Optymalizacja ciągła

2. Podstawy matematyczne

Wojciech Kotłowski

Instytut Informatyki PP
http://www.cs.put.poznan.pl/wkotlowski/

7.03.2019

Plan wykładu

1. Wektory i macierze
2. Pochodne cząstkowe, gradient i hesjan

Wektory i macierze

Przestrzeń wektorowa $\mathbb{R}^n$

Przestrzeń $\mathbb{R}^n$ definiujemy jako n-krotny iloczyn kartezjański:

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n}$$

Elementy tej przestrzeni $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ nazywamy **wektorami**.

Przestrzeń wektorowa $\mathbb{R}^n$

Przestrzeń $\mathbb{R}^n$ definiujemy jako n-krotny iloczyn kartezjański:

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n}$$

Elementy tej przestrzeni $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ nazywamy **wektorami**.

Działania na wektorach:

- Dodawanie: dla $\boldsymbol{x} = (x_1, \ldots, x_n)$ i $\boldsymbol{y} = (y_1, \ldots, y_n)$

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, \ldots, x_n + y_n)$$

- Mnożenie przez skalar: dla $\boldsymbol{x} = (x_1, \ldots, x_n)$ i $\alpha \in \mathbb{R}$

$$\alpha \boldsymbol{x} = (\alpha x_1, \ldots, \alpha x_n)$$

Przestrzeń wektorowa $\mathbb{R}^n$

Przestrzeń $\mathbb{R}^n$ definiujemy jako n-krotny iloczyn kartezjański:

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n}$$

Elementy tej przestrzeni $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ nazywamy **wektorami**.

Działania na wektorach:

- Dodawanie: dla $\boldsymbol{x} = (x_1, \ldots, x_n)$ i $\boldsymbol{y} = (y_1, \ldots, y_n)$

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, \ldots, x_n + y_n)$$

- Mnożenie przez skalar: dla $\boldsymbol{x} = (x_1, \ldots, x_n)$ i $\alpha \in \mathbb{R}$

$$\alpha\boldsymbol{x} = (\alpha x_1, \ldots, \alpha x_n)$$

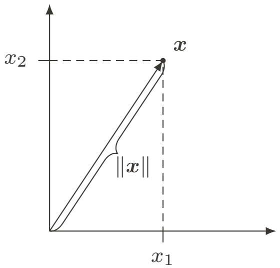
Specjalne oznaczenia:

$$\mathbf{0} = (0, 0, \ldots, 0), \qquad \boldsymbol{e}_k = (0, \ldots, 0, \underbrace{1}_{k\text{-ta pozycja}}, 0, \ldots, 0)$$

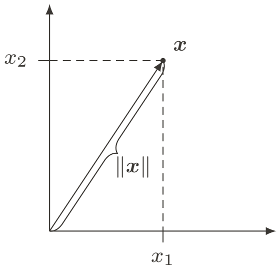
Norma wektora

Norma (długość) wektora:

$$\|\boldsymbol{x}\| \;=\; \sqrt{x_1^2 + \ldots + x_n^2}$$



Norma wektora



Norma (długość) wektora:

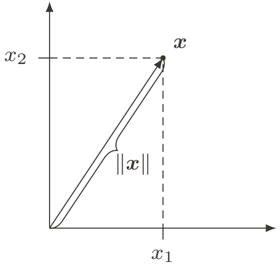
$$\|\boldsymbol{x}\| \;=\; \sqrt{x_1^2 + \ldots + x_n^2}$$

Fakt: Zachodzi $\|\alpha \boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$ dla dowolnego $\alpha \in \mathbb{R}$

Norma wektora

Norma (długość) wektora:

$$\|\boldsymbol{x}\| = \sqrt{x_1^2 + \ldots + x_n^2}$$



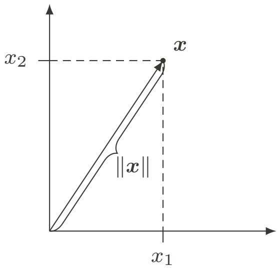
Fakt: Zachodzi $\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|$ dla dowolnego $\alpha \in \mathbb{R}$

Wektor $\boldsymbol{v}$ taki, że $\|\boldsymbol{v}\| = 1$, nazywamy **wektorem jednostkowym**

Norma wektora

Norma (długość) wektora:

$$\|\boldsymbol{x}\| \; = \; \sqrt{x_1^2 + \ldots + x_n^2}$$



Fakt: Zachodzi $\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|$ dla dowolnego $\alpha \in \mathbb{R}$

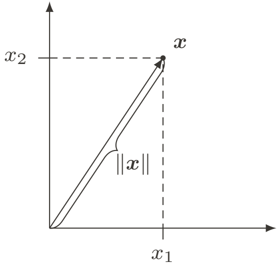
Wektor $\boldsymbol{v}$ taki, że $\|\boldsymbol{v}\| = 1$, nazywamy **wektorem jednostkowym**

Fakt: Wektor $\boldsymbol{v} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ jest jednostkowy

Norma wektora

Norma (długość) wektora:

$$\|\boldsymbol{x}\| \; = \; \sqrt{x_1^2 + \ldots + x_n^2}$$



Fakt: Zachodzi $\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$ dla dowolnego $\alpha \in \mathbb{R}$

Wektor $\boldsymbol{v}$ taki, że $\|\boldsymbol{v}\| = 1$, nazywamy **wektorem jednostkowym**

Fakt: Wektor $\boldsymbol{v} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ jest jednostkowy
Dowód:

$$\|\boldsymbol{v}\| \; = \; \left\| \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \right\| \; = \; \frac{1}{\|\boldsymbol{x}\|}\|\boldsymbol{x}\| \; = \; 1$$

Dygresja: norma ℓ_p

Normę wektora można uogólnić na tzw. normę ℓ_p, dla $p \in [1, \infty]$

$$\|\boldsymbol{x}\|_p = \sqrt[p]{|x_1|^p + \ldots + |x_n|^p}$$

- $\|\boldsymbol{x}\|_1 = |x_1| + \ldots + |x_n|$
- $\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + \ldots + x_n^2} \equiv \|\boldsymbol{x}\|$
- $\|\boldsymbol{x}\|_\infty = \max_{i=1,\ldots,n} |x_i|$

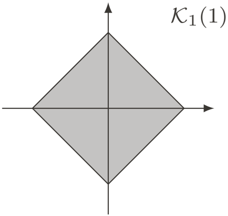
Dygresja: norma ℓ_p

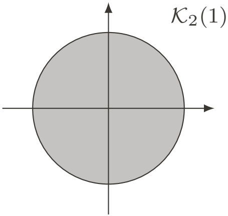
Normę wektora można uogólnić na tzw. normę ℓ_p, dla $p \in [1, \infty]$

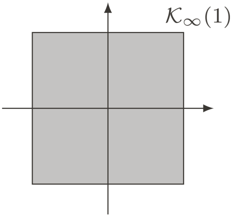
$$\|\boldsymbol{x}\|_p = \sqrt[p]{|x_1|^p + \ldots + |x_n|^p}$$

- $\|\boldsymbol{x}\|_1 = |x_1| + \ldots + |x_n|$
- $\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + \ldots + x_n^2} \equiv \|\boldsymbol{x}\|$
- $\|\boldsymbol{x}\|_\infty = \max_{i=1,\ldots,n} |x_i|$

Kula względem p-normy: $\mathcal{K}_p(r) = \{\boldsymbol{x} \colon |x_1|^p + \ldots + |x_n|^p \leqslant r^p\}$

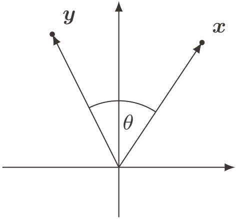






Iloczyn skalarny

Iloczyn skalarny między wektorami:

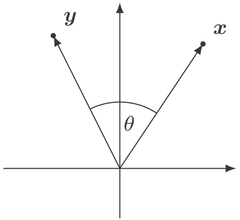


$$\boldsymbol{x} \cdot \boldsymbol{y} \;=\; \boldsymbol{y} \cdot \boldsymbol{x} \;=\; x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

W szczególności: $\boldsymbol{x} \cdot \boldsymbol{x} = \|\boldsymbol{x}\|^2$

Iloczyn skalarny

Iloczyn skalarny między wektorami:



$$\boldsymbol{x} \cdot \boldsymbol{y} \;=\; \boldsymbol{y} \cdot \boldsymbol{x} \;=\; x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

W szczególności: $\boldsymbol{x} \cdot \boldsymbol{x} = \|\boldsymbol{x}\|^2$

Zachodzi:

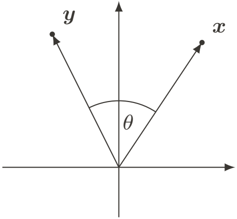
$$\cos\theta = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \cdot \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}$$

