131890-S Metody Optymalizacji **Letni 2023/2024**

Zajęcia 1: Optymalizacja analityczna bez ograniczeń

Daniel Kaszyński

Kwestie organizacyjne

- **Prowadzący zajęcia**: mgr Daniel Kaszyński, dkaszy[@]sgh.waw.pl
- **Konsultacje**: poniedziałek 11:00-12:00, G-213.
- **Tryb zaliczenia przedmiotu**: egzamin pisemny w sesji z treści przedstawianych w trakcie zajęć lub zawartych w materiałach + zerówka na ostatnim wykładzie.

  Przykłady zadań: opisz metodę XYZ, wskaż na różnice między metodami ABC i XYZ, rozwiąż przedstawione zadanie optymalizacji analitycznej, przeprowadź dwie iteracje metody XYZ.
- **Wymagana literatura**:
  - [KW19] Kochenderfer, M.J. and Wheeler, T.A., 2019. *Algorithms for optimization.* Mit Press.
  - [CZ04] Chong, E.K. and Zak, S.H., 2004. *An introduction to optimization.* John Wiley & Sons.
  - [BI86] Birkholc, A., 1986. *Analiza matematyczna: funkcje wielu zmiennych.* Państwowe Wydawnictwo Naukowe.
  - [SS08] Sydsæter, K., Hammond, P., Seierstad, A. and Strom, A., 2008. *Further mathematics for economic analysis.* Pearson education.
  - [CO14] Cortez, P., 2014. *Modern optimization with R.* New York: Springer.

1.1 Definicja ekstremum

Terminem centralnym w zakresie optymalizacji analitycznej jest pojęcie ekstremum: rozwiązania, które dla zdefiniowanej funkcji oceniającej, tzw. **funkcji celu**, najczęściej oznaczanej jako f, generuje 'najlepszą' wartość. Rozważmy przykład $f : \mathbb{R} \to \mathbb{R}$.

Definicja 1: Definicja lokalnego ekstremum (właściwego)

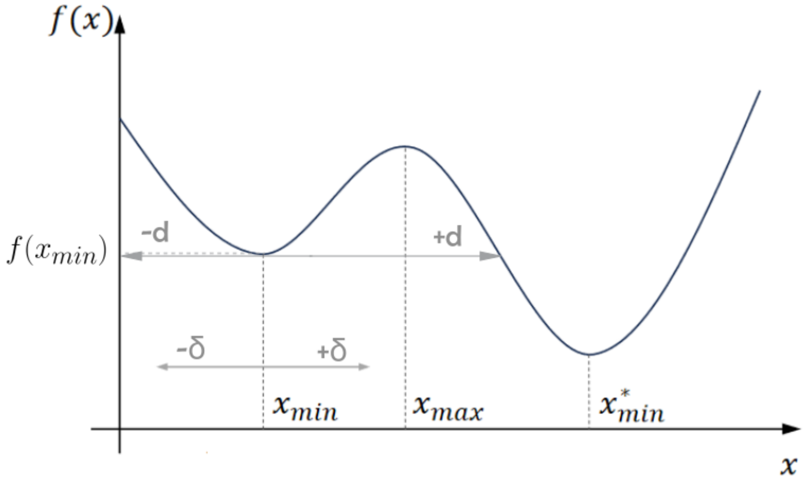
Lokalnym ekstremum typu minimum właściwe nazwiemy punkt x^*, który:

$$\exists\, d > 0 \quad \forall\, 0 < |\delta| < |d| \quad \Rightarrow \quad f(x^* + \delta) \geqslant f(x^*) \tag{1.1}$$

Definicja 2: Definicja globalnego ekstremum (właściwego)

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

$$\forall\, d > 0 \quad \forall\, |\delta| > 0 \quad \Rightarrow \quad f(x^* + \delta) > f(x^*) \tag{1.2}$$



Rysunek 1.1: Ekstremum lokalne funkcji $f : \mathbb{R} \rightarrow \mathbb{R}$

Zauważmy, że różnica między ekstremum *lokalnym* i *globalny* to obszar, w którym uzyskujemy *lepsze* (tj. z perspektywy funkcji celu, funkcji oceniającej) rozwiązanie. W przypadku ekstremum lokalnego, można wskazać otoczenie (może być to bardzo mały obszar!) sąsiedztwa w okół ekstremum x_{min}, w którym utrzymujemy stricte *gorsze* rozwiązania. W przypadku ekstremum *globalnego* x^*_{min}, takie sąsiedztwo to dowolne otoczenie rozwiązania ekstremum.

