(x)|$ .

Twierdzenie to ma dużo wspólnego z twierdzeniem Lagrange'a o wartości średniej.

**Zad. 4.1** Dla par punktów znaleźć wielomiany interpolacyjne Lagrange'a w postaci Newtona odpowiedniego stopnia.

a)  $\begin{array}{c|c|c|c} x_i & 1 & 2 & 3 \\ \hline y_i & 5 & 8 & 2 \end{array}$

b)  $\begin{array}{c|c|c|c|c} x_i & 1 & 2 & 5 & 6 \\ \hline y_i & 0 & 4 & 6 & -2 \end{array}$

**Rozwiązanie.**

a) Tworzymy tablicę ilorazów różnicowych:

$$\begin{array}{cccc} 1 & 5 & & \\ & & 3 & \\ 2 & 8 & & \frac{-9}{2} \\ & & -6 & \\ 3 & 2 & & \end{array}$$

Z tabeli odczytujemy wielomian interpolacyjny:  $L(x) = 5 + 3(x - 1) - \frac{9}{2}(x - 1)(x - 2)$ .

b) Tworzymy tablicę ilorazów różnicowych:

$$\begin{array}{cccccc} 1 & 0 & & & & \\ & & 4 & & & \\ 2 & 4 & & -\frac{5}{6} & & \\ & & \frac{2}{3} & & -\frac{4}{15} & \\ 5 & 6 & & \frac{-13}{6} & & \\ & & -8 & & & \\ 6 & -2 & & & & \end{array}$$

Odczytujemy wielomian interpolacyjny:  $L(x) = 4(x - 1) - \frac{5}{6}(x - 1)(x - 2) - \frac{4}{15}(x - 1)(x - 2)(x - 5)$ .

**Zad. 4.2** Za pomocą wzoru interpolacyjnego Lagrange'a oblicz przybliżoną wartość  $\sqrt{115}$ . Oszacuj błąd otrzymanego w ten sposób przybliżenia?

**Rozwiązanie.** Tworzymy tablicę interpolującą funkcję pierwiastkową w węzłach 100, 121, 144:

$$\begin{array}{cc} 100 & 10 \\ & \frac{1}{21} \\ 121 & 11 & -\frac{1}{21 \cdot 22 \cdot 23} \\ & \frac{1}{23} \\ 144 & 12 \end{array}$$

Wielomian interpolacyjny to  $L(x) = 10 + \frac{x-100}{21} - \frac{(x-100)(x-121)}{21 \cdot 22 \cdot 23}$ . Więc

$$\sqrt{115} \cong L(115) = 10 + \frac{15}{21} + \frac{6 \cdot 15}{21 \cdot 22 \cdot 23} \cong 10,7228.$$

Jednocześnie pierwiastek jest funkcją klasy  $C^\infty$ . Aby skorzystać z twierdzenia, musimy policzyć trzecią pochodną:

$$(x^{1/2})''' = \left( \frac{1}{2} x^{-1/2} \right)'' = - \left( \frac{1}{4} x^{-3/2} \right)' = \frac{3}{8} x^{-5/2}.$$

Mamy więc, że  $\max_{x \in [100, 144]} (x^{1/2})''' = \frac{3}{8} 100^{-5/2} = 3/8 \cdot 10^{-5}$ . Z twierdzenia dostajemy, że

$$|\sqrt{115} - L(x)| \leq \frac{3 \cdot 10^{-5}}{8 \cdot 3!} 15 \cdot 6 \cdot 29 < 0,0017.$$

**Zad. 4.3** Znajdź wielomian interpolacyjny Lagrange'a dla funkcji  $f(x) = 1/x$  i 3 węzłów:  $2, \frac{5}{2}, 4$ .

**Rozwiązanie.**

$$\begin{array}{cccc} 2 & & \frac{1}{2} & \\ & & & -\frac{1}{5} \\ \frac{5}{2} & & \frac{2}{5} & & \frac{1}{20} \\ & & & -\frac{1}{10} \\ 4 & & \frac{1}{4} & \end{array}$$

Zatem poszukiwany wielomian to  $L(x) = \frac{1}{2} - \frac{1}{5}(x-2) + \frac{1}{20}(x-2)(x-\frac{5}{2})$ .

**Zad. 4.4** Mając dane (przybliżone) wartości funkcji  $f(x) = e^x$ :

$x_i$	0	$\frac{1}{2}$	1	2
$y_i$	1	1,64872	2,71828	7,38906

Oblicz:

- $f(1/4)$  korzystając z interpolacji liniowej (węzły  $0, \frac{1}{2}$ );
- $f(3/4)$  korzystając z interpolacji liniowej (węzły  $\frac{1}{2}, 1$ );
- $f(1/4), f(3/4)$  korzystając z interpolacji kwadratowej w węzłach  $0, \frac{1}{2}, 2$ .

Które przybliżenia są dokładniejsze?

**Rozwiązanie.** Dla interpolacji liniowej dostajemy łatwo:

- $f(1/4) \cong 1,32436$ ;
- $f(3/4) \cong 2,1835$ .

Aby otrzymać wielomian interpolacyjny, tworzymy tablicę różnic:

0	1	
		1,29744
$\frac{1}{2}$	1,64872	1,26473
		3,82689
2	7,38906	

Stąd  $L(1/4) \cong 1,24531, L(3/4) \cong 2,21022$ .

Szacując błędy mamy odpowiednio dla wartości  $1/4$ :

- błąd interpolacji liniowej:  $\frac{e^{1/2}}{2!} (1/4)^2 < 0,052$
- błąd interpolacji kwadratowej:  $\frac{e^2}{3!} (1/4)(1/4)(7/4) < 0,14$

Dla  $3/4$ :

- błąd interpolacji liniowej:  $\frac{e}{2!} (1/4)^2 < 0,085$
- błąd interpolacji kwadratowej:  $\frac{e^2}{3!} (3/4)(1/4)(5/4) < 0,29$

Widzimy więc, że oszacowania błędów dla interpolacji kwadratowej są mniej optymistyczne – mimo większej ilości pracy. Wynika to z faktu, że mamy znacznie większy przedział  $[0, 2]$ .

**Zad. 4.5** Oszacuj błąd obliczania wartości  $\ln 100,5$  z użyciem interpolacji Lagrange’a, gdy dane są  $\ln 100, \ln 101, \ln 102, \dots, \ln(100 + n)$ . Dla jakiej wartości  $n$  błąd jest najmniejszy?

**Rozwiązanie.** Musimy obliczyć pochodne funkcji  $\ln x$ :

$$\begin{aligned}(\ln x)' &= x^{-1} \\(\ln x)'' &= -x^{-2} \\(\ln x)''' &= 2x^{-3} \\(\ln x)^{IV} &= -2 \cdot 3x^{-4} \\&\dots \\(\ln x)^{(n)} &= (-1)^{n+1} (n-1)! x^{-n}\end{aligned}$$

Wszystkie pochodne są więc malejące (co do wartości bezwzględnej). Dla wielomianu  $L_n$  opartego o węzły  $100, 101, \dots, 100 + n$  mamy

$$\begin{aligned}|L_n(100,5) - \ln 100,5| &\leq \frac{M_{n+1}}{(n+1)!} \frac{1}{2} \frac{1}{2} \frac{3}{2} \cdot \dots \cdot \frac{2n-1}{2} \\&= \frac{n! 100^{-n-1}}{(n+1)! 2^{n+1}} \frac{1 \cdot 2 \cdot 3 \cdot \dots \cdot (2n)}{2 \cdot 4 \cdot \dots \cdot (2n)} \\&= \frac{1}{10^{2n+2} 2^{2n+1} (n+1)} \frac{(2n)!}{2^n n!} = \frac{1}{10^{2n+2} 2^{2n+1}} \frac{(2n)!}{(n+1)!} \xrightarrow{n \rightarrow \infty} \infty.\end{aligned}$$

Zachowane tej funkcji moglibyśmy badać wykorzystując np. wzór Stirlinga:  $n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$ , jednak łatwiej będzie na skorzystać z Julii lub innego podobnego narzędzia. Można przekonać się, że funkcja ta osiąga minimum dla  $n = 101$ . Zwiększanie liczby węzłów ponad 101 powiększa błąd (a przynajmniej nasze oszacowanie błędu).

**Zadanie domowe.** Napisz program obliczający współczynniki wielomianu interpolacyjnego w postaci Newtona i wartość tego wielomianu w zadanym punkcie schematem Hornera.

## Zadanie interpolacyjne Hermite’a

Dany jest ciąg  $(n+1)$  punktów (niekoniecznie różnych):

$$x_0 \leq x_1 \leq x_2 \leq \dots \leq x_n$$

i odpowiadający im ciąg  $(n+1)$  wartości

$$y_0, y_1, \dots, y_n.$$

Szukamy wielomianu  $H(x)$  takiego, że gdy ta sama liczba  $x$  powtarza się  $m$  razy w ciągu  $(x_0, \dots, x_n)$  jako  $(x_k, x_{k+1}, \dots, x_{k+m-1})$ , to:

$$\begin{aligned}H(x) &= y_l = f(x_k) \\H'(x) &= y_{k+1} = f'(x_{k+1}) \\H''(x) &= y_{k+2} = f''(x_{k+2}) \\&\dots \\H^{m-1}(x) &= y_{k+m-1} = f^{(m-1)}(x_{k+m-1})\end{aligned}$$

Prawdziwy jest analog twierdzenia o postaci Newtona dla Wielomianu Lagrange’a, z dodatkowym zastrzeżeniem dla ilorazów różnicowych, mianowicie gdy natrafilibyśmy na konieczność dzielenia przez zero według starej definicji, to kładziemy:

$$f[\underbrace{x, x, \dots, x}_{i \text{ razy}}] = \frac{f^{(i-1)}(x)}{(i-1)!}.$$

Proszę zwrócić uwagę, że odcięte węzłów  $x_i$  są **uporządkowane rosnąco!**

Analogicznie jak dla interpolacji Lagrange’a mamy:

**Twierdzenie 4.6.** Niech  $f$  będzie funkcją klasy  $C^K$  na przedziale  $[a, b]$  i niech  $H$  będzie wielomianem interpolacyjnym Hermite'a dla  $f$  spełniającym razem  $K$  warunków na wartości i wartości pochodnych, przy czym wszystkie węzły leżą w przedziale  $[a, b]$ . Wtedy dla  $x \in [a, b]$ :

$$|f(x) - H(x)| \leq \frac{M_K}{K!} \prod_i |x - x_i|,$$

gdzie  $M_K = \max_{x \in [a, b]} |f^{(K)}(x)|$ , a każdy punkt  $x_i$  występuje tyle razy w ilu danych występuje.