Iloczyn skalarny

Iloczyn skalarny między wektorami:

$$\boldsymbol{x} \cdot \boldsymbol{y} \;=\; \boldsymbol{y} \cdot \boldsymbol{x} \;=\; x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

W szczególności: $\boldsymbol{x} \cdot \boldsymbol{x} = \|\boldsymbol{x}\|^2$



Zachodzi:

$$\cos\theta = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \cdot \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}$$

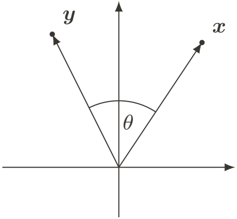
Wektory dla których $\boldsymbol{x} \cdot \boldsymbol{y} = 0$ nazywamy **ortogonalnymi** (prostopadłymi)

Iloczyn skalarny

Iloczyn skalarny między wektorami:

$$\boldsymbol{x} \cdot \boldsymbol{y} \; = \; \boldsymbol{y} \cdot \boldsymbol{x} \; = \; x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

W szczególności: $\boldsymbol{x} \cdot \boldsymbol{x} = \|\boldsymbol{x}\|^2$



Zachodzi:

$$\cos\theta = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \cdot \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}$$

Wektory dla których $\boldsymbol{x} \cdot \boldsymbol{y} = 0$ nazywamy **ortogonalnymi** (prostopadłymi)

Nierówność Cauchy'ego-Schwarza:

$$|\boldsymbol{x} \cdot \boldsymbol{y}| \leqslant \|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

(interpretacja: $|\cos\theta| \leqslant 1$)

Iloczyn skalarny

Zadanie: dla zadanego wektora $\boldsymbol{x}$, znajdź wektor jednostkowy $\boldsymbol{v}$, który (i) maksymalizuje (ii) minimalizuje iloczyn skalarny $\boldsymbol{x} \cdot \boldsymbol{v}$

Iloczyn skalarny

Zadanie: dla zadanego wektora $\boldsymbol{x}$, znajdź wektor jednostkowy $\boldsymbol{v}$, który (i) maksymalizuje (ii) minimalizuje iloczyn skalarny $\boldsymbol{x} \cdot \boldsymbol{v}$

Odpowiedź: z nierówności Cauchy'ego-Schwarza:

$$-\|\boldsymbol{x}\| \underbrace{\|\boldsymbol{v}\|}_{=1} \leqslant \boldsymbol{x} \cdot \boldsymbol{v} \leqslant \|\boldsymbol{x}\| \underbrace{\|\boldsymbol{v}\|}_{=1}$$

Iloczyn skalarny

Zadanie: dla zadanego wektora $\boldsymbol{x}$, znajdź wektor jednostkowy $\boldsymbol{v}$, który (i) maksymalizuje (ii) minimalizuje iloczyn skalarny $\boldsymbol{x} \cdot \boldsymbol{v}$

Odpowiedź: z nierówności Cauchy'ego-Schwarza:

$$-\|\boldsymbol{x}\| \leqslant \boldsymbol{x} \cdot \boldsymbol{v} \leqslant \|\boldsymbol{x}\|$$

Iloczyn skalarny

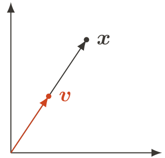
Zadanie: dla zadanego wektora $\boldsymbol{x}$, znajdź wektor jednostkowy $\boldsymbol{v}$, który (i) maksymalizuje (ii) minimalizuje iloczyn skalarny $\boldsymbol{x} \cdot \boldsymbol{v}$

Odpowiedź: z nierówności Cauchy'ego-Schwarza:

$$-\|\boldsymbol{x}\| \leqslant \boldsymbol{x} \cdot \boldsymbol{v} \leqslant \|\boldsymbol{x}\|$$

Biorąc wektor $\boldsymbol{v} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ mamy:

$$\boldsymbol{x} \cdot \boldsymbol{v} = \frac{\boldsymbol{x} \cdot \boldsymbol{x}}{\|\boldsymbol{x}\|} = \|\boldsymbol{x}\|$$



Wektor $\boldsymbol{v} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ **maksymalizuje** $\boldsymbol{x} \cdot \boldsymbol{v}$

Iloczyn skalarny

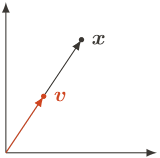
Zadanie: dla zadanego wektora $\boldsymbol{x}$, znajdź wektor jednostkowy $\boldsymbol{v}$, który (i) maksymalizuje (ii) minimalizuje iloczyn skalarny $\boldsymbol{x} \cdot \boldsymbol{v}$

Odpowiedź: z nierówności Cauchy'ego-Schwarza:

$$-\|\boldsymbol{x}\| \leqslant \boldsymbol{x} \cdot \boldsymbol{v} \leqslant \|\boldsymbol{x}\|$$

Biorąc wektor $\boldsymbol{v} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ mamy:

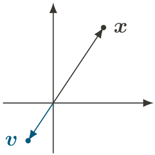
$$\boldsymbol{x} \cdot \boldsymbol{v} = \frac{\boldsymbol{x} \cdot \boldsymbol{x}}{\|\boldsymbol{x}\|} = \|\boldsymbol{x}\|$$



Wektor $\boldsymbol{v} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ **maksymalizuje** $\boldsymbol{x} \cdot \boldsymbol{v}$

Podobnie, biorąc wektor $\boldsymbol{v} = -\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ mamy:

$$\boldsymbol{x} \cdot \boldsymbol{v} = -\|\boldsymbol{x}\|$$



Wektor $\boldsymbol{v} = -\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ **minimalizuje** $\boldsymbol{x} \cdot \boldsymbol{v}$

Macierze

Macierz wymiaru $m \times n$ to prostokątna tablica mn liczb o m wierszach i n kolumnach:

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}$$

Przestrzeń macierzy oznaczamy przez $\mathbb{R}^{m \times n}$

Działania na macierzach

- Mnożenie przez skalar: dla $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ i $\alpha \in \mathbb{R}$

$$\boldsymbol{B} = \alpha \boldsymbol{A} \in \mathbb{R}^{m \times n} \quad \text{ma postać} \quad B_{ij} = \alpha A_{ij}$$

Działania na macierzach

- Mnożenie przez skalar: dla $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ i $\alpha \in \mathbb{R}$

$$\boldsymbol{B} = \alpha \boldsymbol{A} \in \mathbb{R}^{m\times n} \quad \text{ma postać} \quad B_{ij} = \alpha A_{ij}$$

- Dodawanie macierzy: dla $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ i $\boldsymbol{B} \in \mathbb{R}^{m\times n}$

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B} \in \mathbb{R}^{m\times n} \quad \text{ma postać} \quad C_{ij} = A_{ij} + B_{ij}$$

Działania na macierzach

- Mnożenie przez skalar: dla $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ i $\alpha \in \mathbb{R}$

$$\boldsymbol{B} = \alpha \boldsymbol{A} \in \mathbb{R}^{m \times n} \quad \text{ma postać} \quad B_{ij} = \alpha A_{ij}$$

- Dodawanie macierzy: dla $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ i $\boldsymbol{B} \in \mathbb{R}^{m \times n}$

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B} \in \mathbb{R}^{m \times n} \quad \text{ma postać} \quad C_{ij} = A_{ij} + B_{ij}$$

- Mnożenie macierzy: dla $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ i $\boldsymbol{B} \in \mathbb{R}^{n \times p}$

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} \in \mathbb{R}^{m \times p} \quad \text{ma postać} \quad C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

Uwaga: mnożenie macierzy nie jest przemienne!

Działania na macierzach

- Mnożenie przez skalar: dla $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ i $\alpha \in \mathbb{R}$

$$\boldsymbol{B} = \alpha \boldsymbol{A} \in \mathbb{R}^{m\times n} \quad \text{ma postać} \quad B_{ij} = \alpha A_{ij}$$

- Dodawanie macierzy: dla $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ i $\boldsymbol{B} \in \mathbb{R}^{m\times n}$

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B} \in \mathbb{R}^{m\times n} \quad \text{ma postać} \quad C_{ij} = A_{ij} + B_{ij}$$

- Mnożenie macierzy: dla $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ i $\boldsymbol{B} \in \mathbb{R}^{n\times p}$

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} \in \mathbb{R}^{m\times p} \quad \text{ma postać} \quad C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

  Uwaga: mnożenie macierzy nie jest przemienne!

