

Metody Optymalizacji, Wykład 1

Temat: Optymalizacja analityczna (bez ograniczeń)

5 października 2022 r.

mgr Daniel Kaszyński

Kwestie organizacyjne

Zespół prowadzących:

- dr Grzegorz Kołoch (<https://web.sgh.waw.pl/~gkoloch/> (<https://web.sgh.waw.pl/~gkoloch/>))
- mgr. Daniel Kaszyński (<https://kaszyn.ski/> (<https://kaszyn.ski/>))

Literatura:

- Chong, E.K. and Zak, S.H., 2004. **An introduction to optimization**. John Wiley & Sons.
- Kochenderfer, M.J. and Wheeler, T.A., 2019. **Algorithms for optimization**. Mit Press.
- Sydsæter, K., Hammond, P., Seierstad, A. and Strom, A., 2008. **Further mathematics for economic analysis**. Pearson education.

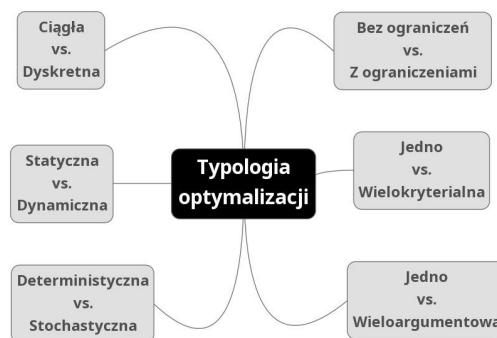
Rozkład spotkań:

Data	Temat
2022-10-05	Optymalizacja analityczna (bez ograniczeń)
2022-10-19	Optymalizacja analityczna (z ograniczeniami)
2022-11-02	Wstęp do obliczeń numerycznych
2022-11-16	Metody Bracketingowe
2022-11-30	Metody Gradientowe
2022-12-14	Metody wyższego rzędu
2023-01-04	Metaheurystyki / metody zaawansowane
2023-01-18	Zerówka

Czemu ustrukturyzowane podejście do optymalizacji?

1. Problem Komiwojażera dla 100 miast.
2. Funkcja $\operatorname{argmin} f(x, y)$.
3. Kształtowanie polityki połowów ryb w danym obszarze.
4. Kalibracja głębokiej sieci neuronowej, sterującej autonomicznym samochodem.

Topologia optymalizacji



Podejście do optymalizacji nieliniowej (analitycznej)

W przypadku rozwiązywania problemów optymalizacji nieliniowej, przy wykorzystaniu narzędzi analitycznych, wykorzystywane są tzw. **Warunki Pierwszego Rzędu** (ang. FOC - First Order Conditions), oraz **Warunki Drugiego Rzędu** (ang. SOC - Second Order Conditions).

FOC

Warunki Pierwszego Rzędu definiują **warunki konieczne** zaistnienia ekstremum w danym punkcie. W przypadku tych warunków, w dowodzie wychodzimy od definicji ekstremum, i wskazujemy na następstwa występowania w danym punkcie (x^*) ekstremum.

Struktura tych twierdzeń / warunków jest następująca: *Jeżeli w punkcie x^* jest ekstremum funkcji f , to*

Przykład FOC dla $f : \mathbb{R} \rightarrow \mathbb{R}$

Jeżeli w punkcie x^* jest ekstremum funkcji $f : \mathbb{D} \subset \mathbb{R} \rightarrow \mathbb{R}$, gdzie $\mathbb{D}$ to **zbiór otwarty** dziedziny funkcji f , to $f'(x) = \frac{df}{dx}(x) = 0$.

SOC

Warunki Drugiego Rzędu definiują **warunki dostateczne** zaistnienia ekstremum w danym punkcie. W przypadku tych warunków, w dowodzie wychodzimy od warunków FOC, do których dodawane są warunki dot. lokalnej krzywizny funkcji, w okół analizowane punktu (x^*).