**Zad. 4.6** Znajdź wielomian interpolacyjny Hermite'a dla

- a)  $f(1) = 0, f(2) = 1, f'(1) = 1, f'(2) = 0$ ;
- b)  $f(0) = 0, f(1) = 3, f'(0) = 1, f'(1) = 6$ ;
- c)  $f(1) = 0, f(2) = 6, f'(1) = 3, f'(2) = 7, f''(2) = 8$ ;
- d)  $f(1) = 0, f(2) = 0, f'(1) = 3, f'(2) = 2, f''(1) = 2$ .

**Zad. 4.7** Znajdź przybliżenie  $\sqrt{e}$  korzystając z wielomianu interpolacyjnego Hermite'a stopnia 4 dla funkcji  $f(x) = e^x$  z wielokrotnym węzłem w punkcie 0. Oszacuj błąd znalezionego przybliżenia.

**Rozwiązanie.** Zauważmy, że  $\sqrt{e} = e^{1/2}$ . Szukamy wielomianu  $H$ , który zgadza się z  $e^x$  w punkcie zero (wartość i kolejne pochodne). Ponieważ chcemy mieć wielomian stopnia 4, potrzebujemy 5 punktów danych:

$$\begin{aligned} H(0) &= e^0 \\ H'(0) &= (e^x)'|_{x=0} = 1 \\ H''(0) &= (e^x)''|_{x=0} = 1 \\ H'''(0) &= (e^x)'''|_{x=0} = 1 \\ H''''(0) &= (e^x)''''|_{x=0} = 1. \end{aligned}$$

Tworzymy tabelkę, powtarzając 5 razy węzeł 0:

0	1			
		1		
0	1		1/2	
		1		1/6
0	1		1/2	1/24
		1		1/6
0	1		1/2	
		1		
0	1			

Odczytujemy wielomian Hermite'a:

$$H(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4.$$

Zatem

$$\sqrt{e} = e^{1/2} \cong 1 + \frac{1}{2} + \frac{1}{8} + \frac{1}{48} + \frac{1}{384} = \frac{633}{384} = \frac{211}{128}.$$

Oszacujemy teraz błąd, korzystając z podanego wyżej twierdzenia. Zauważmy, że interesujący nas przedział to  $[0, \frac{1}{2}]$ :

$$|\sqrt{e} - H(1/2)| \leq \frac{M_5}{5!} \frac{1}{2^5} = \frac{e^{1/2}}{32 \cdot 5!} < \frac{1}{16 \cdot 120} = \frac{1}{1920}.$$

**Zadanie domowe.** Napisz program obliczający współczynniki wielomianu interpolacyjnego Hermite'a w postaci Newtona i wartość tego wielomianu w zadanym punkcie algorytmem przypominającym schemat Hornera.

```

import Polynomials
import Plots

Plots.plotly()

runge(x) = sin(x)

function rungeApproximatingPolynomial(a, b, n)
    x = range(a, stop=b, length=n+1)
    y = runge.(x)
    return Polynomials.polyfit(x, y)
end

d = -10:0.01:10

Plots.plot(d, [runge.(d),
    rungeApproximatingPolynomial(-10,10,2).(d),
    rungeApproximatingPolynomial(-10,10,4).(d),
    rungeApproximatingPolynomial(-10,10,6).(d),
    rungeApproximatingPolynomial(-10,10,8).(d),
    rungeApproximatingPolynomial(-10,10,10).(d),
    rungeApproximatingPolynomial(-10,10,12).(d)])

```

Kod źródłowy 2: Zjawisko Rungego

## Zjawisko Rungego

**Zad. 4.8** Uruchom Kod źródłowy 2 w Julii. Zaobserwuj jak zachowują się przybliżenia wielomianami coraz wyższych stopni. Sprawdź eksperymentalnie czy podobne zjawisko będzie zachodzić jeśli zmniejszymy przedział interpolacji. Spróbuj wykonać anlogiczny eksperyment dla interpolacji funkcji sinus.

# 5. Funkcje sklejące

Funkcją sklejaną (ang. *spline function*) stopnia  $k$  nazywamy funkcję  $S$  spełniającą następujące warunki:

- istnieją przedziały  $[x_i, x_{i+1}]$  takie, że  $S$  obcięta do każdego takiego przedziału jest wielomianem stopnia co najwyżej  $k$ .
- funkcja  $S$  ma  $k - 1$  ciągłych pochodnych.

Dla przykładu funkcja:

$$f(x) = \begin{cases} x^2, & x \leq 0; \\ -x^2, & x > 0 \end{cases}$$

jest funkcją sklejaną stopnia 2.

W praktyce najważniejsze są funkcje sklejące stopnia 3. Są one (lub ich bliscy kuzyni) podstawą najczęściej używanych sposobów modelowania krzywych oraz powierzchni w programach graficznych i inżynierskich oraz w fontach komputerowych. Przyjrzyjmy się bliżej tym funkcjom.

Założmy, że dane są liczby  $x_0 < x_1 < \dots < x_n$  oraz odpowiadające im wartości  $y_0, y_1, \dots, y_n$ . Chcemy znaleźć funkcję sklejaną  $S$  stopnia 3 taką, że  $S(x_i) = y_i$  oraz na każdym przedziale  $[x_i, x_{i+1}]$  funkcja  $S$  jest wielomianem stopnia 3. Można podać dokładny algorytm, ale nie będziemy tego robić zadowalając się następującym intuicyjnym “rozwiązaniem” problemu.

Funkcję  $S$  obciętą do przedziału  $[x_i, x_{i+1}]$  oznaczmy przez  $S_i$ , jest to wielomian stopnia 3. Załóżmy, że aby wyznaczyć  $S_i$  musimy znaleźć cztery współczynniki odpowiedniego wielomianu.

W sumie mamy więc  $4(n-1)$  zmiennych. Warunki na ciągłość  $S$  i jej pochodnych dają nam następujące równania:

$$\begin{aligned} S_0(x_1) &= S_1(x_1), S_1(x_2) = S_2(x_2), \dots, S_{n-2}(x_{n-1}) = S_{n-1}(x_{n-1}) \\ S'_0(x_1) &= S'_1(x_1), S'_1(x_2) = S'_2(x_2), \dots, S'_{n-2}(x_{n-1}) = S'_{n-1}(x_{n-1}) \\ S''_0(x_1) &= S''_1(x_1), S''_1(x_2) = S''_2(x_2), \dots, S''_{n-2}(x_{n-1}) = S''_{n-1}(x_{n-1}) \end{aligned}$$

Ponadto chcemy mieć:

$$S_0(x_0) = y_0, S_1(x_1) = y_1, \dots, S_{n-1}(x_{n-1}) = y_{n-1}, S_{n-1}(x_n) = y_n$$

W sumie mamy  $3(n-2) + n = 4n - 6$  równań. Chwila zastanowienia pokazuje że są to równania linowe, można pokazać, że są one niezależne. Ponieważ zmiennych jest  $4n - 4$ , to powyższe równania nie wyznaczają żądanych współczynników jednoznacznie.

W praktyce najczęściej przyjmuje się dodatkowe dwa równania, aby ujednostacznić wybór współczynników. Żądając dodatkowo, by  $S''(x_0) = S''(x_n) = 0$  dostajemy tzw. naturalną funkcję sklejaną stopnia 3. Innym częstym wyborem jest dodanie równań  $S'(x_0) = S'(x_n), S''(x_0) = S''(x_n)$ . Daje to funkcję sklejaną, która dobrze nadają się do przybliżania funkcji okresowej.

**Zad. 5.1** Niech  $S$  będzie funkcją sklejaną stopnia 1 interpolującą funkcję  $f$  w punktach  $0 = x_0 < x_1 < \dots < x_n = 1$ . Ile wynosi  $\int_0^1 S(x) dx$ .

**Zad. 5.2** Rozstrzygnij, czy poniższa funkcja jest funkcją sklejaną stopnia 2. Czy jest funkcją sklejaną stopnia 3?

$$f(x) = \begin{cases} x, & x \in (-\infty, 1] \\ -\frac{1}{2}(2-x)^2 + \frac{3}{2}, & x \in [1, 2] \\ \frac{3}{2}, & x \in [2, \infty). \end{cases}$$

**Zad. 5.3** Znajdź wartości  $a, b, c, d, e$  dla których poniższa funkcja jest funkcją sklejaną stopnia 3:

$$f(x) = \begin{cases} a(x-2)^2 + b(x-1)^3, & x \in (-\infty, 1] \\ c(x-2)^2, & x \in [1, 3] \\ d(x-2)^2 + e(x-3)^3, & x \in [3, \infty). \end{cases}$$

Jakie muszą być te wartości, aby  $f(0) = 26, f(1) = 7, f(4) = 25$ ?

**Zad. 5.4** Czy istnieją wartości  $a, b, c, d$  takie, że

$$f(x) = \begin{cases} 1 - 2x, & x \in (-\infty, -3] \\ a + bx + cx^2 + dx^3, & x \in [-3, 4] \\ 157 - 32x, & x \in [4, \infty) \end{cases}$$

jest naturalną funkcją sklejaną na przedziale  $[-3, 4]$ ?

**Zad. 5.5** Znajdź naturalną funkcję sklejaną stopnia 3 o węzłach  $-1, 0, 1$  przyjmującą w węzłach wartości odpowiednio  $13, 7, 9$ .

**Zad. 5.6** Używając funkcji z pliku [naturalCubic.jl](#) znajdź przybliżenie spirali archimedesowej danej ww współrzędnych biegunowych równaniem parametrycznym:

$$r(\varphi) = \frac{\varphi}{2}, \quad \varphi \in [0, 4\pi].$$

W tym celu przybliż funkcje  $x(\varphi)$  oraz  $y(\varphi)$  funkcjami sklejanymi.

Następnie zrób to samo dla okręgu.