- Transpozycja: dla $\boldsymbol{A} \in \mathbb{R}^{m\times n}$

$$\boldsymbol{B} = \boldsymbol{A}^\top \in \mathbb{R}^{n\times m} \quad \text{ma postać} \quad B_{ij} = A_{ji}$$

  Zachodzi $(\boldsymbol{A} + \boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$ oraz $(\boldsymbol{A}\boldsymbol{B})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$

Macierze kwadratowe

Macierz dla której $n = m$ nazywamy **macierzą kwadratową stopnia** n

Macierze kwadratowe

Macierz dla której $n = m$ nazywamy **macierzą kwadratową stopnia** n

Macierz diagonalna

Niezerowe elementy wyłącznie na **diagonali**

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & 0 & 0 & \cdots & 0 \\ 0 & A_{22} & 0 & \cdots & 0 \\ 0 & 0 & A_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_{nn} \end{bmatrix}$$

Macierze kwadratowe

Macierz dla której $n = m$ nazywamy **macierzą kwadratową stopnia** n

Macierz diagonalna

Niezerowe elementy wyłącznie na **diagonali**

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & 0 & 0 & \cdots & 0 \\ 0 & A_{22} & 0 & \cdots & 0 \\ 0 & 0 & A_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_{nn} \end{bmatrix}$$

Macierz jednostkowa

Jest **elementem neutralnym mnożenia**, tzn. dla każdego $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}$

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Macierze kwadratowe

Macierz dla której $n = m$ nazywamy **macierzą kwadratową stopnia** n

Macierz diagonalna

Niezerowe elementy wyłącznie na **diagonali**

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & 0 & 0 & \cdots & 0 \\ 0 & A_{22} & 0 & \cdots & 0 \\ 0 & 0 & A_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_{nn} \end{bmatrix}$$

Macierz jednostkowa

Jest **elementem neutralnym mnożenia**,
tzn. dla każdego $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}$

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Macierz symetryczna: $\boldsymbol{A}^\top = \boldsymbol{A}$, tzn. $A_{ij} = A_{ji}$.

Wektory jako macierze

Będziemy utożsamiać wektor $\boldsymbol{x} = (x_1, \ldots, x_n)$ z macierzą $n \times 1$ (**wektor kolumnowy**), wtedy $\boldsymbol{x}^\top$ o wymiarze $1 \times n$ jest **wektorem wierszowym**

$$\boldsymbol{x} \;=\; \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad \boldsymbol{x}^\top \;=\; \begin{bmatrix} x_1 \; x_2 \; \cdots \; x_n \end{bmatrix}$$

Wektory jako macierze

Mnożenie macierzy i wektorów: w powyższej reprezentacji dla macierzy kwadratowej $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ i wektorów $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

Wektory jako macierze

Mnożenie macierzy i wektorów: w powyższej reprezentacji dla macierzy kwadratowej $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ i wektorów $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

- $\boldsymbol{A}\boldsymbol{x}$ jest wektorem kolumnowym o wymiarze $n \times 1$

Wektory jako macierze

Mnożenie macierzy i wektorów: w powyższej reprezentacji dla macierzy kwadratowej $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ i wektorów $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

- $\boldsymbol{A}\boldsymbol{x}$ jest wektorem kolumnowym o wymiarze $n \times 1$
- $\boldsymbol{x}^\top \boldsymbol{A}$ jest wektorem wierszowym o wymiarze $1 \times n$

Wektory jako macierze

Mnożenie macierzy i wektorów: w powyższej reprezentacji dla macierzy kwadratowej $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ i wektorów $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

- $\boldsymbol{A}\boldsymbol{x}$ jest wektorem kolumnowym o wymiarze $n \times 1$
- $\boldsymbol{x}^\top \boldsymbol{A}$ jest wektorem wierszowym o wymiarze $1 \times n$
- Mnożenia $\boldsymbol{A}\boldsymbol{x}^\top$ i $\boldsymbol{x}\boldsymbol{A}$ nie mają sensu

Wektory jako macierze

Mnożenie macierzy i wektorów: w powyższej reprezentacji dla macierzy kwadratowej $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ i wektorów $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

- $\boldsymbol{A}\boldsymbol{x}$ jest wektorem kolumnowym o wymiarze $n \times 1$
- $\boldsymbol{x}^\top \boldsymbol{A}$ jest wektorem wierszowym o wymiarze $1 \times n$
- Mnożenia $\boldsymbol{A}\boldsymbol{x}^\top$ i $\boldsymbol{x}\boldsymbol{A}$ nie mają sensu
- $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y}$ jest skalarem (liczbą)

Wektory jako macierze

Mnożenie macierzy i wektorów: w powyższej reprezentacji dla macierzy kwadratowej $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ i wektorów $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

- $\boldsymbol{A}\boldsymbol{x}$ jest wektorem kolumnowym o wymiarze $n \times 1$
- $\boldsymbol{x}^\top \boldsymbol{A}$ jest wektorem wierszowym o wymiarze $1 \times n$
- Mnożenia $\boldsymbol{A}\boldsymbol{x}^\top$ i $\boldsymbol{x}\boldsymbol{A}$ nie mają sensu
- $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y}$ jest skalarem (liczbą)
- $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{A}^\top \boldsymbol{x}$, ponieważ z zasady transpozycji iloczynu $(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})^\top = \boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{A}^\top$ zachodzi:

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y} \;=\; \underbrace{(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y})^\top}_{\text{bo to liczba}} \;=\; \boldsymbol{y}^\top \boldsymbol{A}^\top (\boldsymbol{x}^\top)^\top \;=\; \boldsymbol{y}^\top \boldsymbol{A}^\top \boldsymbol{x}$$

Wektory jako macierze

Mnożenie macierzy i wektorów: w powyższej reprezentacji dla macierzy kwadratowej $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ i wektorów $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

- $\boldsymbol{A}\boldsymbol{x}$ jest wektorem kolumnowym o wymiarze $n \times 1$
- $\boldsymbol{x}^\top \boldsymbol{A}$ jest wektorem wierszowym o wymiarze $1 \times n$
- Mnożenia $\boldsymbol{A}\boldsymbol{x}^\top$ i $\boldsymbol{x}\boldsymbol{A}$ nie mają sensu
- $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y}$ jest skalarem (liczbą)
- $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{A}^\top \boldsymbol{x}$, ponieważ z zasady transpozycji iloczynu $(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})^\top = \boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{A}^\top$ zachodzi:

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y} \; = \; \underbrace{(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y})^\top}_{\text{bo to liczba}} \; = \; \boldsymbol{y}^\top \boldsymbol{A}^\top (\boldsymbol{x}^\top)^\top \; = \; \boldsymbol{y}^\top \boldsymbol{A}^\top \boldsymbol{x}$$

- W szczególności, jeśli $\boldsymbol{A}$ jest symetryczna: $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{x}$

> Dla symetrycznej macierzy $\boldsymbol{A}$ zachodzi
>
> $$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{x}$$

Iloczyn skalarny i iloczyn zewnętrzny

Dla dowolnych $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ zachodzi w reprezentacji macierzowej:

$$\boldsymbol{x}^\top \boldsymbol{y} \;=\; \begin{bmatrix} x_1 \; x_2 \; \cdots \; x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \;=\; x_1 y_1 + \ldots x_n y_n \;=\; \boldsymbol{x} \cdot \boldsymbol{y}$$

Odtąd zapisujemy iloczyn skalarny jako $\boldsymbol{x}^\top \boldsymbol{y}$ (lub $\boldsymbol{y}^\top \boldsymbol{x}$).
W szczególności: $\boldsymbol{x}^\top \boldsymbol{x} = \|\boldsymbol{x}\|^2$

Iloczyn skalarny i iloczyn zewnętrzny

Dla dowolnych $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ zachodzi w reprezentacji macierzowej:

$$\boldsymbol{x}^\top \boldsymbol{y} \;=\; \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \;=\; x_1 y_1 + \ldots x_n y_n \;=\; \boldsymbol{x} \cdot \boldsymbol{y}$$

Odtąd zapisujemy iloczyn skalarny jako $\boldsymbol{x}^\top \boldsymbol{y}$ (lub $\boldsymbol{y}^\top \boldsymbol{x}$).
W szczególności: $\boldsymbol{x}^\top \boldsymbol{x} = \|\boldsymbol{x}\|^2$

Iloczyn $\boldsymbol{x}\boldsymbol{y}^\top \in \mathbb{R}^{n \times n}$ nazywamy **iloczynem zewnętrznym** $\boldsymbol{x}$ i $\boldsymbol{y}$:

$$\boldsymbol{x}\boldsymbol{y}^\top \;=\; \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \;=\; \begin{bmatrix} x_1 y_1 & x_1 y_2 & x_1 y_3 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & x_2 y_3 & \cdots & x_2 y_n \\ x_3 y_1 & x_3 y_2 & x_3 y_3 & \cdots & x_3 y_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & x_n y_3 & \cdots & x_n y_n \end{bmatrix}$$