Struktura tych twierdzeń / warunków jest następująca: **Jeżeli spełnione są warunki FOC, oraz w punkcie x^* spełnia warunki SOC, to w punkcie x^* jest minimum / maksimum funkcji f .*

Przykład SOC dla $f : \mathbb{R} \rightarrow \mathbb{R}$

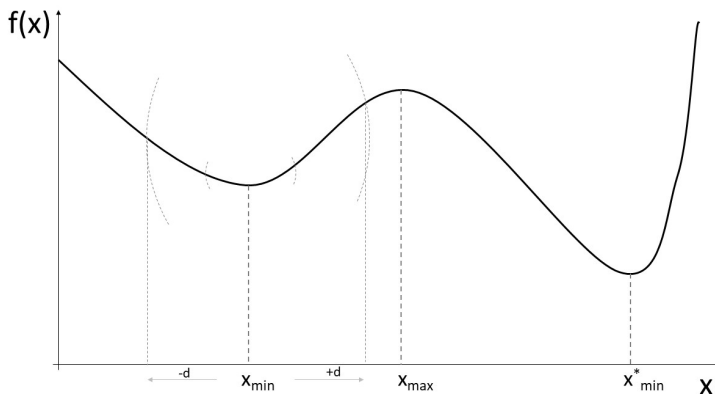
Jeżeli w punkcie x^* funkcji $f : \mathbb{D} \subset \mathbb{R} \rightarrow \mathbb{R}$, gdzie $\mathbb{D}$ to zbiór otwarty dziedziny funkcji f , $f'(x) = \frac{df}{dx}(x) = 0$, oraz dla pewnego $n = \{2, 3, \dots\}$ zachodzi $f^n \neq 0$, to:

- jeżeli n jest parzysta, i $f^n > 0$ to w punkcie x^* jest minimum funkcji f .
- jeżeli n jest parzysta, i $f^n < 0$ to w punkcie x^* jest maksimum funkcji f .
- jeżeli n jest nieparzysta, to w punkcie x^* nie ma ekstremum funkcji f .

Definicja ekstremum

Terminem centralnym w zakresie optymalizacji jest pojęcie ekstremum - rozwiązania, które dla zdefiniowanej funkcji oceniającej (tzw. **funkcji celu**, najczęściej oznaczanej jako f)

Przykład $f : \mathbb{R} \rightarrow \mathbb{R}$



Definicja lokalnego ekstremum (właściwego)

Lokalnym ekstremum typu minimum właściwe nazwiemy punkt x^* , który:

$$\exists_{|d|>0} \forall_{|d|>|\delta|>0} f(x^* + \delta) > f(x^*)$$

Definicja globalnego ekstremum (właściwego)

Globalnym ekstremum typu minimum właściwe nazwiemy punkt x^* , który:

$$\forall_{|d|>0} \forall_{|\delta|>0} f(x^* + \delta) > f(x^*)$$

Przydatne definicje i twierdzenia

Definicja pochodnej

Pochodną funkcji $f : \mathbb{R} \rightarrow \mathbb{R}$ nazwiemy funkcję:

$$f'(x) = \frac{df}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{f(x) - f(x-h)}{h}$$

Ciągłość funkcji f jest warunkiem koniecznym różniczkowalności!

W analogiczny sposób można wyprowadzić wzór na pochodną dowolnego rzędu:

$$f''(x) = \frac{d^2 f}{dx^2} = (f'(x))' = \lim_{h \rightarrow 0} \frac{f'(x+h) - f'(x)}{h} = \lim_{h \rightarrow 0} \frac{f(x+2h) - 2f(x+h) + f(x)}{h^2}$$