**Zad. 5.7** Dokonaj interpolacji wartości funkcji  $f(x) = \frac{1}{1+5x^2}$  na przedziale  $[-1, 1]$  na trzy sposoby:

- znajdując wielomian interpolacyjny Lagrange'a dla  $f$  w 21 równomiernie rozłożonych węzłach;
- znajdując wielomian interpolacyjny Lagrange'a dla  $f$  w 21 węzłach Czebyszewa:  $\cos \frac{i\pi}{2}$  dla  $i = 0, 1, \dots, 20$ ;
- znajdując naturalną funkcję sklejaną dla interpolującą  $f$  w 21 równomiernie rozłożonych węzłach.

Porównaj dokładność przybliżenia w 41 równomiernie rozłożonych punktach.

**Zadanie domowe.** Napisz w dowolnym języku program, który dla zadanych ciągów  $x_0, x_1, \dots, x_n$  oraz  $y_0, y_1, \dots, y_n$  znajdzie funkcję sklejaną  $S$  stopnia 3 taką, że  $S(x_i) = y_i$ . Jak wiadomo funkcja taka nie jest wyznaczona jednoznacznie. Swój program napisz używając jednego z warunków brzegowych:

- okresowość:  $S'(x_0) = S'(x_n), S''(x_0) = S''(x_n)$ ;
- “not-a-knot”:  $S'''(x_1 - 0) = S'''(x_1 + 0), S'''(x_{n-1} - 0) = S'''(x_{n-1} + 0)$ ;
- wartości pochodnej na brzegu:  $S'(x_0) = a, S'(x_n) = b$  (gdzie  $a$  i  $b$  są parametrami procedury).

W programie powinny być przynajmniej dwie funkcje: jedna obliczająca i zwracająca funkcję  $S$  (dla której trzeba wymyślić jakąś odpowiednią reprezentację) i druga obliczająca wartość funkcji w zadanym punkcie – podobnie jak w pliku [naturalCubic.jl](#). Proszę użyć tych funkcji dla przykładowych danych.

**Zadanie domowe.** Dla przedziału  $[a, b]$ , gdy mamy dane  $S(a), S(b), S'(a), S'(b)$  możemy wyznaczyć jednoznacznie wielomian  $S$  stopnia 3 spełniający te warunki (szczególny przypadek interpolacji Hermite’a). Napisz program, który dla zadanych ciągów  $x_0, x_1, \dots, x_n$  oraz  $y_0, y_1, \dots, y_n$  znajdzie funkcję interpolującą  $S$ , która na każdym przedziale  $[x_i, x_{i+1}]$  jest wielomianem stopnia 3, oraz spełnia warunki:

- $S(x_i) = y_i$
- $S'(x_i) = \frac{1}{2} \left( \frac{y_{i+1} - y_i}{x_{i+1} - x_i} + \frac{y_i - y_{i-1}}{x_i - x_{i-1}} \right)$  dla  $i = 1, 2, \dots, n-1$
- $S'(x_0) = \frac{y_1 - y_0}{x_1 - x_0}$
- $S'(x_n) = \frac{y_n - y_{n-1}}{x_n - x_{n-1}}$

*Uwaga: co prawda obliczona funkcja jest nazywana funkcją sklejaną Hermite’a, prawdopodobnie nie będzie jednak funkcją sklejaną stopnia 3 w sensie przyjętej przez nas definicji – jej druga pochodna nie musi być ciągła.*

W programie powinny być przynajmniej dwie funkcje: jedna obliczająca i zwracająca funkcję  $S$  (dla której trzeba wymyślić jakąś odpowiednią reprezentację) i druga obliczająca wartość funkcji w zadanym punkcie – podobnie jak w pliku [naturalCubic.jl](#). Proszę użyć tych funkcji dla przykładowych danych.

## 6. Kwadratury interpolacyjne

### Kwadratury proste

Będziemy szukać sposobu obliczenia przybliżenia wartości całki oznaczonej:

$$\int_a^b f(x) dx. \quad (6.1)$$

Wiadomo, że jeżeli znajdziemy funkcję  $F$  taką, że  $F' = f$ , to

$$\int_a^b f(x) dx = F(b) - F(a),$$

jednak w praktyce znalezienie postaci funkcji  $F$  może nie być proste, lub wręcz niemożliwe (tak jest np. dla bardzo istotnych funkcji postaci  $x \mapsto e^{ax^2}$ ). Dlatego skazani jesteśmy na obliczanie całek oznaczonych używając przybliżeń.

Pomysł na obliczenie całki jest prosty: zastąpimy  $f$  przez wielomian interpolacyjny na przedziale  $[a, b]$  i całkę (6.1) obliczymy jako całkę z tego wielomianu interpolacyjnego. Będziemy używać tylko interpolacji Lagrange’a (w praktyce wartości funkcji  $f$  są często znajdowane eksperymentalnie, obliczanie pochodnych byłoby bardzo trudne).

Przypomnijmy, że wielomian interpolacyjny Lagrange’a dla węzłów  $x_0, x_1, \dots, x_n \in [a, b]$  dany jest w bazie Lagrange’a wzorem:

$$L(x) = \sum_{j=0}^n f(x_j) l_j(x),$$

gdzie  $l_j$  są wielomianami przyjmującymi w  $x_j$  wartość 1, a w pozostałych węzłach wartość 0. Podstawiając to do całki (6.1) otrzymujemy:

$$\int_a^b f(x) dx \approx \int_a^b L(x) dx = \sum_{j=0}^n f(x_j) \int_a^b l_j(x) dx = \sum_{j=0}^n f(x_j) A_j.$$

Wyrażenia postaci  $\sum_{j=0}^n f(x_j) A_j$  nazywamy **kwadraturami**. Jeżeli węzły  $x_j$  są równoodległe, to mówimy o kwadraturach Newtona-Cotesa. Jedyne co musimy zrobić to znaleźć odpowiednie wartości  $A_j$ .

Wartości te można znaleźć na dwa sposoby.

Jeden z nich to obliczenie całek  $\int_a^b l_j(x) dx$ . Wydaje się to skomplikowane, ale w praktyce nie jest takie trudne. Skorzystamy z zamiany zmiennych w całce oznaczonej:

$$\begin{aligned} \int_a^b f(x) dx &= \int_0^n f\left(\frac{(b-a)t}{n} + a\right) \frac{b-a}{n} dt \\ &\approx \frac{b-a}{n} \sum_{j=0}^n f\left(\frac{(b-a)k}{n} + a\right) \int_0^n l_j(x) dx, \end{aligned}$$

gdzie tym razem wielomiany  $l_j$  są zdefiniowane na przedziale  $[0, n]$  i każdy z nich przyjmuje w punktach całkowitych wartość 0, poza punktem  $j$ , gdzie przyjmuje wartość 1.

Drugi sposób znajdowania wartości  $A_j$  wymaga pojęcia rzędu kwadratury.

**Definicja 6.1.** Mówimy, że kwadratura  $\sum_{k=0}^n f(x_k) A_k$  jest dokładna dla funkcji  $f$ , jeżeli

$$\sum_{k=0}^n f(x_k) A_k = \int_a^b f(x) dx.$$

Mówimy, że kwadratura ma rząd  $k$ , gdy jest dokładna dla wszystkich wielomianów stopnia mniejszego od  $k$  i nie jest dokładna dla pewnego wielomianu stopnia  $k$ .

Ponieważ z definicji kwadratura Newtona-Cotesa oparta na węzłach  $[x_0, \dots, x_n]$  jest rzędu co najmniej  $n + 1$ , rozważenie wartości kwadratury dla wielomianów  $1, x, \dots, x^n$  prowadzi do  $n + 1$  równań, które pozwalają obliczyć potrzebne wartości  $A_j$ .

*Uwaga 6.2.* Zamiast prowadzić obliczenia na przedziale  $[a, b]$ , wygodniej jest przekształcić całkę przez zamianę zmiennych do przedziału  $[0, n]$  lub  $[0, 1]$ .

*Uwaga 6.3.* Analogiczne kwadratury można wyprowadzać dla obliczania całek postaci  $\int_a^b f(x) w(x) dx$ , gdzie  $w(x)$  jest funkcją wagową.

*Uwaga 6.4.* Kwadratura zależy od wyboru punktów  $x_j$  oraz od wyboru liczb  $A_j$ . Powyżej punkty  $x_j$  były ustalone, co nieco ograniczało pole manewru. Pozwalając na zmianę punktów, mamy do dyspozycji  $2(n + 1)$  parametrów, i okazuje się, że można je wybrać tak, by dla tej samej ilości węzłów otrzymać wyższy rząd niż dają kwadratury Newtona-Cotesa. Prowadzi to do tzw. kwadratur Gaussa. Zainteresowanych odsyłam do literatury.

## Kwadratury złożone

Dla kwadratur Newtona-Cotesa zwiększanie liczby węzłów prowadzi paradoksalnie do zmniejszania dokładności (dla funkcji, które nie są wielomianami). Aby mimo to móc obliczać te całki w miarę dokładnie, postępujemy inaczej.

W najprostszym przypadku zadany przedział  $[a, b]$  dzielimy na wiele mniejszych przedziałów i na każdym z nich osobno stosujemy tę samą kwadraturę Newtona-Cotesa niezbyt wysokiego stopnia. Jest to tzw. kwadratura złożona i przypomina sumy Riemanna pojawiające się przy definicji całki Riemanna. Bez względu na to jaką kwadraturę weźmiemy, przy wzroście liczby przedziałów wynik dąży do prawdziwej wartości całki.

W praktyce dzisiaj stosuje się tzw. kwadratury adaptacyjne, które używają krótszych przedziałów tam, gdzie to konieczne i same dostosowują się do całkowanej funkcji. Pozwalają one uzyskać dokładniejszy wynik przy mniejszym nakładzie pracy (a więc w krótszym czasie).



## Zadania

- Zad. 6.1** Wyprowadź tzw. wzór trapezów, czyli kwadraturę Newtona-Cotesa opartą o dwa węzły. Sprawdź, że wzór trapezów jest kwadraturą rzędu 2.
- Zad. 6.2** Wyprowadź tzw. wzór Simpsona, czyli kwadraturę Newtona-Cotesa opartą o trzy węzły. Sprawdź, że wzór Simpsona jest kwadraturą rzędu 4(!). Dlaczego może to być zaskakujące?
- Zad. 6.3** Wyprowadź tzw. wzór 3/8-Simpsona, czyli kwadraturę Newtona-Cotesa opartą o cztery węzły. Jaki jest rząd tej kwadratury?
- Zad. 6.4** Wyprowadź wzór na kwadraturę otwartą Newtona-Cotesa opartą o 3 węzły. Jaki jest rząd tej kwadratury?