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

W zakresie podstawowych twierdzeń, musimy wpierw wskazać podstawową dla nas definicję – definicję pochodnej funkcji jednej zmiennej.

Definicja 3: Pochodna funkcji

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

$$f'(x) = \frac{df}{dx}(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{f(x) - f(x-h)}{h} \tag{1.3}$$

Przykład 1. Przyjmijmy $f(x) = x^2 + 2x$. Wyznacz pochodną funkcji w punkcie $x_0 = 2$ z definicji:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{(x+h)^2 + 2(x+h) - x^2 - 2x}{h} = \tag{1.4}$$

$$= \lim_{h \to 0} \frac{x^2 + 2xh + h^2 + 2x + 2h - x^2 - 2x}{h} = \lim_{h \to 0} \frac{2xh + h^2 + 2h}{h} = \tag{1.5}$$

$$= \lim_{h \to 0} 2x + h + 2 = 2x + 2 \tag{1.6}$$

$$f'(2) = 2 \times 2 + 2 = 6 \tag{1.7}$$

Ciągłość funkcji f jest warunkiem koniecznym różniczkowalności!

W analogiczny sposób można wyprowadzić wzór na pochodną dowolnego rzędy:

$$f''(x) = \frac{d^2 f}{dx^2}(x) = (f'(x))' = \lim_{h\to 0} \frac{f'(x+h) - f'(x)}{h} = \lim_{h\to 0} \frac{f(x+2h) - 2f(x+h) + f(x)}{h^2} \tag{1.8}$$

1.2.1 Warunki Pierwszego Rzędu $f : \mathbb{R} \to \mathbb{R}$

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

Dowód. Ponieważ funkcja f ,a w punkcie x minumim, 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:

$$\frac{f(x+\delta) - f(x)}{\delta} > 0 \qquad \wedge \qquad \frac{f(x+\delta) - f(x)}{\delta} < 0$$

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

$$\lim_{\delta\to 0^+} \frac{f(x+\delta) - f(x)}{\delta} = f'_+(x) \geqslant 0 \qquad \wedge \qquad \lim_{\delta\to 0^-} \frac{f(x+\delta) - f(x)}{\delta} = f'_-(x) \leqslant 0$$

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

Warunki Pierwszego Rzędu dają nam możliwość przefiltrowania przestrzeni rozwiązań, umożliwiając selekcję rozwiązań, w których pochodna funkcji jest równa zero (tzw. punkty *stacjonarne*). **Uwaga!** To, że w danym punkcie mamy $f'(x) = 0$ nie oznacza, że w tym punkcie występuje ekstremum funkcji. Jako przykład rozważ funkcje $f(x) = x^2$ oraz $f(x) = x^3$.

1.2.2 Twierdzenie Taylora $f : \mathbb{R} \to \mathbb{R}$

Żebyśmy mogli ruszyć dalej, potrzebne nam jest jedno z istotniejszych twierdzeń analizy matematycznej!