Wyznacznik macierzy

Wyznacznik macierzy kwadratowej $\boldsymbol{A}$ to funkcja $\det \boldsymbol{A}$ o wartościach rzeczywistych, określona indukcyjnie:

1. Dla $\boldsymbol{A} \in \mathbb{R}^{1\times 1}$, $\det \boldsymbol{A} = A_{11}$
2. Dla $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ z $n \geqslant 2$:

$$\det \boldsymbol{A} \;=\; \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det \boldsymbol{A}_{ij},$$

   gdzie i jest dowolnym wierszem, a $\boldsymbol{A}_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ jest macierzą powstałą z $\boldsymbol{A}$ poprzez skreślenie i-tego wiersza i j-tej kolumny

Wyznacznik macierzy

Wyznacznik macierzy kwadratowej $\boldsymbol{A}$ to funkcja $\det \boldsymbol{A}$ o wartościach rzeczywistych, określona indukcyjnie:

1. Dla $\boldsymbol{A} \in \mathbb{R}^{1\times 1}$, $\det \boldsymbol{A} = A_{11}$
2. Dla $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ z $n \geqslant 2$:

$$\det \boldsymbol{A} \;=\; \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det \boldsymbol{A}_{ij},$$

   gdzie i jest dowolnym wierszem, a $\boldsymbol{A}_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ jest macierzą powstałą z $\boldsymbol{A}$ poprzez skreślenie i-tego wiersza i j-tej kolumny

Przykład dla macierzy $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$:

$$\det \boldsymbol{A} \;=\; \det \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \;=\; A_{11}A_{22} - A_{12}A_{21}$$

Macierz odwrotna

Niech $\boldsymbol{A}$ będzie macierzą kwadratową stopnia n. **Macierz odwrotną** do $\boldsymbol{A}$ oznaczamy przez $\boldsymbol{A}^{-1}$ i definiujemy jako:

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

Macierz odwrotna

Niech $\boldsymbol{A}$ będzie macierzą kwadratową stopnia n. **Macierz odwrotną** do $\boldsymbol{A}$ oznaczamy przez $\boldsymbol{A}^{-1}$ i definiujemy jako:

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

Nie każda macierz ma macierz odwrotną!

Macierz, która ma macierz odwrotną, nazywamy **odwracalną**, a jeśli nie ma, nazywamy ją **osobliwą**

Macierz odwrotna

Niech $\boldsymbol{A}$ będzie macierzą kwadratową stopnia n. **Macierz odwrotną** do $\boldsymbol{A}$ oznaczamy przez $\boldsymbol{A}^{-1}$ i definiujemy jako:

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

Nie każda macierz ma macierz odwrotną!

Macierz, która ma macierz odwrotną, nazywamy **odwracalną**, a jeśli nie ma, nazywamy ją **osobliwą**

Fakt: macierz kwadratowa $\boldsymbol{A}$ jest odwracalna wtedy i tylko wtedy gdy $\det \boldsymbol{A} \neq 0$.

Wyznaczanie macierzy odwrotnej

W ogólności macierz odwrotną można wyznaczyć np. **metodą eliminacji Gaussa** w czasie $O(n^3)$

Rozważymy dwa szczególne przypadki:

Wyznaczanie macierzy odwrotnej

W ogólności macierz odwrotną można wyznaczyć np. **metodą eliminacji Gaussa** w czasie $O(n^3)$

Rozważymy dwa szczególne przypadki:

- Macierz diagonalna:

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{nn} \end{bmatrix}, \qquad \boldsymbol{A}^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 & \cdots & 0 \\ 0 & A_{22}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{nn}^{-1} \end{bmatrix}$$

Wyznaczanie macierzy odwrotnej

W ogólności macierz odwrotną można wyznaczyć np. **metodą eliminacji Gaussa** w czasie $O(n^3)$

Rozważymy dwa szczególne przypadki:

- Macierz diagonalna:

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{nn} \end{bmatrix}, \qquad \boldsymbol{A}^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 & \cdots & 0 \\ 0 & A_{22}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{nn}^{-1} \end{bmatrix}$$

- Macierz 2×2:

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad \boldsymbol{A}^{-1} = \frac{1}{\det \boldsymbol{A}} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}$$

Wartości i wektory własne macierzy symetrycznych

Niech $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ będzie macierzą symetryczną ($\boldsymbol{A}^\top = \boldsymbol{A}$).
Jeśli dla $\lambda \in \mathbb{R}$ i jednostkowego wektora $\boldsymbol{v} \in \mathbb{R}^n$ zachodzi:

$$\boldsymbol{A}\boldsymbol{v} \;=\; \lambda \boldsymbol{v},$$

to λ i $\boldsymbol{v}$ nazywamy, odpowiednio, **wartością własną** i stowarzyszonym z nią **wektorem własnym**

Wartości i wektory własne macierzy symetrycznych

Niech $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ będzie macierzą symetryczną ($\boldsymbol{A}^\top = \boldsymbol{A}$).
Jeśli dla $\lambda \in \mathbb{R}$ i jednostkowego wektora $\boldsymbol{v} \in \mathbb{R}^n$ zachodzi:

$$\boldsymbol{A}\boldsymbol{v} \;=\; \lambda \boldsymbol{v},$$

to λ i $\boldsymbol{v}$ nazywamy, odpowiednio, **wartością własną** i stowarzyszonym z nią **wektorem własnym**

Fakt: Macierz symetryczna stopnia n posiada n wartości własnych $\lambda_1, \ldots, \lambda_n$ (mogą się one powtarzać), a stowarzyszone z nimi wektory własne $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ są **ortogonalne**, tzn. $\boldsymbol{v}_i^\top \boldsymbol{v}_j = 0$ dla $i \neq j$.

Wartości i wektory własne macierzy symetrycznych

Niech $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ będzie macierzą symetryczną ($\boldsymbol{A}^\top = \boldsymbol{A}$).
Jeśli dla $\lambda \in \mathbb{R}$ i jednostkowego wektora $\boldsymbol{v} \in \mathbb{R}^n$ zachodzi:

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v},$$

to λ i $\boldsymbol{v}$ nazywamy, odpowiednio, **wartością własną** i stowarzyszonym z nią **wektorem własnym**

Fakt: Macierz symetryczna stopnia n posiada n wartości własnych $\lambda_1, \ldots, \lambda_n$ (mogą się one powtarzać), a stowarzyszone z nimi wektory własne $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ są **ortogonalne**, tzn. $\boldsymbol{v}_i^\top \boldsymbol{v}_j = 0$ dla $i \neq j$.

Fakt: Jeśli macierz symetryczna $\boldsymbol{A}$ ma wektory własne $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ i wartości własne $\lambda_1, \ldots, \lambda_n$, to jej odwrotność $\boldsymbol{A}^{-1}$ ma te same wektory własne i wartości własne $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$

Wartości i wektory własne macierzy symetrycznych

Niech $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ będzie macierzą symetryczną ($\boldsymbol{A}^\top = \boldsymbol{A}$).
Jeśli dla $\lambda \in \mathbb{R}$ i jednostkowego wektora $\boldsymbol{v} \in \mathbb{R}^n$ zachodzi:

$$\boldsymbol{A}\boldsymbol{v} \;=\; \lambda \boldsymbol{v},$$

to λ i $\boldsymbol{v}$ nazywamy, odpowiednio, **wartością własną** i stowarzyszonym z nią **wektorem własnym**

Fakt: Macierz symetryczna stopnia n posiada n wartości własnych $\lambda_1, \ldots, \lambda_n$ (mogą się one powtarzać), a stowarzyszone z nimi wektory własne $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ są **ortogonalne**, tzn. $\boldsymbol{v}_i^\top \boldsymbol{v}_j = 0$ dla $i \neq j$.

Fakt: Jeśli macierz symetryczna $\boldsymbol{A}$ ma wektory własne $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ i wartości własne $\lambda_1, \ldots, \lambda_n$, to jej odwrotność $\boldsymbol{A}^{-1}$ ma te same wektory własne i wartości własne $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$

Wniosek: Macierz symetryczna jest odwracalna wtedy i tylko wtedy gdy wszystkie jej wartości własne są **niezerowe**

Forma kwadratowa

Dla symetrycznej macierzy $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, wyrażenie postaci:

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

nazywamy **formą kwadratową** z macierzą formy $\boldsymbol{A}$

Forma kwadratowa

Dla symetrycznej macierzy $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, wyrażenie postaci:

$$\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

nazywamy **formą kwadratową** z macierzą formy $\boldsymbol{A}$

Macierz $\boldsymbol{A}$ jest **dodatnio określona** jeśli dla dowolnego niezerowego wektora $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{v} > 0$, i **dodatnio półokreślona**, jeśli $\boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{v} \geqslant 0$.