- Zad. 6.5** Znajdź kwadraturę interpolacyjną obliczającą całkę  $\int_{-1}^1 f(x) dx$  opartą o punkty  $0$  i  $\pm\sqrt{\frac{3}{5}}$ . Jaki jest rząd tej kwadratury?

- Zad. 6.6** Czy istnieje kwadratura

$$\int_0^1 f(x) dx \approx \alpha (f(x_1) + f(x_2)),$$

która jest dokładna dla wielomianów stopnia 2?

- Zad. 6.7** Podpierając się narzędziami do symbolicznego obliczania całek znajdź kwadraturę interpolacyjną do obliczania całki

$$\int_{-\infty}^{\infty} f(x)e^{-x^2} dx$$

opartą o węzły  $0$  i  $\pm\sqrt{\frac{3}{2}}$ . Zbadaj rząd znalezionej kwadratury.

- Zad. 6.8** Znajdź w literaturze oszacowania na błędy wzorów trapezów i Simpsona.

- Zad. 6.9** Sprawdź, że wzór:

$$\int_0^1 f(x) dx \approx \frac{1}{90} (7f(0) + 32f(1/4) + 12f(1/2) + 32f(3/4) + 7f(1))$$

jest dokładny dla wielomianów stopnia  $\leq 4$ .

- Zad. 6.10** Znajdź kwadraturę

$$\int_0^1 f(x) dx \approx A_0 f(0) + A_1 f(1),$$

która jest dokładna dla wszystkich funkcji postaci  $f(x) = ae^x + b \cos \frac{\pi x}{2}$ .

- Zad. 6.11** Znajdź wzór:

$$\int_0^{2\pi} f(x) dx \approx A_0 f(0) + A_1 f(\pi),$$

który jest dokładny dla wszystkich funkcji postaci  $f(x) = a + b \cos x$ .

Udowodnij, że wzór ten jest dokładny dla wszystkich funkcji postaci:

$$f(x) = \sum_{k=0}^n (a_k \cos(2k+1)x + b_k \sin kx).$$

- Zad. 6.12** Znajdź rozwinięcie funkcji  $e^{x^2}$  w szereg potęgowy. Sprawdź, że szereg ten jest jednostajnie zbieżny na każdym przedziale skończonym. Całkując wyraz po wyrazie otrzymany szereg, oblicz przybliżenie całki  $\int_0^2 e^{x^2} dx$ .

- Zad. 6.13** Znajdź kwadraturę interpolacyjną przybliżającą całkę  $\int_1^3 f(x) dx$  opartą o punkty  $0, 2$  i  $4$ .

- Zad. 6.14** Znajdź kwadraturę interpolacyjną opartą o punkty  $x_0 < x_1 < x_2 < x_3$  do obliczenia całki

$$\int_{x_0}^{x_3} f(x) dx.$$

- Zad. 6.15** Uruchom ponownie Kod źródłowy 2 w Julii. Dlaczego nie używamy w praktyce kwadratur Newtona-Cotesa wysokich stopni?

- Zad. 6.16** Za pomocą wzorów trapezów i Simpsona oblicz przybliżenia całek:

a)  $\int_0^1 \sin x \, dx;$

b)  $\int_0^1 \frac{1}{1+x^2} \, dx;$

c)  $\int_0^1 \frac{1}{1+25x^2} \, dx;$

d)  $\int_0^{\frac{1}{10}} x^{\frac{1}{3}} \, dx;$

e)  $\int_0^{\pi} e^{\sin(x)} \, dx.$

**Zad. 6.17** Poniżej znajduje się przykładowa implementacja złożonej kwadratury trapezów. Na jej podstawie zaimplementuj złożoną kwadraturę Simpsona. Za ich pomocą w Julii oblicz przybliżone wartości całek:

a)  $\int_{-1}^1 \sqrt{1-x^4} \, dx;$

b)  $\int_0^1 e^{e^x} \, dx;$

c)  $\int_2^7 \frac{1}{\ln x} \, dx;$

d)  $\int_e^{e^2} \ln \ln x \, dx;$

trapezoid.jl

```
function trapezoid(f::Function, a::Number, b::Number,
                  n::Integer)
    h = (b - a)/n;
    result = 0
    for i in 0:(n-1)
        result += f(a+i*h) + f(a+(i+1)*h);
    end
    result *= h/2;
end
```

Kod źródłowy 3: Implementacja złożonego wzoru trapezów

## 7. Metody iteracyjne

Metody iteracyjne to metody poszukiwania rozwiązań równania  $f(x) = 0$ . Na wyjściu metody iteracyjnej otrzymamy ciąg zbieżny do nieznanego rozwiązania (miejmy nadzieję).

### Metoda bisekcji

Założmy, że funkcja  $f$  jest ciągła i znamy dwie liczby  $x_1, x_2$  takie, że  $f(x_1) \cdot f(x_2) < 0$ . Wtedy z własności Darboux wynika, że między  $x_1$  i  $x_2$  jest pierwiastek rozważanego równania. Aby go odnaleźć postępujemy następująco: Niech  $x$  będzie połową przedziału  $[x_1, x_2]$ . Zastępujemy  $x_1$  bądź  $x_2$  przez  $x$  tak, aby cały czas zachodził warunek  $f(x_1) \cdot f(x_2) < 0$ . Procedurę kontynuujemy tak długo aż będziemy usatysfakcjonowani uzyskaną dokładnością.

### Metoda siecznych

Założmy, że dana jest funkcja  $f$  oraz dwie liczby  $x_1, x_2$ . Prowadzimy prostą przechodzącą przez punkty  $(x_1, f(x_1))$  i  $(x_2, f(x_2))$ . Niech  $x$  będzie miejscem zerowym tej prostej. Zastępujemy  $x_1$  przez  $x_2$  i  $x_2$  przez obliczoną wartość  $x$  i powtarzamy procedurę.

### Metoda stycznych (Newtona)

Założmy, że dana jest funkcja różniczkowalna  $f$  oraz liczba  $x_1$ . Prowadzimy styczną do wykresu  $f$  przechodzącą przez punkt  $(x_1, f(x_1))$ . Niech  $x$  będzie miejscem zerowym tej stycznej. Zastępujemy  $x_1$  przez obliczoną wartość  $x$  i powtarzamy procedurę.

## Rząd zbieżności ciągu

**Definicja 7.1.** Niech  $x_n \rightarrow x$ . Rzędem zbieżności ciągu  $(x_n)$  nazywamy supremum zbioru liczb  $\alpha$  dla których istnieje stała  $C > 0$  taka, że

$$|x_{n+1} - x| \leq C|x_n - x|^\alpha.$$

Żądamy przy tym ponadto, aby dla  $\alpha \leq 1$  stała  $C$  była mniejsza od 1.

*Uwaga 7.2.* Jeżeli ciąg ma rząd zbieżności równy 1, to mówimy, że jest zbieżny liniowo, gdy rząd zbieżności jest równy 2, to mówimy o zbieżności kwadratowej. Gdy rząd zbieżności jest między 1 i 2 mówimy o zbieżności superliniowej.

*Uwaga 7.3.* Metoda bisekcji daje ciąg przybliżeń zbieżny liniowo do rozwiązania.

Jeżeli funkcja  $f$  jest klasy  $C^2$  i pierwiastek równania  $f(x) = 0$  jest pierwiastkiem pojedynczym, to istnieje otoczenie pierwiastka w którym metoda Newtona daje ciąg zbieżny do pierwiastka kwadratowo.

Przy podobnych założeniach metoda siecznych daje ciąg zbieżny superliniowo (rząd zbieżności równy  $\frac{\sqrt{5}+1}{2}$ ).

## Twierdzenie o kontrakcji

**Definicja 7.4.** Jeżeli dla funkcji  $F$  istnieje liczba  $\lambda \in [0, 1)$  taka, że

$$|F(x) - F(y)| \leq \lambda|x - y|,$$

to  $F$  nazywamy kontrakcją (odwzorowaniem zwężającym).

*Uwaga 7.5.* Pojęcie kontrakcji ma sens dla dowolnych przestrzeni metrycznych, my ograniczymy się do rozważania ich jedynie na prostej rzeczywistej.

**Twierdzenie 7.6** (Twierdzenie Banacha o kontrakcji). *Niech  $F: X \rightarrow X$  będzie kontrakcją na zupełnej przestrzeni metrycznej  $X$  (np. na odcinku domkniętym, na prostej). Wtedy istnieje dokładnie jeden punkt stały dla  $F$ , to znaczy punkt  $s$  taki, że  $F(s) = s$ . Punkt  $s$  jest granicą każdego ciągu spełniającego dla  $n \geq 0$ :*

$$x_{n+1} = F(x_n).$$

## Zadania

**Zad. 7.1** Znajdź rząd zbieżności ciągu

a)  $q^n, |q| < 1$ ;

b)  $n^k q^n, |q| < 1, k \in \mathbb{Z}$ ;

c)  $q^{n^k}, |q| < 1$ ;

d)  $q^{k^n}, |q| < 1$ ;

**Zad. 7.2** Wzorując się na podanych przykładach implementacji metody bisekcji i Newtona napisać implementację metody siecznych w funkcji `secant`:

```
function secant(f::Function, x0::Number, x1::Number;  
    delta::AbstractFloat=eps(x0),  
    epsilon::AbstractFloat=eps(x0),  
    maxiter::Integer=50)
```

**Zad. 7.3** Znajdź dodatni pierwiastek równania

$$x^2 - 4x \sin x + 4 \sin^2 x = 0$$

metodami bisekcji, siecznych i stycznych. Porównaj liczbę kroków.