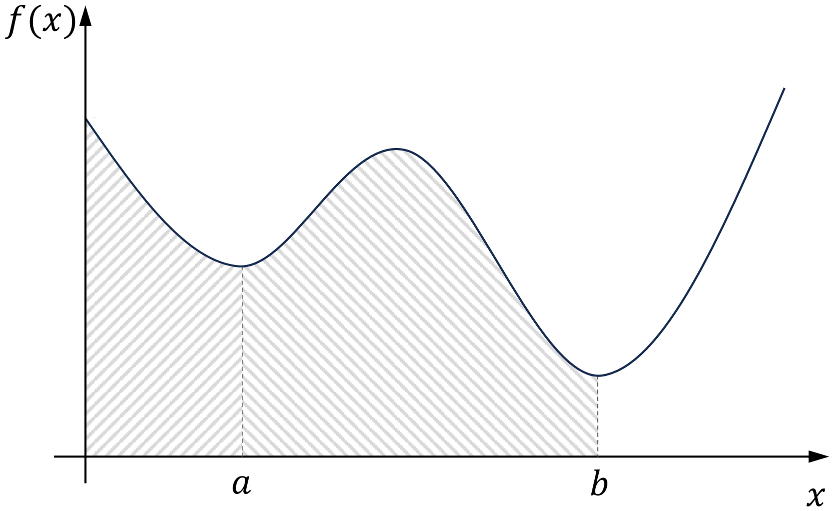
Rozważmy **Podstawowe twierdzenie rachunku całkowego**, które brzmi następująco:

Twierdzenie 2: Fundamental theorem of calculus.

$$\int_a^b f(x)\, dx = F(b) - F(a) \tag{1.9}$$

gdzie $F(x)$ to całka nieoznaczona wywołana w x. W skrócie powyższe oznacza, że żeby wyznaczyć pole powierzchni pod wykresem funkcji f na przebiegu między a i b, należy obliczyć: pole powierzchni od $-\infty$ do a, pole powierzchni pod f od $-\infty$ do b, i odjąć te dwie wartości. Poniżej intuicja tego wzoru.

Dla uproszczenia oznaczeń, przyjmijmy, że $\int f'(x)dx = f(x)$, wtedy powyższy wzór zmienia się w:



Rysunek 1.2: Całka oznaczona
Źródło: Opracowanie własne na podstawie

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

Uzyskujemy dalej taką samą intuicję, przy czym trochę uprościliśmy przyjmowaną notację. Idąc dalej, przyjmijmy następujące oznaczenia: niech $a = x$, $b = x + h$. Mówimy tyle: punkt a, to jest punkt startowy (porównawczy, referencyjny), a punkt b to punkt uzyskany po przesunięci o wartość h względem punktu a. Zwróć uwagę, że modyfikując wartość h, przyjmijmy że ta wartość jest równa zero, i zaczynamy ją coraz bardziej zwiększać, powyższy wzór odpowiada na pytanie: jak się zwiększa pole powierzchni pod krzywą funkcji f, ze względu na zwiększanie odległości od punktu referencyjnego (czyli zwiększając h).

$$f(x + h) - f(x) = \int_x^{x+h} f'(a)da$$

Porządkując powyższe równanie, oraz sprowadzając całkę oznaczoną do początku układu współrzędnych (teraz jest sztucznie zaczepiona w punkcie x – zwróć uwagę, że początek całki jak i jej koniec są funkcjami x), otrzymujemy:

$$f(x + h) = f(x) + \int_0^h f'(x + a)da \tag{1.10}$$

Powyższy wzór 1.10 jest istotny z perspektywy naszych dalszych przekształceń. Można zauważyć w nim, że x jest interpretowany jako stała wartość (patrz: całka jest po a). Co więcej, można zauważyć że wyrażenie po lewej stronie równości wstępuje w zbliżonej formie pod znakiem całki. Spróbujmy zatem wyrażenie pod całką rozwinąć zgodnie ze wzorem który otrzymaliśmy:

$$f(x + h) = f(x) + \int_0^h \left[f'(x) + \int_0^a f''(x + b)db \right] da$$

Korzystając z własności addytywności całki, uzyskujemy:

$$f(x+h) = f(x) + \int_0^h f'(x)da + \int_0^h \left[\int_0^a f''(x+b)db\right] da$$