Forma kwadratowa

Dla symetrycznej macierzy $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, wyrażenie postaci:

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

nazywamy **formą kwadratową** z macierzą formy $\boldsymbol{A}$

Macierz $\boldsymbol{A}$ jest **dodatnio określona** jeśli dla dowolnego niezerowego wektora $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} > 0$, i **dodatnio półokreślona**, jeśli $\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} \geqslant 0$.

Fakt: Jeśli $\lambda_{\min}$ i $\lambda_{\max}$ to najmniejsza i największa wartość własna macierzy, to dla dowolnego wektora $\boldsymbol{v}$:

$$\lambda_{\min} \|\boldsymbol{v}\|^2 \leqslant \boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} \leqslant \lambda_{\max} \|\boldsymbol{v}\|^2$$

Forma kwadratowa

Dla symetrycznej macierzy $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, wyrażenie postaci:

$$\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

nazywamy **formą kwadratową** z macierzą formy $\boldsymbol{A}$

Macierz $\boldsymbol{A}$ jest **dodatnio określona** jeśli dla dowolnego niezerowego wektora $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{v} > 0$, i **dodatnio półokreślona**, jeśli $\boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{v} \geqslant 0$.

Fakt: Jeśli $\lambda_{\min}$ i $\lambda_{\max}$ to najmniejsza i największa wartość własna macierzy, to dla dowolnego wektora $\boldsymbol{v}$:

$$\lambda_{\min} \|\boldsymbol{v}\|^2 \leqslant \boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{v} \leqslant \lambda_{\max} \|\boldsymbol{v}\|^2$$

Fakt: Macierz $\boldsymbol{A}$ jest dodatnio określona (półokreślona), jeśli wszystkie jej wartości własne są dodatnie (nieujemne)

Macierze dodatnio określone

Testowanie czy $\boldsymbol{A}$ jest dodatnio określone jest skomplikowane: wymaga np. wyznaczenia najmniejszej wartości własnej

Macierze dodatnio określone

Testowanie czy $\boldsymbol{A}$ jest dodatnio określone jest skomplikowane: wymaga np. wyznaczenia najmniejszej wartości własnej

Dla macierzy 2×2 jest znacznie prostszy test: Macierz jest dodatnio określona wtedy i tylko wtedy gdy $\det \boldsymbol{A} > 0$ i $A_{11} > 0$

Macierze dodatnio określone

Testowanie czy $\boldsymbol{A}$ jest dodatnio określone jest skomplikowane: wymaga np. wyznaczenia najmniejszej wartości własnej

Dla macierzy 2×2 jest znacznie prostszy test: Macierz jest dodatnio określona wtedy i tylko wtedy gdy $\det \boldsymbol{A} > 0$ i $A_{11} > 0$

Przykład: Rozważmy macierz:

$$\boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}$$

Macierze dodatnio określone

Testowanie czy $\boldsymbol{A}$ jest dodatnio określone jest skomplikowane: wymaga np. wyznaczenia najmniejszej wartości własnej

Dla macierzy 2×2 jest znacznie prostszy test: Macierz jest dodatnio określona wtedy i tylko wtedy gdy $\det \boldsymbol{A} > 0$ i $A_{11} > 0$

Przykład: Rozważmy macierz:

$$\boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \qquad \det \boldsymbol{A} = 5,\ A_{11} = 3$$

Wniosek: macierz $\boldsymbol{A}$ jest dodatnio określona
(faktycznie, wartości własne to $\lambda_1 = 3.62, \lambda_2 = 1.38$)

Macierze dodatnio określone

Testowanie czy $\boldsymbol{A}$ jest dodatnio określone jest skomplikowane: wymaga np. wyznaczenia najmniejszej wartości własnej

Dla macierzy 2×2 jest znacznie prostszy test: Macierz jest dodatnio określona wtedy i tylko wtedy gdy $\det \boldsymbol{A} > 0$ i $A_{11} > 0$

Przykład: Rozważmy macierz:

$$\boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \qquad \det \boldsymbol{A} = 5,\ A_{11} = 3$$

Wniosek: macierz $\boldsymbol{A}$ jest dodatnio określona
(faktycznie, wartości własne to $\lambda_1 = 3.62, \lambda_2 = 1.38$)

Przykład: Rozważmy macierz:

$$\boldsymbol{A} = \begin{bmatrix} 1 & -3 \\ -3 & 4 \end{bmatrix}$$

Macierze dodatnio określone

Testowanie czy $\boldsymbol{A}$ jest dodatnio określone jest skomplikowane: wymaga np. wyznaczenia najmniejszej wartości własnej

Dla macierzy 2×2 jest znacznie prostszy test: Macierz jest dodatnio określona wtedy i tylko wtedy gdy $\det \boldsymbol{A} > 0$ i $A_{11} > 0$

Przykład: Rozważmy macierz:

$$\boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \qquad \det \boldsymbol{A} = 5,\ A_{11} = 3$$

Wniosek: macierz $\boldsymbol{A}$ jest dodatnio określona
(faktycznie, wartości własne to $\lambda_1 = 3.62, \lambda_2 = 1.38$)

Przykład: Rozważmy macierz:

$$\boldsymbol{A} = \begin{bmatrix} 1 & -3 \\ -3 & 4 \end{bmatrix} \qquad \det \boldsymbol{A} = -5,\ A_{11} = 1$$

Wniosek: macierz $\boldsymbol{A}$ nie jest dodatnio określona
(faktycznie, wartości własne to $\lambda_1 = 5.85, \lambda_2 = -0.85$)

Pochodne cząstkowe, gradient i hesjan

Pochodne cząstkowe

Funkcje jednej zmiennej. Pochodna funkcji $f\colon \mathbb{R} \to \mathbb{R}$ w punkcie x:

$$f'(x) \;=\; \frac{\mathrm{d}f(x)}{\mathrm{d}x} \;=\; \lim_{\Delta \to 0} \frac{f(x+\Delta) - f(x)}{\Delta}$$

Pochodne cząstkowe

Funkcje jednej zmiennej. Pochodna funkcji $f\colon \mathbb{R} \to \mathbb{R}$ w punkcie x:

$$f'(x) = \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \lim_{\Delta \to 0} \frac{f(x+\Delta) - f(x)}{\Delta}$$

Funkcje wielu zmiennych. Pochodna cząstkowa funkcji $f\colon \mathbb{R}^n \to \mathbb{R}$ w punkcie $\boldsymbol{x}$ ze względu na zmienną x_i:

$$\begin{aligned} \frac{\partial f(\boldsymbol{x})}{\partial x_i} &= \lim_{\Delta \to 0} \frac{f(x_1, x_2, \ldots, x_i + \Delta, \ldots, x_n) - f(\boldsymbol{x})}{\Delta} \\ &= \lim_{\Delta \to 0} \frac{f(\boldsymbol{x} + \Delta \boldsymbol{e}_i) - f(\boldsymbol{x})}{\Delta} \end{aligned}$$

Wszystkie zmienne poza x_i traktujemy jako stałe

Pochodne cząstkowe

Funkcje jednej zmiennej. Pochodna funkcji $f\colon \mathbb{R} \to \mathbb{R}$ w punkcie x:

$$f'(x) = \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \lim_{\Delta \to 0} \frac{f(x+\Delta) - f(x)}{\Delta}$$

Funkcje wielu zmiennych. Pochodna cząstkowa funkcji $f\colon \mathbb{R}^n \to \mathbb{R}$ w punkcie $\boldsymbol{x}$ ze względu na zmienną x_i:

$$\begin{aligned}\frac{\partial f(\boldsymbol{x})}{\partial x_i} &= \lim_{\Delta \to 0} \frac{f(x_1, x_2, \ldots, x_i + \Delta, \ldots, x_n) - f(\boldsymbol{x})}{\Delta} \\ &= \lim_{\Delta \to 0} \frac{f(\boldsymbol{x} + \Delta \boldsymbol{e}_i) - f(\boldsymbol{x})}{\Delta}\end{aligned}$$

Wszystkie zmienne poza x_i traktujemy jako stałe

Przykład: dla $f(x_1, x_2) = x_1^2 e^{x_2} + x_2$

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 2x_1 e^{x_2}, \qquad \frac{\partial f(x_1, x_2)}{\partial x_2} = x_1^2 e^{x_2} + 1$$

Gradient

Gradient funkcji $f(\boldsymbol{x})$ w punkcie $\boldsymbol{x}$ to wektor pochodnych cząstkowych:

$$\nabla f(\boldsymbol{x}) \;=\; \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \ldots, \frac{\partial f(\boldsymbol{x})}{\partial x_n}\right)$$