**Zad. 7.4** Wykonaj ręcznie 3 kroki metody bisekcji dla równania  $x^2 - 5 = 0$ .

**Zad. 7.5** Znajdź przybliżony pierwiastek równania  $x^3 - 3x^2 - 2x + 5 = 0$  metodą stycznych z dokładnością 0.1, wiedząc, że jest on w przedziale  $[1, 2]$ .

```

using Printf

function bisect(f::Function, a::Number, b::Number;
               delta::Number=eps(1.0),
               epsilon::Number=eps(1.0),
               maxiter::Integer=100)

    k = 0;
    fa = f(a);
    fb = f(b);
    @printf("a      = %30.26f, f(a    ) = %30.26f\n", a, fa)
    @printf("b      = %30.26f, f(b    ) = %30.26f\n", b, fb)
    if sign(fa) == sign(fb)
        error("Między zadanymi wartościami " *
              "niekoniecznie znajdziemy pierwiastek")
    end
    e = b - a;
    for i in 1:maxiter
        e = e/2;
        c = a + e;
        fc = f(c);
        @printf("x%-3d = %30.26f, f(x%-3d) = %30.26f\n", i, c, i, fc)
        if abs(fc) < epsilon || abs(e) < delta
            break
        end
        if sign(fc) != sign(fa)
            b = c;
            fb = fc;
        else
            a = c;
            fa = fc;
        end
    end
end
end

```

Kod źródłowy 4: Przykładowa implementacja metody bisekcji.

**Zad. 7.6** Znajdź przybliżony pierwiastek równania  $x^3 + x - 1$  metodą siecznych zaczynając od  $x_0 = 0, x_1 = 1$ .

**Zad. 7.7** Znajdź pierwiastek równania w zadanym przedziale:

$$\frac{1}{x} - \operatorname{tg} x = 0, \quad [0, \frac{\pi}{2}];$$

$$\frac{1}{x} - 2^x = 0, \quad [0, 1];$$

$$2^{-x} + e^x + 2 \cos x - 6 = 0, \quad [1, 3];$$

$$\frac{x^3 + 4x^2 + 3x + 5}{2x^3 - 9x^2 + 18x - 2} = 0, \quad [0, 4].$$

**Zad. 7.8** Spróbuj znaleźć pierwiastek równania

$$x - \operatorname{tg} x$$

w przedziale  $[1, 2]$  metodą bisekcji. Wytlumacz otrzymane wyniki.

**Zad. 7.9** Niech

$$\begin{cases} x_0 = 0, \\ x_{n+1} = x_n - (\operatorname{tg} x_n - 1)(\cos^2 x_n). \end{cases}$$

Ile wynosi  $\lim x_n$ ? Pokaż, że jest to zastosowanie metody iteracji Newtona dla pewnego równania.

```

using Printf

function newton(f::Function, fp::Function, x::Number;
               delta::Number=eps(1.0),
               epsilon::Number=eps(1.0),
               maxiter::Integer=100)
    fx = f(x)
    @printf("x0 = %30.26f, f(x0) = %30.26f\n", x, fx)
    for i in 1:maxiter
        if abs(fx) < epsilon
            return x
        end
        next = x - fx/fp(x)
        fx = f(next)
        @printf("x%-3d = %30.26f, f(x%-3d) = %30.26f\n", i, next, i, fx)
        if abs(next-x) < delta
            return x
        end
        x = next
    end
    error("Max iteration exceeded")
end

```

Kod źródłowy 5: Przykładowa implementacja metody Newtona.

**Zad. 7.10** Zaproponuj sposób obliczania odwrotności zadanej liczby bez używania dzielenia. Użyj metody Newtona dla równania  $1/x - y = 0$ .

**Zad. 7.11** Znaleźć pierwiastek równania:

$$x^8 - 36x^7 + 546x^6 - 4536x^5 + 22449x^4 - 67284x^3 + 118124x^2 - 109584x + 40320 = 0$$

w przedziale  $[5.5, 6.5]$ . Następnie zamienić współczynnik  $-36$  na  $-36.001$  i powtórzyć.

**Zad. 7.12** Korzystając z metody Newtona podać algorytm obliczania  $\sqrt{R}$ , gdzie  $R$  jest liczbą dodatnią.

**Zad. 7.13** Wielomian  $x^3 + 94x^2 - 389x + 294$  ma pierwiastki  $1, 3, -98$ . Korzystając z metody Newtona dla tego równania i startując z punktu  $2$  zobacz co się dzieje.

**Zad. 7.14** Znajdź pierwiastek równania  $x = 1 + \frac{\sin x}{4}$  metodą iteracji prostej.

**Zad. 7.15** Na rachunek oszczędnościowy z miesięcznym okresem kapitalizacji odsetek wpłacano przez 60 miesięcy środki, w kolejnych latach odpowiednio po 200, 250, 320, 375, 350 złotych. Po ostatniej wpłacie na rachunku znajdowało się 20180 złotych. Jaka była średnia wysokość odsetek na rachunku w tym okresie? Użyj odpowiedniej, wybranej przez siebie metody iteracyjnej.

**Zad. 7.16** Napisz (w dowolnym języku programowania) rekurencyjną wersję algorytmu bisekcji.

---

## 8. Lokalizacja zer wielomianów

---

**Twierdzenie 8.1** (Metoda Sturma). Niech  $P$  będzie wielomianem o współczynnikach rzeczywistych. Zdefiniujmy ciąg wielomianów, tzw. ciąg Sturma:

$$\begin{aligned} s_1(x) &= P(x) \\ s_2(x) &= P'(x) \\ s_n(x) &= \text{reszta z dzielenia } s_{n-2} \text{ przez } s_{n-1} \text{ ze znakiem przeciwnym, } n \geq 3 \\ &\vdots \\ s_p(x) &= 0. \end{aligned}$$

Niech  $a$  i  $b$  nie będą miejscami zerowymi wielomianu. Wtedy liczba miejsc zerowych wielomianu  $P$  w przedziale  $(a, b)$  jest równa

$$S(a) - S(b),$$

gdzie  $S(x)$  jest liczbą zmian znaku w ciągu  $s_1(x), s_2(x), \dots, s_p(x)$  (przy czym nie bierzemy pod uwagę pojawiających się być może zer).

**Twierdzenie 8.2** (Metoda Fouriera). Niech  $P$  będzie wielomianem o współczynnikach rzeczywistych stopnia  $n$ . Weźmy ciąg pochodnych  $P$ :

$$\begin{aligned} f_1(x) &= P(x) \\ f_2(x) &= P'(x) \\ f_3(x) &= P''(x) \\ &\vdots \\ f_{n+1}(x) &= 0. \end{aligned}$$

Niech  $a$  i  $b$  nie będą pierwiastkami wielomianu. Wtedy liczba **pierwiastków** wielomianu  $P$  w przedziale  $(a, b)$  jest równa

$$F(a) - F(b),$$

gdzie  $F(x)$  jest liczbą zmian znaku w ciągu  $f_1(x), f_2(x), \dots, f_n(x)$  (przy czym nie bierzemy pod uwagę pojawiających się być może zer), **lub jest od niej mniejsza o liczbę parzystą**.

**Twierdzenie 8.3** (Metoda Budana). Niech  $P(x) = \sum_{k=0}^n a_k x^k$  będzie wielomianem o współczynnikach rzeczywistych. Dla ustalonej liczby  $r$  tworzymy wielomian  $P_r(t)$  podstawiając za  $x$  wartość  $t + r$ . Oznaczmy liczbę zmian znaku (z pominięciem zer) w ciągu współczynników wielomianu  $P_r(t)$  przez  $B(r)$ . Niech  $a$  i  $b$  nie będą pierwiastkami wielomianu. Wtedy liczba **pierwiastków** wielomianu  $P$  w przedziale  $(a, b)$  jest równa

$$B(a) - B(b),$$

**lub jest od niej mniejsza o liczbę parzystą**.

**Wniosek 8.4** (Reguła Kartezjusza). Niech  $P$  będzie wielomianem i niech  $P(0) \neq 0$ . Wtedy liczba pierwiastków dodatnich  $P$  jest równa liczbie zmian znaku w ciągu współczynników  $P$  bądź jest od niej mniejsza o liczbę parzystą.

*Uwaga 8.5.* Aby zbadać liczbę pierwiastków ujemnych, wystarczy rozważyć wielomian  $P(-x)$ .

### Zadania

**Zad. 8.1** Określ lokalizację pierwiastków rzeczywistych następujących wielomianów (korzystając z dogodnej metody):

a)  $x^3 + x^2 - x - 2$

b)  $x^3 - 5x^2 - 5x + 5$

c)  $x^4 - 2x^2 - 6x + 5$

d)  $x^5 + 4x^4 - 3x^3 + 5x^2 + 3x + 2$

e)  $3x^5 - 4x^4 - 3x^3 + 2x - 1$

**Rozwiązanie.**

a) Korzystamy z najdokładniejszej ale i najbardziej pracochłonnej metody Sturma. Mamy:

$$s_1(x) = x^3 + x^2 - x - 2$$

$$s_2(x) = 3x^2 + 2x - 1$$

$$s_3(x) = \frac{8x + 17}{9} \rightarrow 8x + 17$$

$$s_4(x) = -\frac{531}{64}$$

$$s_5(x) = 0$$

Obliczamy znaki w kilku punktach i liczymy liczby zmian znaku:

	$-\infty$	0	1	2	$\infty$
$s_1(x)$	-	-	-	+	+
$s_2(x)$	+	-	+	+	+
$s_3(x)$	-	+	+	+	+
$s_4(x)$	-	-	-	-	-
$S(x)$	2	2	2	1	1

b) Korzystamy z metody Fouriera. Mamy:

$$f_1(x) = x^3 - 5x^2 - 5x + 5$$

$$f_2(x) = 3x^2 - 10x - 5$$

$$f_3(x) = 6x - 10$$

$$f_4(x) = 6$$

Obliczamy znaki w kilku punktach i liczymy liczby zmian znaku:

	$-\infty$	-2	-1	0	1	2	$\infty$
$f_1(x)$	-	-	+	+	-	-	+
$f_2(x)$	+	+	+	-	-	-	+
$f_3(x)$	-	-	-	-	-	+	+
$f_4(x)$	+	+	+	+	+	+	+
$F(x)$	3	3	2	2	1	1	0

c) Korzystamy z metody Budana. Mamy:

$$P_{-1}(t) = P(t-1) = (t-1)^4 - 2(t-1)^2 - 6(t-1) + 5 = t^4 - 4t^3 + 4t^2 - 6t + 10$$

$$P_0(t) = P(t) = t^4 - 2t^2 - 6t + 5$$

$$P_1(t) = P(t+1) = (t+1)^4 - 2(t+1)^2 - 6(t+1) + 5 = t^4 + 4t^3 + 4t^2 - 6t - 2$$

$$P_2(t) = P(t+2) = (t+2)^4 - 2(t+2)^2 - 6(t+2) + 5 = t^4 + 8t^3 + 22t^2 + 18t + 1$$