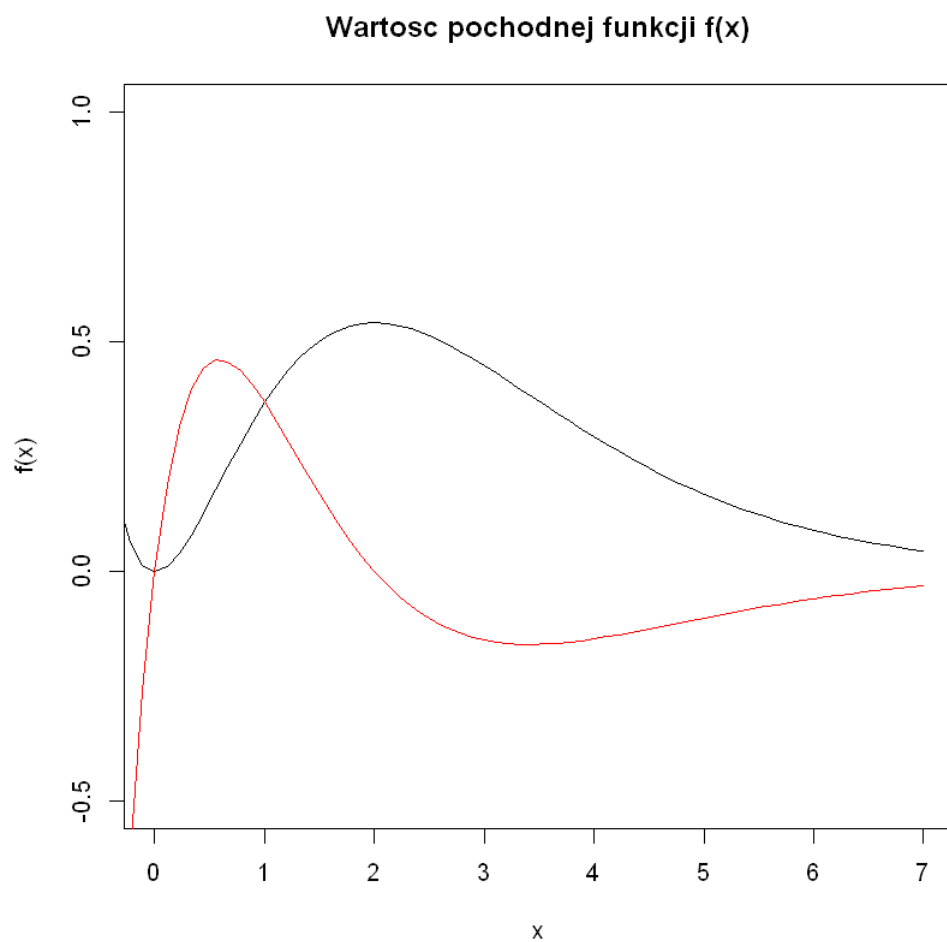
Przykład

Niech $f(x) = \frac{x^2}{\exp(x)}$.

```

In [1]: 1 my_fun <- function(x) x^2 / exp(x);
2
3 # UWAGA! Pochodna numeryczna, tzw. forward finite difference
4 df <- function(f, x, h = 10^-6){
5   return((f(x + h) - f(x)) / (h))
6 }
7
8 x_seq <- seq(-4, 7, length = 100)
9
10 # Wykres funkcji my_fun
11 plot(x_seq, my_fun(x_seq),
12      type = 'l',
13      col = 'black',
14      xlim=c(0, 7),
15      ylim=c(-0.5, 1),
16      ylab = 'f(x)',
17      xlab='x',
18      main='Wartość pochodnej funkcji f(x)')
19
20 # Wykres Pochodnej
21 lines(x_seq, df(my_fun, x_seq),
22      type = 'l',
23      col = 'red')

```



Twierdzenie Taylora

Niech $f : \mathbb{D} \subset \mathbb{R} \rightarrow \mathbb{R}$ oraz $f \in C^{n+1}$ w każdym punkcie odcinka $[x, x+h]$. Wówczas dla pewnego θ mamy:

$$f(x+h) = f(x) + \sum_{k=1}^n \frac{1}{k!} f^{(k)}(x) h^k + \frac{1}{(n+1)!} f^{(n+1)}(x+\theta h) h^{n+1}$$