Gradient – przykłady

Zadanie: policz gradient funkcji liniowej

$$f(\boldsymbol{x}) \;=\; c + \boldsymbol{b}^\top \boldsymbol{x} \;=\; c + \sum_{i=1}^{n} b_i x_i$$

Gradient – przykłady

Zadanie: policz gradient funkcji liniowej

$$f(\boldsymbol{x}) \;=\; c + \boldsymbol{b}^\top \boldsymbol{x} \;=\; c + \sum_{i=1}^{n} b_i x_i$$

Odpowiedź: wykorzystując liniowość pochodnej:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_k} = \frac{\partial}{\partial x_k}\left(c + \sum_{i=1}^{n} b_i x_i\right) = \underbrace{\frac{\partial c}{\partial x_k}}_{0} + \sum_{i=1}^{n} b_i \underbrace{\frac{\partial x_i}{\partial x_k}}_{\delta_{ik}} = b_k,$$

gdzie δ_{ik} jest tzw. **deltą Kroneckera**: $\delta_{ik} = \begin{cases} 0 & i \neq k \\ 1 & i = k \end{cases}$

Gradient – przykłady

Zadanie: policz gradient funkcji liniowej

$$f(\boldsymbol{x}) \;=\; c + \boldsymbol{b}^\top \boldsymbol{x} \;=\; c + \sum_{i=1}^{n} b_i x_i$$

Odpowiedź: wykorzystując liniowość pochodnej:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_k} = \frac{\partial}{\partial x_k}\left(c + \sum_{i=1}^{n} b_i x_i\right) = \underbrace{\frac{\partial c}{\partial x_k}}_{0} + \sum_{i=1}^{n} b_i \underbrace{\frac{\partial x_i}{\partial x_k}}_{\delta_{ik}} = b_k,$$

gdzie δ_{ik} jest tzw. **deltą Kroneckera**: $\delta_{ik} = \begin{cases} 0 & i \neq k \\ 1 & i = k \end{cases}$

Stąd $\nabla f(\boldsymbol{x}) \;=\; \boldsymbol{b}$

Gradient – przykłady

Zadanie: policz gradient funkcji liniowej

$$f(\boldsymbol{x}) \;=\; c + \boldsymbol{b}^\top \boldsymbol{x} \;=\; c + \sum_{i=1}^{n} b_i x_i$$

Odpowiedź: wykorzystując liniowość pochodnej:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_k} = \frac{\partial}{\partial x_k}\left(c + \sum_{i=1}^{n} b_i x_i\right) = \underbrace{\frac{\partial c}{\partial x_k}}_{0} + \sum_{i=1}^{n} b_i \underbrace{\frac{\partial x_i}{\partial x_k}}_{\delta_{ik}} = b_k,$$

gdzie δ_{ik} jest tzw. **deltą Kroneckera**: $\delta_{ik} = \begin{cases} 0 & i \neq k \\ 1 & i = k \end{cases}$

Stąd $\nabla f(\boldsymbol{x}) \;=\; \boldsymbol{b}$

$$\boxed{\nabla(c + \boldsymbol{b}^\top \boldsymbol{x}) \;=\; \boldsymbol{b}}$$

Gradient – przykłady

Zadanie: policz gradient formy kwadratowej z symetryczną macierzą $\boldsymbol{A}$

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j$$

Gradient – przykłady

Zadanie: policz gradient formy kwadratowej z symetryczną macierzą $\boldsymbol{A}$

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j$$

Odpowiedź: wykorzystując liniowość pochodnej:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_k} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\frac{\partial(x_i x_j)}{\partial x_k} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\left(\underbrace{\frac{\partial x_i}{\partial x_k}}_{\delta_{ik}} x_j + x_i \underbrace{\frac{\partial x_j}{\partial x_k}}_{\delta_{jk}}\right)$$

Gradient – przykłady

Zadanie: policz gradient formy kwadratowej z symetryczną macierzą $\boldsymbol{A}$

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j$$

Odpowiedź: wykorzystując liniowość pochodnej:

$$\begin{aligned}
\frac{\partial f(\boldsymbol{x})}{\partial x_k} \;&=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\frac{\partial (x_i x_j)}{\partial x_k} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\Bigg(\underbrace{\frac{\partial x_i}{\partial x_k}}_{\delta_{ik}} x_j + x_i\,\underbrace{\frac{\partial x_j}{\partial x_k}}_{\delta_{jk}}\Bigg) \\
&=\; \sum_{j=1}^{n} A_{kj}x_j + \sum_{i=1}^{n} \underbrace{A_{ik}}_{A_{ki}} x_i \;=\; 2\sum_{j=1}^{n} A_{kj}x_j
\end{aligned}$$

Gradient – przykłady

Zadanie: policz gradient formy kwadratowej z symetryczną macierzą $\boldsymbol{A}$

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} x_i x_j$$

Odpowiedź: wykorzystując liniowość pochodnej:

$$\begin{aligned}
\frac{\partial f(\boldsymbol{x})}{\partial x_k} \;&=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\frac{\partial(x_i x_j)}{\partial x_k} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\Bigg(\underbrace{\frac{\partial x_i}{\partial x_k}}_{\delta_{ik}} x_j + x_i\,\underbrace{\frac{\partial x_j}{\partial x_k}}_{\delta_{jk}}\Bigg) \\
&=\; \sum_{j=1}^{n} A_{kj}x_j + \sum_{i=1}^{n}\underbrace{A_{ik}}_{A_{ki}}x_i \;=\; \underbrace{2\sum_{j=1}^{n} A_{kj}x_j}_{\text{macierz} \,\times\, \text{wektor}}
\end{aligned}$$

Gradient – przykłady

Zadanie: policz gradient formy kwadratowej z symetryczną macierzą $\boldsymbol{A}$

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j$$

Odpowiedź: wykorzystując liniowość pochodnej:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_k} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\frac{\partial (x_i x_j)}{\partial x_k} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\Bigg(\underbrace{\frac{\partial x_i}{\partial x_k}}_{\delta_{ik}} x_j + x_i \underbrace{\frac{\partial x_j}{\partial x_k}}_{\delta_{jk}}\Bigg)$$

$$= \;\sum_{j=1}^{n} A_{kj}x_j + \sum_{i=1}^{n} \underbrace{A_{ik}}_{A_{ki}} x_i \;=\; \underbrace{2\sum_{j=1}^{n} A_{kj}x_j}_{\text{macierz} \times \text{wektor}}$$

Stąd: $\nabla f(\boldsymbol{x}) \;=\; 2\boldsymbol{A}\boldsymbol{x}$

Gradient – przykłady

Zadanie: policz gradient formy kwadratowej z symetryczną macierzą $\boldsymbol{A}$

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \;=\; \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$

Odpowiedź: wykorzystując liniowość pochodnej:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_k} \;=\; \sum_{i=1}^n \sum_{j=1}^n A_{ij} \frac{\partial (x_i x_j)}{\partial x_k} \;=\; \sum_{i=1}^n \sum_{j=1}^n A_{ij} \left(\underbrace{\frac{\partial x_i}{\partial x_k}}_{\delta_{ik}} x_j + x_i \underbrace{\frac{\partial x_j}{\partial x_k}}_{\delta_{jk}} \right)$$

$$= \;\sum_{j=1}^n A_{kj} x_j + \sum_{i=1}^n \underbrace{A_{ik}}_{A_{ki}} x_i \;=\; \underbrace{2 \sum_{j=1}^n A_{kj} x_j}_{\text{macierz} \,\times\, \text{wektor}}$$

Stąd: $\nabla f(\boldsymbol{x}) \;=\; 2\boldsymbol{A}\boldsymbol{x}$

$$\boxed{\nabla \left(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \right) \;=\; 2\boldsymbol{A}\boldsymbol{x}}$$

Reguła łańcuchowa

Złożenie dwóch funkcji jednej zmiennej
Pochodna funkcji złożonej $g(x) = f\big(y(x)\big)$ dla $x \in \mathbb{R}$

$$g'(x) \;=\; f'(y(x)) \cdot y'(x),$$

lub w notacji Leibniza:

$$\frac{\mathrm{d}g(x)}{\mathrm{d}x} \;=\; \frac{\mathrm{d}f(y)}{\mathrm{d}y} \cdot \frac{\mathrm{d}y(x)}{\mathrm{d}x}$$

Reguła łańcuchowa

Złożenie dwóch funkcji jednej zmiennej
Pochodna funkcji złożonej $g(x) = f\big(y(x)\big)$ dla $x \in \mathbb{R}$

$$g'(x) \;=\; f'(y(x)) \cdot y'(x),$$

lub w notacji Leibniza:

$$\frac{\mathrm{d}g(x)}{\mathrm{d}x} \;=\; \frac{\mathrm{d}f(y)}{\mathrm{d}y} \cdot \frac{\mathrm{d}y(x)}{\mathrm{d}x}$$