Liczmy liczby zmian znaku:

	$-\infty$	-1	0	1	2	$\infty$
$B(x)$	4	4	2	1	0	0

d) Skorzystamy z metody Sturma:

$$s_1(x) = x^5 + 4x^4 - 3x^3 + 5x^2 + 3x + 2$$

$$s_2(x) = 5x^4 + 16x^3 - 9x^2 + 10x + 3$$

$$s_3(x) = \frac{94x^3 - 111x^2 - 20x - 38}{25} \rightarrow 94x^3 - 111x^2 - 20x - 38$$

$$s_4(x) = -\frac{158425x^2 + 147400x + 104750}{8836} \rightarrow -158425x^2 - 147400x - 104750$$

$$s_5(x) = -\frac{4115950176x + 3743486268}{40157569} \rightarrow -4115950176x - 3743486268$$

$$s_6(x) = \frac{2452849413000625}{24109393984}$$

Obliczamy znaki w kilku punktach i liczymy liczby zmian znaku:

	$-\infty$	0	$\infty$
$s_1(x)$	—	+	+
$s_2(x)$	+	+	+
$s_3(x)$	—	—	+
$s_4(x)$	—	—	—
$s_5(x)$	+	—	—
$s_6(x)$	+	+	+
$S(x)$	3	2	2

Z reguły Kartezjusza dostalibyśmy parzystą liczbę pierwiastków dodatnich i nieparzystą liczbę ujemnych.

e) Skorzystamy z metody Budana:

$$P_{-1}(t) = 3(t-1)^5 - 4(t-1)^4 - 3(t-1)^3 + 2(t-1) - 1 = 3t^5 - 19t^4 + 43t^3 - 45t^2 + 24t - 7$$

$$P_0(t) = 3t^5 - 4t^4 - 3t^3 + 2t - 1$$

$$P_1(t) = 3(t+1)^5 - 4(t+1)^4 - 3(t+1)^3 + 2(t+1) - 1 = 3t^5 + 11t^4 + 11t^3 - 3t^2 - 8t - 3$$

$$P_2(t) = 3(t+2)^5 - 4(t+2)^4 - 3(t+2)^3 + 2(t+2) - 1 = 3t^5 + 26t^4 + 85t^3 + 126t^2 + 78t + 11$$

Liczymy liczby zmian znaku:

	$-\infty$	-1	0	1	2	$\infty$
$B(x)$	5	5	3	1	0	0

**Zad. 8.2** Korzystając z metody Fouriera określ liczbę pierwiastków wielomianu  $x^4 - 5x^3 + 2x^2 - x + 3$  na przedziale  $(-5, 5)$ .

**Rozwiązanie.** Sprawdzamy, że  $\pm 5$  nie są pierwiastkami danego wielomianu. Mamy:

$$f_1(x) = x^4 - 5x^3 + 2x^2 - x + 3$$

$$f_2(x) = 4x^3 - 15x^2 + 4x - 1$$

$$f_3(x) = 12x^2 - 30x + 4$$

$$f_4(x) = 24x - 30$$

$$f_5(x) = 24$$

Obliczamy znaki liczymy liczby zmian znaku:

	-5	5
$f_1(x)$	+	+
$f_2(x)$	—	+
$f_3(x)$	+	+
$f_4(x)$	—	+
$f_5(x)$	+	+
$F(x)$	4	0



## 9. Rozkład PLU macierzy kwadratowych

Zajmiemy się rozwiązywaniem układów równań liniowych

$$Ax + b,$$

gdzie:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

i  $A$  jest macierzą nieosobliwą.

Zauważmy, że jeżeli  $A$  jest macierzą górnotrójkątną lub macierzą dolnotrójkątną z niezerową główną przekątną, to (odpowiednio):

$$x_i = \frac{b_i - \sum_{j=1}^{i-1} a_{ij}x_j}{a_{ii}}, \quad i = 1, \dots, n$$

$$x_i = \frac{b_i - \sum_{j=i+1}^n a_{ij}x_j}{a_{ii}}, \quad i = n, \dots, 1,$$

zatem układ łatwo już rozwiązać, gdy  $A$  jest takiej postaci. Rozwiązywanie danego układu będziemy starali się sprowadzić do rozwiązywania układów tego typu.

Przypuśćmy, że  $a = LU$ , gdzie  $L$  – dolnotrójkątna i  $U$  – górnotrójkątna. Wtedy mamy:

$$\begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

Stąd

$$a_{ij} = \sum_{s=1}^n l_{is}u_{sj} = \sum_{s=1}^{\min(i,j)} l_{is}u_{sj}.$$

Założmy teraz, że obliczyliśmy już kolumny do  $k-1$  włącznie w macierzy  $L$  i wiersze do  $k-1$  włącznie w macierzy  $U$ . Wtedy  $a_{kk} = \sum_{s=1}^{k-1} l_{ks}u_{sk} + l_{kk}u_{kk}$ , skąd możemy obliczyć  $l_{kk}$  i  $u_{kk}$  (możliwe konwencje poniżej).

Wyznaczwszy  $l_{kk}$  i  $u_{kk}$  mamy dla  $k$ -tego wiersza i  $k$ -tej kolumny w  $A$ :

$$a_{kj} + \sum_{s=1}^{k-1} l_{ks}u_{sj} + l_{kk}u_{kj}, \quad j = k+1, \dots, n$$

$$a_{ik} + \sum_{s=1}^{k-1} l_{is}u_{sk} + l_{ik}u_{kk}, \quad i = k+1, \dots, n,$$

skąd możemy obliczyć potrzebne  $u_{kj}$  i  $l_{ik}$ .

W powyższej procedurze mieliśmy pewną dowolność przy ustalaniu  $u_{kk}$  i  $l_{kk}$ . Odpowiednie konwencje mają swoje nazwy:

- $l_{kk} = 1$  – rozkład Doolittle’a;
- $u_{kk} = 1$  – rozkład Crouta;
- $l_{kk} = u_{kk}$  – rozkład Cholesky’ego (możliwy tylko dla niektórych macierzy).

Niezależnie od konwencji mówimy o rozkładzie LU. Na tych zajęciach będziemy wybierać konwencję pierwszą. Mając jeden z rozkładów łatwo uzyskać pozostałe wyłączając z macierzy czynnik diagonalny.

Nie jest prawdą, że dowolna nieosobliwa macierz ma rozkład LU. Jeżeli  $A = LU$ , to  $a_{11} = l_{11}u_{11}$ , więc gdy  $a_{11} = 0$ , to stąd  $l_{11} = 0$  lub  $u_{11} = 0$ , to jednak oznaczałoby, że  $L$  bądź  $U$  jest osobliwa, co jest niemożliwe. Prawdziwe jest jednak twierdzenie:

**Twierdzenie 9.1.** Jeżeli wszystkie minory główne macierzy  $A$  są niezerowe, to  $A$  ma rozkład LU.

Ponadto (będziemy tego potrzebować później):

**Twierdzenie 9.2.** Jeżeli  $A$  jest symetryczna i dodatnio określona (tzn. dla dowolnego niezerowego wektora  $x$  mamy  $x^T A x > 0$ ), to  $A$  ma jednoznaczny rozkład postaci  $A = LL^T$ , gdzie  $L$  jest macierzą dolnotrójkątną o dodatniej przekątnej.

Wyżej opisana metoda jest w istocie metodą eliminacji Gaussa, bardziej jednak nadaje się dla komputerów niż do obliczeń praktycznych. Praktyczne obliczenia polegają na konstruowaniu ciągu macierzy

$$A = A^{(1)} \rightarrow A^{(2)} \rightarrow \dots \rightarrow A^{(n)} = U.$$

takich, że

$$A^{(k)} = \left[ \begin{array}{ccc|ccc} a_{11}^{(k)} & \dots & a_{1,k-1}^{(k)} & a_{1k}^{(k)} & \dots & a_{1n}^{(k)} \\ \vdots & & \ddots & \vdots & & \vdots \\ 0 & \dots & a_{k-1,k-1}^{(k)} & a_{k-1,k}^{(k)} & \dots & a_{k-1,n}^{(k)} \\ \hline 0 & \dots & 0 & a_{kk}^{(k)} & \dots & a_{kn}^{(k)} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & a_{nk}^{(k)} & \dots & a_{nn}^{(k)} \end{array} \right]$$

i danych rekurencyjną zależnością:

$$a_{ij}^{(k+1)} = \begin{cases} a_{ij}^{(k)}, & i \leq k; \\ a_{ij}^{(k)} - \frac{a_{ik}^{(k)} a_{kj}^{(k)}}{a_{kk}^{(k)}}, & i, j \geq k+1; \\ 0, & i \geq k+1 \text{ i } j \leq k. \end{cases}$$

Równolegle obliczamy elementy  $k$ -tej kolumny macierzy  $L$ :

$$l_{ik} = \begin{cases} 0, & i \leq k-1; \\ 1, & i = k; \\ \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}, & i \geq k+1. \end{cases}$$

W praktyce, ponieważ w macierzy  $A^{(k)}$  mamy duży blok zer, tam właśnie zapisujemy już obliczone elementy macierzy  $L$ , otaczając je kółkami. Elementy niezapisane w ten sposób to zera i jedynki na głównej przekątnej, więc łatwo je odtworzyć.

**Uwaga 9.3.** Przy wykonywaniu powyższego algorytmu może się zdarzyć, że przyszłoby nam podzielić przez zero (jedna z liczb  $a_{kk}^{(k)}$  mogłaby równać się zero). To znak, że dana macierz nie ma rozkładu LU. Można wtedy rozkład LU wykonywać dla macierzy  $PA$ , gdzie  $P$  jest zero-jedynkową macierzą przestawiającą wiersze. Dla każdej nieosobliwej macierzy istnieje macierz permutacji  $P$  taka, że  $PA$  ma rozkład LU.