Spróbujmy rozpisać pierwszą z całek w powyższym wzorze:

$$\int_0^h f'(x)da = [f'(x)a]_0^h = f'(x)h$$

Czyli łącznie uzyskujemy:

$$f(x+h) = f(x) + f'(x)h + \int_0^h \int_0^a f''(x+b)db\, da$$

Uwaga: przecież wyrażenie pod znakiem podwójnej całki możemy również rozpisać przy pomocy naszej wcześniejszej formuły:

$$f(x+h) = f(x) + f'(x)h + \int_0^h \int_0^a f''(x) + \left[\int_0^b f''(x+c)dc\right] db\, da$$

Spróbujmy teraz rozpisać ponownie, pierwszą całkę we wzorze:

$$\int_0^h \int_0^a f''(x)db\, da = \int_0^h [f''(x)b]_0^a\, da = \int_0^h f''(x)a\, da = \int_0^h \left[\frac{1}{2}f''(x)a^2\right]_0^h = f''(x)h^2$$

Łącząc to wyprowadzenie z naszym wcześniejszym równaniem, otrzymujemy:

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \int_0^h \int_0^a \int_0^b f''(x+c)dc\, db\, da$$

Łatwo zauważyć, że podobnie jak wcześniej – wyrażenie w całce jest rozpisywalne przy pomocy naszego wcześniejszego równania 1.10; co więcej, taką procedurę można powtarzać w nieskończoność (technicznie, tyle razy ile razy różniczkowalna jest funkcja $f(x)$ – w powyższym wzorze to łatwo zauważyć). W ogólności wzór ten można sprowadzić do:

$$f(x+h) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!}h^n$$

Powyższy wzór jest nazywany *równaniem / wzorem Taylora.*

Zwróćmy uwagę, że w praktycznym wykorzystaniu tego wzoru, często nie będziemy mogli nieskończenie wiele razy różniczkować f: możemy to zrobić kilka razy, ale od pewnego progu bólu przestaniemy już stosować ten wzór. Stad można spróbować rozpisać ten wzór dla konkretnego, N-tego rozwinięcia funkcjo wzorem Taylora:

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

gdzie $R_N(x,h) = \frac{f^{(N)}(x+\theta h)}{N!}h^N$ jest resztą Lagrange'a. Można również pokazać, że reszta ta ma następującą własność:

$$\lim_{h\to 0} \frac{R_N(x,h)}{h^N} = 0$$

Czyli reszta $R_N(x,h)$ aproksymacji wzorem Taylora maleje do 0 w tempie szybszym niż wielomian n-tego stopnia.

Twierdzenie 3: Wzór Taylora $f : \mathbb{R} \to \mathbb{R}$.
Niech $f : \mathbb{D} \subset \mathbb{R} \to \mathbb{R}$ oraz $f \in C^N$ w każdym punkcie odcinka $[x, x+h]$. Wówczas dla pewnego θ mamy:

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

Twierdzenie Taylora to bardzo ważny istotny wynik, wykorzystywany często!

Rozwinięcie funkcji w szereg Taylora 1 i 2 stopnia:

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

Przykład 2. Rozwiń funkcję $f(x) = \sin x$ w szereg Taylora w okolicy punktu $x_0 = 0$.

Wyznacz pochodne funkcji:

$$\begin{aligned}(\sin x)' &= \cos x\\ (\sin x)'' &= -\sin x\\ (\sin x)''' &= -\cos x\\ (\sin x)^{(4)} &= \sin x\end{aligned}$$

Można zauważyć, że wzorzec te powtarza się dla pochodnych wyższego rzedu. Wyznaczym teraz pochodną w punkcie 0.

$$\begin{aligned}\sin 0 &= 0\\ (\sin 0)' &= 1\\ (\sin 0)'' &= 0\\ (\sin 0)''' &= -1\\ (\sin 0)^{(4)} &= 0\end{aligned}$$

Używając wzoru Taylora otrzymujemy:

$$\begin{aligned}sin(x) &= 0 + 1x - 0x^2 + \frac{-1}{3!}x^3 + 0x^4 + \dots\\ &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots\end{aligned}$$

Przy wykorzystaniu języka **R** rozwińmy przykładową funkcję w szereg Taylora 1 i 2 stopnia: $f(x) = \frac{x^2}{e^x}$:

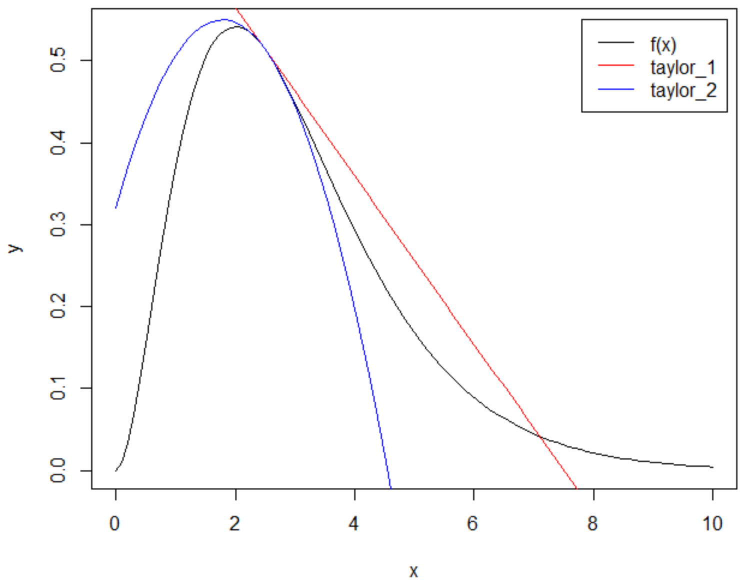
```r
# Dane wejsciowe
f  <- function(x) x^2/exp(x)
x0 <- 2.5
h_seq <- seq(0, 10, length = 100)

# Pochodne numeryczne
d1f <- function(f, x, h = 10^-6) (f(x+h)-f(x))/h
d2f <- function(f, x, h = 10^-6) (f(x+2*h)-2*f(x+h)+f(x))/h^2

# Aproksymacja Taylora funkcji f wokol x0
taylor_1 <- function(f, x, h) f(x)+d1f(f, x)*(h-x)
taylor_2 <- function(f, x, h) f(x)+d1f(f, x)*(h-x)+1/2*d2f(f, x)*(h-x)^2

# Wykresy
plot(h_seq, f(h_seq), type='l', col='black', xlab = 'x', ylab = 'y')
lines(h_seq, taylor_1(f, x0, h_seq), col='red')
lines(h_seq, taylor_2(f, x0, h_seq), col='blue')
legend(7.8, 0.55, legend=c('f(x)', 'taylor_1', 'taylor_2'),
       col=c('black', 'red', 'blue'), lty=1, cex=1)
```

Listing 1: Przykład rozwiniecia funkcji w szereg Taylora



Rysunek 1.3: Przykład: rozwinięcie funkcji w szereg Taylora

1.2.3 Warunki Drugiego Rzędu $f : \mathbb{R} \to \mathbb{R}$

Wskazaliśmy wcześniej, że Warunki Pierwszego Rzędu stanowią warunki konieczne (tj. każde ekstremum będzie posiadać taką własność), ale niewystarczające (tj. są punkty, które ekstremami nie są, a również posiadają taką własność). W celu weryfikacja czy dany punkt stacjonarny jest ekstremum, należy wykorzystać warunki *dostateczne* – Warunki Drugiego Rzędu.

Twierdzenie 4: Warunki Drugiego Rzędu $f : \mathbb{R} \to \mathbb{R}$.
Niech $f : \mathbb{D} \subset \mathbb{R} \to \mathbb{R}$, $f \in C^n$. Jeżeli dla pewnego $x \in \mathbb{D}$ zachodzi: $f'(x) = 0, f''(x) = 0, \ldots, f^{(n-1)}(x) = 0$, ale $f^{(n)}(x) \neq 0$, to:

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

Dowód. 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 $f'(x) = 0, f''(x) = 0, \ldots, f^{(n-1)}(x) = 0$ 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. Analogiczne możemy wykazać przypadek maksimum. □

1.3 Optymalizacja nieliniowa bez ograniczeń, przypadek *wielowymiarowy*, tj. $f : \mathbb{R}^n \to \mathbb{R}$

We wcześniejszej części wykazaliśmy warunki optymalizacji dla przypadku jednowymiarowego – tj. sytuacji, w które mamy do czynienia z jedną zmienną decyzyjną. Natura problemów optymalizacyjnych jest z reguły jednak wielowymiarowa. Poniżej przedstawiamy analogiczny wywód w zakresie problemów wielowymiarowych.