Złożenie funkcji wielu zmiennej z funkcjami jednej zmiennej
Pochodna funkcji złożonej $g(x) = f\big(\boldsymbol{y}(x)\big)$ dla $x \in \mathbb{R}$,
gdzie $\boldsymbol{y}(x) = \big(y_1(x), \ldots, y_n(x)\big)$

$$\frac{\mathrm{d}g(x)}{\mathrm{d}x} \;=\; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot \frac{\mathrm{d}y_i(x)}{\mathrm{d}x}$$

Reguła łańcuchowa – przykład

Rozważmy funkcję $g(\alpha)$ dla $\alpha \in \mathbb{R}$ postaci:

$$g(\alpha) \;=\; f(\boldsymbol{x} + \alpha \boldsymbol{v})$$

dla ustalonych wektorów $\boldsymbol{x}, \boldsymbol{v} \in \mathbb{R}^n$

Reguła łańcuchowa – przykład

Rozważmy funkcję $g(\alpha)$ dla $\alpha \in \mathbb{R}$ postaci:

$$g(\alpha) \;=\; f(\boldsymbol{x} + \alpha\boldsymbol{v})$$

dla ustalonych wektorów $\boldsymbol{x}, \boldsymbol{v} \in \mathbb{R}^n$

Definiujemy $\boldsymbol{y}(\alpha) = \boldsymbol{x} + \alpha\boldsymbol{v}$, stąd $g(\alpha) = f\big(\boldsymbol{y}(\alpha)\big)$

Reguła łańcuchowa – przykład

Rozważmy funkcję $g(\alpha)$ dla $\alpha \in \mathbb{R}$ postaci:

$$g(\alpha) \;=\; f(\boldsymbol{x} + \alpha \boldsymbol{v})$$

dla ustalonych wektorów $\boldsymbol{x}, \boldsymbol{v} \in \mathbb{R}^n$

Definiujemy $\boldsymbol{y}(\alpha) = \boldsymbol{x} + \alpha \boldsymbol{v}$, stąd $g(\alpha) = f\big(\boldsymbol{y}(\alpha)\big)$

Z reguły łańcuchowej:

$$\frac{\mathrm{d}g(\alpha)}{\mathrm{d}\alpha} \;=\; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot \frac{\mathrm{d}y_i(\alpha)}{\mathrm{d}\alpha}$$

Reguła łańcuchowa – przykład

Rozważmy funkcję $g(\alpha)$ dla $\alpha \in \mathbb{R}$ postaci:

$$g(\alpha) \;=\; f(\boldsymbol{x} + \alpha \boldsymbol{v})$$

dla ustalonych wektorów $\boldsymbol{x}, \boldsymbol{v} \in \mathbb{R}^n$

Definiujemy $\boldsymbol{y}(\alpha) = \boldsymbol{x} + \alpha \boldsymbol{v}$, stąd $g(\alpha) = f\big(\boldsymbol{y}(\alpha)\big)$

Z reguły łańcuchowej:

$$\frac{\mathrm{d}g(\alpha)}{\mathrm{d}\alpha} \;=\; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot \frac{\mathrm{d}\,\overbrace{y_i(\alpha)}^{x_i + \alpha v_i}}{\mathrm{d}\alpha}$$

Reguła łańcuchowa – przykład

Rozważmy funkcję $g(\alpha)$ dla $\alpha \in \mathbb{R}$ postaci:

$$g(\alpha) \;=\; f(\boldsymbol{x} + \alpha\boldsymbol{v})$$

dla ustalonych wektorów $\boldsymbol{x}, \boldsymbol{v} \in \mathbb{R}^n$

Definiujemy $\boldsymbol{y}(\alpha) = \boldsymbol{x} + \alpha\boldsymbol{v}$, stąd $g(\alpha) = f\big(\boldsymbol{y}(\alpha)\big)$

Z reguły łańcuchowej:

$$\begin{aligned} \frac{\mathrm{d}g(\alpha)}{\mathrm{d}\alpha} \;&=\; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot \frac{\mathrm{d}\, \overbrace{y_i(\alpha)}^{x_i + \alpha v_i}}{\mathrm{d}\alpha} \\ &=\; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot v_i \;=\; \nabla f(\boldsymbol{y})^\top \boldsymbol{v} \end{aligned}$$

Reguła łańcuchowa – przykład

Rozważmy funkcję $g(\alpha)$ dla $\alpha \in \mathbb{R}$ postaci:

$$g(\alpha) \;=\; f(\boldsymbol{x} + \alpha \boldsymbol{v})$$

dla ustalonych wektorów $\boldsymbol{x}, \boldsymbol{v} \in \mathbb{R}^n$

Definiujemy $\boldsymbol{y}(\alpha) = \boldsymbol{x} + \alpha \boldsymbol{v}$, stąd $g(\alpha) = f\big(\boldsymbol{y}(\alpha)\big)$

Z reguły łańcuchowej:

$$\begin{aligned} \frac{\mathrm{d}g(\alpha)}{\mathrm{d}\alpha} \;&=\; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot \frac{\mathrm{d}\,\overbrace{y_i(\alpha)}^{x_i+\alpha v_i}}{\mathrm{d}\alpha} \\ &=\; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot v_i \;=\; \nabla f(\boldsymbol{y})^\top \boldsymbol{v} \end{aligned}$$

$$\boxed{\frac{\mathrm{d}f(\boldsymbol{x} + \alpha \boldsymbol{v})}{\mathrm{d}\alpha} \;=\; \nabla f(\boldsymbol{x} + \alpha \boldsymbol{v})^\top \boldsymbol{v}}$$

Pochodna kierunkowa

Pochodną kierunkową funkcji $f(\boldsymbol{x})$ wzdłuż wektora $\boldsymbol{v}$ w punkcie $\boldsymbol{x}$ nazywamy granicę:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{v}} = \lim_{\Delta \to 0} \frac{f(\boldsymbol{x} + \Delta \boldsymbol{v}) - f(\boldsymbol{x})}{\Delta}$$

Uwaga: zauważmy, że $\frac{\partial f(\boldsymbol{x})}{\partial x_i}$ jest pochodną kierunkową wzdłuż wektora $\boldsymbol{e}_i$

Pochodna kierunkowa

Pochodną kierunkową funkcji $f(\boldsymbol{x})$ wzdłuż wektora $\boldsymbol{v}$ w punkcie $\boldsymbol{x}$ nazywamy granicę:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{v}} = \lim_{\Delta \to 0} \frac{f(\boldsymbol{x} + \Delta \boldsymbol{v}) - f(\boldsymbol{x})}{\Delta}$$

Uwaga: zauważmy, że $\frac{\partial f(\boldsymbol{x})}{\partial x_i}$ jest pochodną kierunkową wzdłuż wektora $\boldsymbol{e}_i$

Definiując funkcję $g(\Delta) = f(\boldsymbol{x} + \Delta \boldsymbol{v})$, otrzymujemy:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{v}} = \lim_{\Delta \to 0} \frac{g(\Delta) - g(0)}{\Delta} = g'(0)$$

Pochodna kierunkowa

Pochodną kierunkową funkcji $f(\boldsymbol{x})$ wzdłuż wektora $\boldsymbol{v}$ w punkcie $\boldsymbol{x}$ nazywamy granicę:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{v}} = \lim_{\Delta \to 0} \frac{f(\boldsymbol{x} + \Delta \boldsymbol{v}) - f(\boldsymbol{x})}{\Delta}$$

Uwaga: zauważmy, że $\frac{\partial f(\boldsymbol{x})}{\partial x_i}$ jest pochodną kierunkową wzdłuż wektora $\boldsymbol{e}_i$

Definiując funkcję $g(\Delta) = f(\boldsymbol{x} + \Delta \boldsymbol{v})$, otrzymujemy:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{v}} = \lim_{\Delta \to 0} \frac{g(\Delta) - g(0)}{\Delta} = g'(0)$$

Używając reguły z poprzedniego slajdu (z Δ w roli α):

$$g'(0) = \frac{\mathrm{d} f(\boldsymbol{x} + \Delta \boldsymbol{v})}{\mathrm{d}\Delta}\Big|_{\Delta=0} = \nabla f(\boldsymbol{x} + \Delta \boldsymbol{v})^\top \boldsymbol{v}\Big|_{\Delta=0} = \nabla f(\boldsymbol{x})^\top \boldsymbol{v}$$

Pochodna kierunkowa

Pochodną kierunkową funkcji $f(\boldsymbol{x})$ wzdłuż wektora $\boldsymbol{v}$ w punkcie $\boldsymbol{x}$ nazywamy granicę:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{v}} = \lim_{\Delta \to 0} \frac{f(\boldsymbol{x} + \Delta \boldsymbol{v}) - f(\boldsymbol{x})}{\Delta}$$