**Uwaga 9.4.** Przy implementacji powyższego algorytmu, pożyteczne jest aby nie dzielić przez liczby bardzo małe, gdyż to znacząco pogarsza dokładność uzyskiwanych wyników. W tym celu stosuje się różne strategie wyboru tzw. elementu głównego (elementu, który zastąpi  $a_{kk}^{(k)}$ ):

- częściowy wybór elementu głównego – w  $k$ -tym kroku algorytmu wybieramy najpierw spośród liczb  $a_{ik}^{(k)}$  ( $i \geq k$ ) tę o największym module (niech będzie to  $a_{pk}^{(k)}$ ), a następnie zamieniamy  $k$ -ty wiersz z  $p$ -tym, co odpowiada przemnożeniu naszej macierzy przez odpowiednią macierz permutacji z lewej strony.
- całkowity wybór elementu głównego – w  $k$ -tym kroku algorytmu wybieramy najpierw spośród liczb  $a_{ij}^{(k)}$  ( $i, j \geq k$ ) tę o największym module (niech będzie to  $a_{pl}^{(k)}$ ), a następnie zamieniamy  $k$ -ty wiersz z  $p$ -tym i  $k$ -tą kolumnę z  $l$ -tą, co odpowiada przemnożeniu naszej macierzy przez odpowiednią macierz permutacji z lewej strony i inną macierz permutacji z prawej strony.

Dokonując przestawiania wierszy lub kolumn należy pamiętać o przemnożeniu przez odpowiednie macierze także wektorów  $b$  i  $x$ , w przeciwnym przypadku dostaniemy błędne rozwiązania.

**Zad. 9.1** Rozwiąż układ równań:

a)

$$\begin{aligned}4x_1 + 4x_2 - 3x_3 &= 3 \\2x_1 - 3x_2 + x_3 &= -1 \\2x_1 + 3x_2 - x_3 &= 5\end{aligned}$$

b)

$$\begin{aligned}4x_1 - 2x_2 + 4x_3 - 2x_4 &= 8 \\3x_1 + x_2 + 4x_3 + 2x_4 &= 7 \\2x_1 + 4x_2 + 2x_3 + x_4 &= 10 \\2x_1 - 2x_2 + 4x_3 + 2x_4 &= 2\end{aligned}$$

c)

$$\begin{aligned}x_1 + x_3 &= 3 \\2x_1 + 6x_2 + 3x_3 &= 1 \\x_1 + 3x_2 + 3x_3 &= 2\end{aligned}$$

**Rozwiązanie.** a) Zapisujemy układ w postaci macierzowej:

$$\begin{bmatrix} 4 & 4 & -3 \\ 2 & -3 & 1 \\ 2 & 3 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}.$$

Teraz rozłożymy macierz układu na macierz dolnotrójkątną  $L$  i górną  $U$ .

$$\begin{bmatrix} 4 & 4 & -3 \\ 2 & -3 & 1 \\ 2 & 3 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} \textcolor{red}{4} & 4 & -3 \\ \textcircled{\frac{1}{2}} & -5 & \frac{5}{2} \\ \textcircled{\frac{1}{2}} & 1 & \frac{1}{2} \end{bmatrix} \rightarrow \begin{bmatrix} \textcolor{red}{4} & 4 & -3 \\ \textcircled{\frac{1}{2}} & \textcolor{red}{-5} & \frac{5}{2} \\ \textcircled{\frac{1}{2}} & \textcircled{-\frac{1}{5}} & 1 \end{bmatrix} \rightarrow \begin{bmatrix} \textcolor{red}{4} & 4 & -3 \\ \textcircled{\frac{1}{2}} & \textcolor{red}{-5} & \frac{5}{2} \\ \textcircled{\frac{1}{2}} & \textcircled{-\frac{1}{5}} & \textcolor{red}{1} \end{bmatrix}$$

Dostajemy więc rozkład:

$$\begin{bmatrix} 4 & 4 & -3 \\ 2 & -3 & 1 \\ 2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & -\frac{1}{5} & 1 \end{bmatrix} \begin{bmatrix} 4 & 4 & -3 \\ 0 & -5 & \frac{5}{2} \\ 0 & 0 & 1 \end{bmatrix} = LU$$

Aby rozwiązać układ równań musimy najpierw rozwiązać pomocniczy układ  $Ly = b$ , czyli

$$\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & -\frac{1}{5} & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix}$$

Podstawiając po kolei dostajemy, że

$$\begin{aligned}y_1 &= 3 \\y_2 &= -1 - \frac{1}{2} \cdot 3 = -\frac{5}{2} \\y_3 &= 5 - \frac{1}{2} \cdot 3 + \frac{1}{5} \cdot \left(-\frac{5}{2}\right) = 3\end{aligned}$$

Pozostaje nam do rozwiązania układ  $Ux = y$ :

$$\begin{bmatrix} 4 & 4 & -3 \\ 0 & -5 & \frac{5}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 3 \\ -\frac{5}{2} \\ 3 \end{bmatrix}$$

Dostajemy, że:

$$\begin{aligned}x_3 &= 3 \\x_2 &= \frac{-\frac{5}{2} - 3 \cdot \frac{5}{2}}{-5} = 2 \\x_1 &= \frac{3 - 4 \cdot 2 + 3 \cdot 3}{4} = 1\end{aligned}$$

b) Skorzystamy z częściowego wyboru elementu głównego (elementy główne są zaznaczone na czerwono):

$$\begin{bmatrix} 4 & -2 & 4 & -2 \\ 3 & 1 & 4 & 2 \\ 2 & 4 & 2 & 1 \\ 2 & -2 & 4 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} \textcolor{red}{4} & -2 & 4 & -2 \\ \textcircled{\frac{3}{4}} & \frac{5}{2} & 1 & \frac{7}{2} \\ \textcircled{\frac{1}{2}} & 5 & 0 & 2 \\ \textcircled{\frac{1}{2}} & -1 & 2 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} \textcolor{red}{4} & -2 & 4 & -2 \\ \textcircled{\frac{3}{4}} & \textcircled{\frac{1}{2}} & 1 & \frac{5}{2} \\ \textcircled{\frac{1}{2}} & \textcolor{red}{5} & 0 & 2 \\ \textcircled{\frac{1}{2}} & \textcircled{-\frac{1}{5}} & 2 & \frac{17}{5} \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} \textcolor{red}{4} & -2 & 4 & -2 \\ \textcircled{\frac{3}{4}} & \textcircled{\frac{1}{2}} & \textcircled{\frac{1}{2}} & \frac{4}{5} \\ \textcircled{\frac{1}{2}} & \textcolor{red}{5} & 0 & 2 \\ \textcircled{\frac{1}{2}} & \textcircled{-\frac{1}{5}} & \textcolor{red}{2} & \frac{17}{5} \end{bmatrix} \rightarrow \begin{bmatrix} \textcolor{red}{4} & -2 & 4 & -2 \\ \textcircled{\frac{3}{4}} & \textcircled{\frac{1}{2}} & \textcircled{\frac{1}{2}} & \textcolor{red}{\frac{4}{5}} \\ \textcircled{\frac{1}{2}} & \textcolor{red}{5} & 0 & 2 \\ \textcircled{\frac{1}{2}} & \textcircled{-\frac{1}{5}} & \textcolor{red}{2} & \frac{17}{5} \end{bmatrix}$$

Mamy zatem:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 1 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{5} & 1 & 0 \\ \frac{3}{4} & \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 4 & -2 & 4 & -2 \\ 0 & 5 & 0 & 2 \\ 0 & 0 & 2 & \frac{17}{5} \\ 0 & 0 & 0 & \frac{4}{5} \end{bmatrix} = PLU$$

Przekształcamy nasze równanie:

$$Ax = PLUx = b$$

$$LUx = P^{-1}b = P^T b$$

Rozwiązujemy najpierw równanie  $Ly = P^T b$ :

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 1 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{5} & 1 & 0 \\ \frac{3}{4} & \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}^T \begin{bmatrix} 8 \\ 7 \\ 10 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 8 \\ 7 \\ 10 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 10 \\ 2 \\ 7 \end{bmatrix}$$

Dostajemy

$$y_1 = 8$$

$$y_2 = 10 - \frac{1}{2} \cdot 8 = 6$$

$$y_3 = 2 + \frac{1}{5} \cdot 6 - \frac{1}{2} \cdot 8 = -\frac{4}{5}$$

$$y_4 = 7 - \frac{3}{4} \cdot 8 - \frac{1}{2} \cdot 6 + \frac{1}{2} \cdot \frac{4}{5} = -\frac{8}{5}$$

Stąd  $Ux = y$ :

$$\begin{bmatrix} 4 & -2 & 4 & -2 \\ 0 & 5 & 0 & 2 \\ 0 & 0 & 2 & \frac{17}{5} \\ 0 & 0 & 0 & \frac{4}{5} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \\ -\frac{4}{5} \\ -\frac{8}{5} \end{bmatrix}$$

Zatem:

$$x_4 = -2$$

$$x_3 = \frac{-\frac{4}{5} + \frac{17}{5} \cdot 2}{2} = 3$$

$$x_2 = \frac{6 + 2 \cdot 2}{5} = 2$$

$$x_1 = \frac{8 - 2 \cdot 2 - 4 \cdot 3 + 2 \cdot 2}{4} = -1$$

c) Użyjemy całkowitego wyboru elementu głównego.

$$\begin{aligned}
 \begin{bmatrix} 1 & 0 & 1 \\ 2 & 6 & 3 \\ 1 & 3 & 3 \end{bmatrix} &\rightarrow \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 6 & 3 \\ 1 & 0 & 1 \\ 1 & 3 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 6 & 2 & 3 \\ 0 & 1 & 1 \\ 3 & 1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
 &\rightarrow \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 6 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
 &\rightarrow \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 6 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
 &\rightarrow \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 6 & 3 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
 &\rightarrow \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 6 & 3 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}
 \end{aligned}$$

Zatem

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 6 & 3 \\ 1 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 6 & 3 & 2 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = PLUQ$$

Znowu więc mamy, że:

$$\begin{aligned}
 Ax &= PLUQx = b \\
 LUQx &= P^{-1}b = P^T b
 \end{aligned}$$

Rozwiązujemy najpierw równanie  $Ly = P^T b$ :

$$\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}^T \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Stąd

$$\begin{aligned}
 y_1 &= 1 \\
 y_2 &= 2 - \frac{1}{2} \cdot 1 = \frac{3}{2} \\
 y_3 &= 3 - \frac{2}{3} \cdot \frac{3}{2} = 2
 \end{aligned}$$

Mamy więc  $UQx = y$ :

$$\begin{bmatrix} 6 & 3 & 2 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 & 3 & 2 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{3}{2} \\ 2 \end{bmatrix}$$

Stąd kolejno:

$$\begin{aligned}
 x_1 &= 2 \\
 x_3 &= \frac{3}{2} \cdot \frac{2}{3} = 1 \\
 x_2 &= \frac{1 - 2 \cdot 2 - 3 \cdot 1}{6} = -1
 \end{aligned}$$