$$f(x+h) \approx f(x) + \sum_{k=1}^n \frac{1}{k!} f^{(k)}(x) h^k$$

$$f(x+h) \approx f(x) + f'(x)h$$

$$f(x+h) \approx f(x) + f'(x)h + \frac{1}{2} f''(x)h^2$$

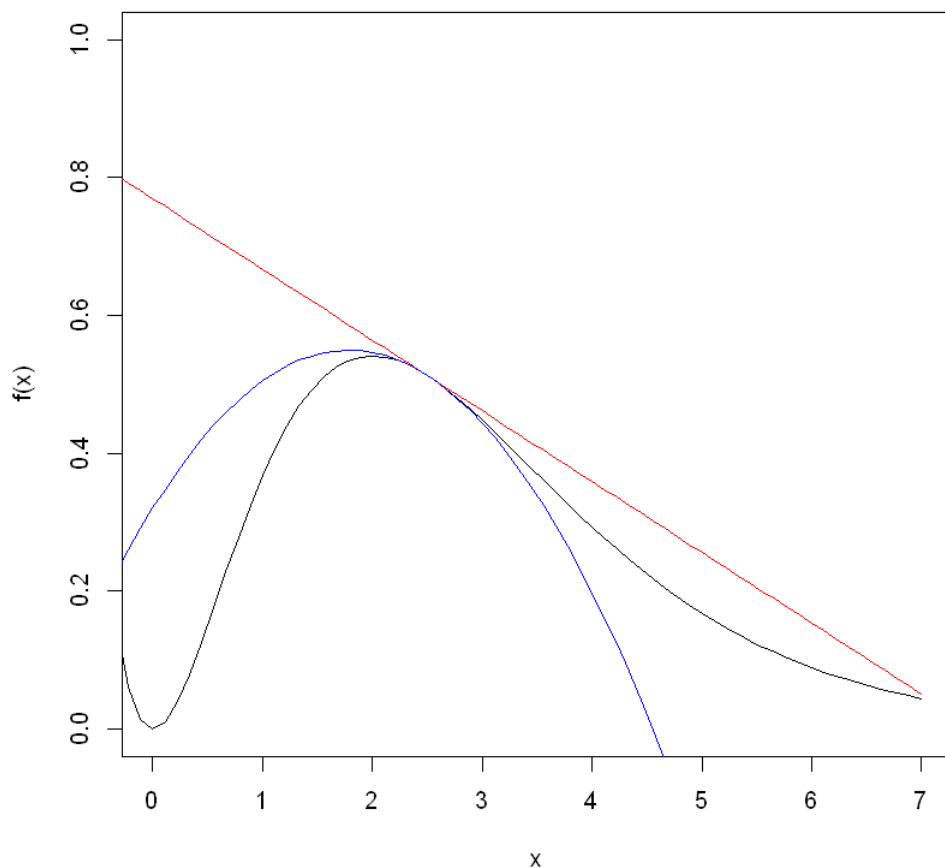
Przykład

```

In [5]: 1 my_fun <- function(x) x^2 / exp(x);
        2
        3 x = 2.5
        4
        5 df <- function(f, x, h = 10^-6){
        6   return((f(x + h) - f(x)) / (h))
        7 }
        8
        9 d2f <- function(f, x, h = 10^-6){
       10   return((f(x + 2*h) - 2*f(x+h) + f(x))/(h^2))
       11 }
       12
       13 taylor_1 <- function(f, x, h){
       14   return(f(x) + df(f, x)*(h - x))
       15 }
       16
       17 taylor_2 <- function(f, x, h){
       18   return(f(x)+df(f, x)*(h-x)+1/2*d2f(f, x)*(h-x)^2)
       19 }
       20
       21
       22 x_seq <- seq(-4, 7, length = 100)
       23
       24 # Wykres funkcji my_fun
       25 plot(x_seq, my_fun(x_seq),
       26   type = 'l',
       27   col = 'black',
       28   xlim=c(0, 7),
       29   ylim=c(0, 1),
       30   ylab = 'f(x)',
       31   xlab='x',
       32   main='Aproksymacja Taylora')
       33
       34 # xlim=c(2, 3),
       35 # ylim=c(0.45, 0.55),
       36
       37 # Wykres aproksymacji Taylora 1
       38 lines(x_seq, taylor_1(my_fun, x, x_seq),
       39   type = 'l',
       40   col = 'red')
       41
       42 # Wykres aproksymacji Taylora 2
       43 lines(x_seq, taylor_2(my_fun, x, x_seq),
       44   type = 'l',
       45   col = 'blue')

```

Aproksymacja Taylora



Pochodna kierunkowa

Niech $f : \mathbb{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $x \in \mathbb{D}$ oraz $h \in \mathbb{R}^n : x + h \in \mathbb{D}$. Pochodną kierunkową funkcji f w punkcie x w kierunku h nazywamy funkcję:

$$\frac{df}{dh}(x) = \lim_{t \rightarrow 0} \frac{f(x+th) - f(x)}{t}$$

Pochodna cząstkowa

Niech $f : \mathbb{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $x \in \mathbb{D}$ oraz $h \in \mathbb{R}^n : x+h \in \mathbb{D}$. Pochodną cząstkową funkcji f w punkcie x względem zmiennej x_i , $i = 1, 2, \dots, n$ nazywamy funkcję:

$$\frac{\partial f}{\partial x_i}(x) = \frac{df}{de_i}(x)$$

gdzie e_i jest i tym wersorem przestrzeni $\mathbb{R}^n$. Pochodna cząstkowa f względem x_i jest więc pochodną kierunkową f w kierunku i tego wersora, tj. przy $h = e_i$.