Na początek jednak musimy wykazać kilka podstawowych definicji obszaru rachunku różniczkowego funkcji wielu zmiennych; polecanym w tym zakresie podręcznikiem jest [BI86] – w szczególności początkowe rozdziały.

Definicja 4: Pochodna kierunkowa

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

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

Definicja 5: Pochodna cząstkowa

Niech $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{D}$ oraz $h \in \mathbb{R}^n : x + h \in \mathbb{D}$. Pochodną cząstkową funkcji f w punkcie x względem zmiennej x_i, $i = 1, 2, ...n$ 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 wersowa, tj. przy $h = e_i$

Definicja 6: Gradient funkcji

Niech $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{D}$. Gradientem funkcji f, jest funkcja $\nabla_f(x) : \mathbb{R}^n \to \mathbb{R}^n$ w punkcie x nazywamy funkcję:

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

Związek między Pochodną Kierunkową a Gradientem? Jeżeli istnieje gradient funkcji $\nabla_f(\mathbf{x})$ w punkcie $\mathbf{x}$ (co oznacza, że f jest różniczkowalna w $\mathbf{x}$)

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

to pochodna kierunkowa funkcji f w kierunku wektora $\mathbf{h}$ jest równa iloczynowi skalarnemu gradientu funkcji f i wektora $\mathbf{h}$

$$\frac{df}{dh} = \nabla_f(\mathbf{x}|\mathbf{h}) = \nabla_f(\mathbf{x}) \times \mathbf{h}$$

,

1.3.1 Warunki Pierwszego Rzędu $f : \mathbb{R}^n \to \mathbb{R}$

Twierdzenie 5: Warunki Pierwszego Rzędu $f : \mathbb{R}^n \to \mathbb{R}$.
Niech $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$. Jeżeli funkcja f posiada ekstremum w punkcie x, to $\nabla_f(x) = \mathbf{0}$

Dowód. Rozważmy funkcję $g_x(t) = f(x+th)$ oraz $h \in \mathbb{R}^n : x+h \in \mathbb{D}$. Ponieważ f ma ekstremum w x, to g ma ekstremum w $t = 0$, a więc $g'(t) = 0$ dla $t = 0$, co oznacza $g'(t)|_{t=0} = 0$. W konsekwencji:

$$g'(t)|_{t=0} = \lim_{\Delta \to 0} \frac{g(t+\Delta) - g(t)}{\Delta} = \lim_{\Delta \to 0} \frac{f(x+\Delta h) - f(x)}{\Delta} = \frac{df}{dh}(x) = \nabla_f(x)h = \mathbf{0}$$

□

Przykład 3. Przykład. Niech $f(x) = x_1^2 + x_2^2$. Znajdź ekstrema funkcji $f(x)$.

$$\nabla_f(x) = \left[\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x)\right] = [2x_1, 2x_2] = \mathbf{0} \implies [x_1, x_2] = [0, 0]$$

1.3.2 Warunki Drugiego Rzędu $f : \mathbb{R}^n \to \mathbb{R}$

Do wyprowadzenia warunków drugiego rzędu potrzebujemy wskazać na uogólnienie pochodnej drugiego rzędu dla przypadku funkcji wielu zmiennych: tzw. macierz Hessego.

Definicja 7: Macierz Hessego

Niech $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{D}$. Macierzą Hessego $H_f(x)$ nazywamy macierz:

$$H_f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Uwaga! Macierz Hessego, jest macierzą symetryczną wtedy, gdy wszystkie pochodne cząstkowe drugiego stopnia są ciągłe (tzw. twierdzenie Schwarza). Oznacz to, że: $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$

Dodatkowo, wcześniej wyprowadziliśmy wzór Taylora dla przypadku jednowymiarowego; poniżej jest wzór Taylora w przypadku wielowymiarowym.