**Zad. 9.2** Znaleźć rozkład LU dla macierzy

a)

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

b)

$$\begin{bmatrix} 2 & 1 & 1 & 1 \\ 4 & 3 & 5 & 6 \\ 6 & 5 & 11 & 6 \\ 4 & 3 & 9 & 8 \end{bmatrix}$$

**Zad. 9.3** Rozwiąż układ równań:

a)

$$\begin{aligned} 3x_1 + x_2 &= -3 \\ x_1 + 2x_2 - x_3 &= 0 \\ -x_2 + 3x_3 &= 0 \end{aligned}$$

b)

$$\begin{aligned} 3x_1 + x_2 + 6x_3 &= 2 \\ 2x_1 + x_2 - 3x_3 &= 7 \\ x_1 + x_2 + x_3 &= 4 \end{aligned}$$

## 10. Iteracyjne rozwiązywanie układów liniowych

Naszym celem jest znalezienie (prawdopodobnie przybliżonego) rozwiązania układu liniowego

$$Ax = b. \quad (10.1)$$

Podobnie jak przy rozwiązywaniu równań nieliniowych, rozważać będziemy metody iteracyjne, dające na wyjściu ciąg wektorów dążący do rozwiązania. Równanie możemy zapisać w równoważnej postaci:

$$Qx = (Q - A)x + b,$$

gdzie  $Q$  jest odpowiednio wybraną macierzą. To sugeruje wyznaczanie ciągu wektorów  $x^{(k)}$  z równania

$$Qx^{(k+1)} = (Q - A)x^{(k)} + b \quad (10.2)$$

Nas będą interesowały głównie przypadki, gdy ciąg wyznaczony przez równanie (10.2) jest zbieżny do rozwiązania dla każdego punktu startowego  $x^{(0)}$ . Ponadto oczywiście chcemy, żeby równanie (10.2) pozwalało łatwo wyliczyć ciąg  $x^{(k)}$ . Można pokazać, że metoda iteracyjna ma dobre właściwości, gdy  $Q^{-1}$  jest dość bliskie  $A^{-1}$ :

**Fakt 10.1.** Jeśli  $\|I - Q^{-1}A\| < 1$ , gdzie  $\|\cdot\|$  jest macierzową normą operatorową pochodzącą od normy wektorowej  $\|\cdot\|$ , to metoda iteracyjna (10.2) jest zbieżna do rozwiązania układu (10.1) dla dowolnego punktu startowego.

Większość używanych w praktyce metod używa prostego, addytywnego, rozkładu macierzy  $A$  na elementy pod przekątną  $L$ , nad przekątną  $U$  i na samą przekątną  $D$ . Mamy więc:

$$A = L + D + U.$$

*Uwaga 10.2.* Mimo używania tego samego oznaczenia, nie należy mylić tego rozkładu z rozkładem PLU z poprzedniego rozdziału.

### Metoda Richardsona

W metodzie Richardsona przyjmujemy  $Q = I$ . Zatem równanie rekurencyjne to:

$$x^{(k+1)} = (I - A)x^{(k)} + b.$$

Prowadzi to do następującego wzoru:

$$x_i^{(k+1)} = x_i^{(k)} - \sum_{j=1}^n a_{ij}x_j^{(k)} + b_i.$$

## Metoda Jacobiego

W metodzie Jacobiego przyjmujemy  $Q = D$ . Równanie  $Ax = b$  prowadzi do wzoru:

$$x^{(k+1)} = D^{-1}(D - A)x^{(k)} + D^{-1}b.$$

Zauważmy, że macierz  $D$  jako macierz diagonalną bardzo łatwo odwrócić. Prowadzi to do następującego wzoru na elementy wektora  $x^{(k+1)}$ :

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left[ b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^{(k)} \right].$$

Oczywiście metoda ta może być użyta jedynie gdy na przekątnej nie ma zer. Zauważmy, że elementy wektora  $x^{(k)}$  można obliczać równoległe. Liczbę potrzebnych obliczeń można zmniejszyć mnożąc wyjściowe równanie obustronnie przez  $D^{(-1)}$ .

## Metoda Gaussa-Seidela

W metodzie Gaussa-Seidela przyjmujemy  $Q = (L + D)$ . Równanie  $Ax = b$  prowadzi więc do równania:

$$x^{(k+1)} = -(L + D)^{-1}Ux^{(k)} + (L + D)^{-1}b.$$

Ponieważ macierz  $(L + D)^{-1}$  jest dolnotrójkątna, można ją więc stosunkowo łatwo odwrócić (znowu: jedynie wtedy, gdy przekątna nie zawiera zer). Prowadzi to do następującego wzoru na elementy wektora  $x^{(k+1)}$ :

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left[ b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right].$$

Równanie to jest bardzo podobne do równania z metody Jacobiego, tylko tam gdzie to możliwe w rachunkach używamy elementów  $x^{(k+1)}$ , a nie  $x^{(k)}$ .

Po bliższym przyjrzeniu można zauważyć, że nie trzeba przechowywać osobno wektorów  $x^{(k)}$  i  $x^{(k+1)}$ . Odpowiednie elementy można po kolei zastępować, niestety obliczeń nie można prowadzić równoległe.

## Nadrelaksacja

W praktyce bardzo dobrze sprawuje się metoda iteracyjna używająca  $Q = D + \omega L$  do rozwiązania układu  $\omega Ax = \omega b$ . Dostajemy równanie:

$$(D + \omega L)x^{(k+1)} = -[\omega U + (\omega - 1)D]x^{(k)} + \omega b.$$

Prowadzi to do następującego wzoru:

$$x_i^{(k+1)} = \frac{\omega}{a_{ii}} \left[ b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right] + (1 - \omega)x_i^{(k)}.$$

Metoda ma sens dla każdego  $\omega \neq 1$ , a dla  $\omega = 1$  pokrywa się z metodą Gaussa-Seidela.

## Kilka faktów

Dla zbieżności metody Jacobiego i Gaussa-Seidela mamy następujące twierdzenie:

**Twierdzenie 10.3.** Jeśli  $A$  jest macierzą przekątniowo dominującą (tzn.

$$|a_{ii}| > \sum_{\substack{j=1, \dots, n \\ j \neq i}} |a_{ij}|$$

dla wszystkich  $i$ ), to metoda Jacobiego i metoda Gaussa-Seidela są zbieżne dla dowolnego punktu startowego do rozwiązania układu (10.1).

Przypomnijmy, że jeżeli  $Ax = \lambda x$ , to  $x$  nazywamy wektorem własnym, a  $\lambda$  wartością własną macierzy  $A$ . Wartości własne są pierwiastkami tzw. wielomianu charakterystycznego macierzy danego wzorem  $P_A(\lambda) = \det(\lambda I - A)$ . Promieniem spektralnym  $\rho(A)$  macierzy  $A$  nazywamy promień najmniejszego koła o środku w zerze zawierający wszystkie wartości własne.

**Twierdzenie 10.4.** Równanie rekurencyjne  $x^{(k+1)} = Cx^{(k)} + c$  daje ciąg zbieżny do  $(I - C)^{-1}c$  dla dowolnego wektora  $x^{(0)}$  wtedy i tylko wtedy, gdy  $\rho(C) < 1$ .

**Twierdzenie 10.5.** Załóżmy, że dla nieosobliwej macierzy  $A$  macierz  $D^{-1}(D - A)$  występująca w metodzie Jacobiego ma tylko rzeczywiste wartości własne i promień spektralny  $\mu$  mniejszy od 1. Wtedy promień spektralny macierzy występującej w metodzie nadrelaksacji wyraża się wzorem:

$$\rho(A_\omega) = \begin{cases} \frac{1}{4} \left( \omega\mu + \sqrt{\omega^2\mu^2 - 4(\omega - 1)} \right)^2, & 0 < \omega \leq \omega_0, \\ \omega - 1, & \omega_0 < \omega < 2, \end{cases}$$

gdzie optymalny parametr  $\omega_0$  dla którego promień spektralny jest najmniejszy wynosi  $\omega_0 = 1 + \left( \frac{\mu}{1 + \sqrt{1 - \mu^2}} \right)^2$ .

## Zadania

**Zad. 10.1** Napisz w dowolnym znaym Ci języku programowania implementację metody Richardsona.

**Zad. 10.2** Napisz w dowolnym znaym Ci języku programowania implementację metody Jacobiego.

**Zad. 10.3** Napisz w dowolnym znaym Ci języku programowania implementację metody Gaussa-Seidela.

**Zad. 10.4** Napisz w dowolnym znaym Ci języku programowania implementację metody nadrelaksacji.

**Zad. 10.5** Wypróbuj stworzone metody do rowiązania układu:

$$\begin{cases} x_1 + 9x_2 + x_3 = 5 \\ 4x_1 + x_2 - x_3 = 3 \\ x_1 - 3x_2 + 12x_3 = 31 \end{cases}$$

Rozwiąż najpierw układ tak jak jest zapisany, a następnie przestaw równania aby macierz układu była przekątniowo dominująca.

**Zad. 10.6** Zbadaj najpierw eksperymentalnie zachowanie metod Jacobiego i Gaussa-Seidela dla układu o macierzy

$$\begin{cases} 2x_1 + x_2 + x_3 = 4 \\ x_1 + 2x_2 + x_3 = 4 \\ x_1 + x_2 + 2x_3 = 4 \end{cases},$$

gdę startujemy od wektora zerowego. Następnie znajdź promienie spektralne odpowiednich macierzy i wytłumacz zachowanie obu metod.

**Zad. 10.7** Znajdź bez użycia komputera wartości własne macierzy

$$\begin{bmatrix} 4 & 3 \\ 2 & 5 \end{bmatrix}.$$

Ile wynosi promień spektralny tej macierzy?

**Zad. 10.8** Niech

$$A = \begin{bmatrix} 6 & 4 & -3 \\ 2 & 3 & 1 \\ 2 & 2 & 10 \end{bmatrix}.$$

Znajdź promienie spektralne macierzy używanych w metodzie Jacobiego, Gaussa-Seidela i w metodzie nadrelaksacji (dla różnych wartości parametru  $\omega$ ). Sprawdź w praktyce podane twierdzenie dla promienia spektralnego macierzy metody nadrelaksacji.