Gradient

Niech $f : \mathbb{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $x \in \mathbb{D}$. Gradientem funkcji f w punkcie x nazywamy funkcję:

$$\nabla_f(x) = \left[\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right]$$

Optymalizacja nieliniowa bez ograniczeń, przypadek *jednowymiarowy*, tj. $f : \mathbb{R} \rightarrow \mathbb{R}$

Twierdzenie FOC

Niech $f : \mathbb{D} \subset \mathbb{R} \rightarrow \mathbb{R}$, $f \in C^1$. Jeżeli funkcja f posiada ekstremum w punkcie $x \in \mathbb{D}$, to $f'(x) = 0$.

Szkic dowodu:

Ponieważ funkcja f , a w punkcie x minimum, wiemy, że istnieje $d > 0$, takie, że dla każdego $0 < |\delta| < d$, mamy $f(x+\delta) > f(x)$, czyli $f(x+\delta) - f(x) > 0$. Dzieląc obie strony przez δ , otrzymujemy:

$$\frac{f(x+\delta) - f(x)}{\delta} > 0 \quad \text{oraz} \quad \frac{f(x+\delta) - f(x)}{\delta} < 0$$

odpowiednio dla $\delta > 0$ i $\delta < 0$. Odpowiednio przy przejściu granicznym $\delta \rightarrow 0^+$ i $\delta \rightarrow 0^-$ mamy:

$$\lim_{\delta \rightarrow 0^+} \frac{f(x+\delta) - f(x)}{\delta} = f'_+(x) \geq 0 \quad \text{oraz} \quad \lim_{\delta \rightarrow 0^-} \frac{f(x+\delta) - f(x)}{\delta} = f'_-(x) \leq 0$$

Ponieważ jednak funkcja f jest różniczkowalna, mamy $f'(x) = f'_+(x) = f'_-(x) = 0$.

Twierdzenie SOC

Niech $f : \mathbb{D} \subset \mathbb{R} \rightarrow \mathbb{R}$, $f \in C^n$. Jeżeli dla pewnego $x \in \mathbb{D}$ zachodzi: $f'(x) = 0$, $f''(x) = 0$, $\dots$, $f^{(n-1)}(x) = 0$, ale $f^{(n)}(x) \neq 0$, to:

- jeśli n jest parzyste, to funkcja f ma w punkcie x ekstremum; jeśli $f^{(n)}(x) > 0$ to jest to minimum, $f^{(n)}(x) < 0$ to jest to maksimum
- jeśli n jest nieparzyste, to funkcja f ma w punkcie x nie ma ekstremum.

Szkic dowodu:

Ze wzoru Taylora dla pewnego $0 < \theta < 1$ mamy:

$$f(x+h) = \sum_{k=0}^{n-1} \frac{1}{k!} f^{(k)}(x) h^k + \frac{1}{(n)!} f^{(n)}(x+\theta h) h^n$$

a ponieważ $f'(x) = 0$, $f''(x) = 0$, $\dots$, $f^{(n-1)}(x) = 0$ to:

$$f(x+h) = f(x) + \frac{1}{n!} f^{(n)}(x+\theta h) h^n$$

$$f(x+h) - f(x) = \frac{1}{n!} f^{(n)}(x+\theta h) h^n$$

Kiedy n jest parzyste, i $f^{(n)}(x) > 0$, ze względu na parzystość n , mamy $h^n > 0$. Ze względu na ciągłość funkcji $f^{(n)}$ w punkcie n wiemy, że istnieje $\delta > 0$ taka, że dla każdego $h : 0 < |h| < \delta$ mamy $f^{(n)}(x+h) > 0$, a więc także $f^{(n)}(x+\theta h) > 0$. Oznacza to, że funkcja f ma w tym miejscu minimum.