Twierdzenie 6: Wzór Taylora $f : \mathbb{R}^n \to \mathbb{R}$.
Niech $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$ *oraz* $f \in C^2$ *w każdym punkcie odcinka* $[x, x+h]$. *Wówczas dla mamy:*

$$f(x+h) = f(x) + \nabla_f(x)h + \frac{1}{2}h^T H_f(x)h + R_3(x, h)$$

Twierdzenie 7: Warunki Drugiego Rzędu $f : \mathbb{R}^n \to \mathbb{R}$.
Niech $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$. *Jeżeli w* x^* *zachodzi*

1. $\nabla_f(x^*) = \mathbf{0}$, *oraz*
2. $H_f(x^*) > 0$

Wtedy w x^* *jest minimum lokalne* f.

Dowód. Z twierdzenia Taylora dla $\mathbb{R}^n$ mamy:

$$f(x+h) = f(x) + \nabla_f(x)h + \frac{1}{2}h^T H_f(x)h + o(|h|^2) = f(x) + \frac{1}{2}h^T H_f(x)h + o(|h|^2)$$

czyli $f(x+h)-f(x) = \frac{1}{2}h^T H_f(x)h+o(|h|^2)$. Z twierdzenia Rayleigh'a, wartość formy kwadratowej $h^T H_f(x)h$ możemy ograniczyć z dołu przez $\lambda_{min}|h|^2$:

$$f(x+h) - f(x) = \frac{1}{2}h^T H_f(x)h + o(|h|^2) \geqslant \frac{1}{2}\lambda_{min}|h|^2 + o(|h|^2)$$

Dla dostatecznie małych h mamy więc $f(x+h) - f(x) > 0$ $\square$

To co nam jeszcze pozostaje, to sprawdzić kiedy $H_f(x^*) > 0$?

Twierdzenie 8: Kryterium Sylwestera.
Niech A będzie symetryczną macierzą o wartościach rzeczywistych:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{bmatrix}$$

Minorami głównymi macierzy nazwiemy:

$$M_1 = a_{1,1} \quad M_2 = det\left(\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}\right) \quad \dots \quad M_n = det\left(\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{bmatrix}\right)$$

Wówczas:

1. *Macierz jest określona dodatnio wtedy i tylko wtedy, gdy **wszystkie** minory główne M_i macierzy A są dodatnie.*
2. *Macierz jest określona ujemnie wtedy i tylko wtedy, gdy wszystkie **parzyste** minory główne M_i macierzy A są dodatnie, a wszystkie **nieparzyste** minory główne M_i są ujemne.*

Przykład 4. Niech $f(x) = x_1^2 + x_2^2$. Znajdź ekstrema funkcji $f(x)$:

Twierdzenie Warunków Pierwszego Rzędu:

$$\nabla_f(x) = \left[\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x)\right] = [2x_1, 2x_2] = \mathbf{0} \implies [x_1, x_2] = [0, 0]$$

Twierdzenie Warunków Drugiego Rzędu:

$$H_f([0,0]) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} > 0$$

The $f(x)$ function has only one extremum, which is $[0, 0]$ – minimum.

Bibliografia

[KW19] M. Kochenderfer and T. Wheeler, "Algorithms for optimization, "*Mit Press*, 2019.

[BI86] A. Birkholc, "Analiza matematyczna: funkcje wielu zmiennych, "*Państwowe Wydawnictwo Naukowe*, 1986.

[CZ04] E.K. Chong and S.H.. Zak, "An introduction to optimization, "*John Wiley & Sons*, 2004.

[CO14] P. Cortez, "Modern optimization with R, "*Springer*, 2014.

[SS08] K. Sydsæter and P. Hammond and A. Seierstadand A. Strom, "Further mathematics for economic analysis "*Pearson education*, 2008.