Uwaga: zauważmy, że $\frac{\partial f(\boldsymbol{x})}{\partial x_i}$ jest pochodną kierunkową wzdłuż wektora $\boldsymbol{e}_i$

Definiując funkcję $g(\Delta) = f(\boldsymbol{x} + \Delta \boldsymbol{v})$, otrzymujemy:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{v}} = \lim_{\Delta \to 0} \frac{g(\Delta) - g(0)}{\Delta} = g'(0)$$

Używając reguły z poprzedniego slajdu (z Δ w roli α):

$$g'(0) = \frac{\mathrm{d} f(\boldsymbol{x} + \Delta \boldsymbol{v})}{\mathrm{d}\Delta}\Big|_{\Delta=0} = \nabla f(\boldsymbol{x} + \Delta \boldsymbol{v})^\top \boldsymbol{v}\Big|_{\Delta=0} = \nabla f(\boldsymbol{x})^\top \boldsymbol{v}$$

$$\boxed{\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{v}} = \nabla f(\boldsymbol{x})^\top \boldsymbol{v}}$$

Pochodne wyższych rzędów

$$\underbrace{\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i^2} = \frac{\partial}{\partial x_i}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_i}\right)}_{\text{pochodne czyste}}, \qquad \underbrace{\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_j}\right)}_{\text{pochodne mieszane}},$$

Pochodne wyższych rzędów

$$\underbrace{\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i^2} \;=\; \frac{\partial}{\partial x_i}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_i}\right),}_{\text{pochodne czyste}} \qquad \underbrace{\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} \;=\; \frac{\partial}{\partial x_i}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_j}\right),}_{\text{pochodne mieszane}}$$

Przykład: dla $f(x_1, x_2) = x_1^2 e^{x_2} + x_2$

$$\frac{\partial f}{\partial x_1} \;=\; 2x_1 e^{x_2}, \qquad \frac{\partial f}{\partial x_2} \;=\; x_1^2 e^{x_2} + 1$$

$$\frac{\partial^2 f}{\partial x_1^2} = 2e^{x_2}, \quad \frac{\partial^2 f}{\partial x_2 \partial x_1} = 2x_1 e^{x_2}, \quad \frac{\partial^2 f}{\partial x_2^2} = x_1^2 e^{x_2}, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 2x_1 e^{x_2}$$

Pochodne wyższych rzędów

$$\underbrace{\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i^2} = \frac{\partial}{\partial x_i}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_i}\right)}_{\text{pochodne czyste}}, \qquad \underbrace{\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_j}\right)}_{\text{pochodne mieszane}},$$

Przykład: dla $f(x_1, x_2) = x_1^2 e^{x_2} + x_2$

$$\frac{\partial f}{\partial x_1} = 2x_1 e^{x_2}, \qquad \frac{\partial f}{\partial x_2} = x_1^2 e^{x_2} + 1$$

$$\frac{\partial^2 f}{\partial x_1^2} = 2e^{x_2}, \quad \frac{\partial^2 f}{\partial x_2 \partial x_1} = 2x_1 e^{x_2}, \quad \frac{\partial^2 f}{\partial x_2^2} = x_1^2 e^{x_2}, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 2x_1 e^{x_2}$$

Fakt: Jeśli f ma ciągłe drugie pochodne mieszane to

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_j \partial x_i}$$

Hesjan

Hesjan funkcji $f(\boldsymbol{x})$ w punkcie $\boldsymbol{x}$ to macierz drugich pochodnych cząstkowych:

$$\nabla^2 f(\boldsymbol{x}) \;=\; \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial^2 x_n} \end{bmatrix}$$

Hesjan jest macierzą **symetryczną**

Hesjan

Hesjan funkcji $f(\boldsymbol{x})$ w punkcie $\boldsymbol{x}$ to macierz drugich pochodnych cząstkowych:

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial^2 x_n} \end{bmatrix}$$

Hesjan jest macierzą **symetryczną**

Przykład: dla $f(x_1, x_2) = x_1^2 e^{x_2} + x_2$

$$\frac{\partial^2 f}{\partial x_1^2} = 2e^{x_2}, \quad \frac{\partial^2 f}{\partial x_2 \partial x_1} = 2x_1 e^{x_2}, \quad \frac{\partial^2 f}{\partial x_2^2} = x_1^2 e^{x_2}, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 2x_1 e^{x_2}$$

Stąd: $\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 2e^{x_2} & 2x_1 e^{x_2} \\ 2x_1 e^{x_2} & x_1^2 e^{x_2} \end{bmatrix}$

Hesjan – przykład

Wyznacz hesjan funkcji liniowej $f(\boldsymbol{x}) = c + \boldsymbol{b}^\top \boldsymbol{x}$

Hesjan – przykład

Wyznacz hesjan funkcji liniowej $f(\boldsymbol{x}) = c + \boldsymbol{b}^\top \boldsymbol{x}$

Rozwiązanie:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} \ = \ b_i, \qquad \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} \ = \ 0$$

stąd $\nabla^2 f(\boldsymbol{x}) = \mathbf{0}$ jest **macierzą zerową**

Hesjan – przykład

Wyznacz hesjan formy kwadratowej $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ dla symetrycznej macierzy $\boldsymbol{A}$

Hesjan – przykład

Wyznacz hesjan formy kwadratowej $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ dla symetrycznej macierzy $\boldsymbol{A}$

Rozwiązanie: wyznaczyliśmy już poprzednio

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = 2\sum_{j=1}^{n} A_{ij}x_j,$$

stąd:

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = 2A_{ij}$$

Hesjan – przykład

Wyznacz hesjan formy kwadratowej $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ dla symetrycznej macierzy $\boldsymbol{A}$

Rozwiązanie: wyznaczyliśmy już poprzednio

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = 2\sum_{j=1}^{n} A_{ij}x_j,$$

stąd:

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = 2A_{ij}$$

$$\boxed{\nabla^2 \left(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}\right) = 2\boldsymbol{A}}$$

Hesjan – przykład

Wyznacz hesjan formy kwadratowej $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ dla symetrycznej macierzy $\boldsymbol{A}$

Rozwiązanie: wyznaczyliśmy już poprzednio

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = 2\sum_{j=1}^{n} A_{ij}x_j,$$

stąd:

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = 2A_{ij}$$

$$\boxed{\nabla^2 \left(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}\right) = 2\boldsymbol{A}}$$

Wniosek: Hesjan ogólnej funkcji kwadratowej

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c, \qquad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}$$

Hesjan – przykład

Wyznacz drugą pochodną funkcji złożonej $g(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{v})$

Hesjan – przykład

Wyznacz drugą pochodną funkcji złożonej $g(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{v})$

Pokazaliśmy już, że $g'(\alpha) = \nabla f(\boldsymbol{x} + \alpha \boldsymbol{v})^\top \boldsymbol{v}$

Hesjan – przykład

Wyznacz drugą pochodną funkcji złożonej $g(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{v})$

Pokazaliśmy już, że $g'(\alpha) = \nabla f(\boldsymbol{x} + \alpha \boldsymbol{v})^\top \boldsymbol{v}$

Możemy potraktować $g'(\alpha)$ jako funkcję złożoną $h(\boldsymbol{y}(\alpha))$, gdzie

$$\boldsymbol{y}(\alpha) \ = \ \boldsymbol{x} + \alpha \boldsymbol{v}, \qquad h(\boldsymbol{y}) \ = \ \nabla f(\boldsymbol{y})^\top \boldsymbol{v} \ = \ \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot v_i$$

Hesjan – przykład

Wyznacz drugą pochodną funkcji złożonej $g(\alpha) = f(\boldsymbol{x} + \alpha\boldsymbol{v})$

Pokazaliśmy już, że $g'(\alpha) = \nabla f(\boldsymbol{x} + \alpha\boldsymbol{v})^\top \boldsymbol{v}$

Możemy potraktować $g'(\alpha)$ jako funkcję złożoną $h(\boldsymbol{y}(\alpha))$, gdzie

$$\boldsymbol{y}(\alpha) \; = \; \boldsymbol{x} + \alpha\boldsymbol{v}, \qquad h(\boldsymbol{y}) \; = \; \nabla f(\boldsymbol{y})^\top \boldsymbol{v} \; = \; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot v_i$$

Z reguły łańcuchowej:

$$g''(\alpha) \quad = \quad \frac{\mathrm{d}g'(\alpha)}{\mathrm{d}\alpha} \; = \; \sum_{i=1}^{n} \frac{\partial h(\boldsymbol{y})}{\partial y_i} \cdot \frac{\mathrm{d}\overbrace{y_i}^{x_i + \alpha v_i}}{\mathrm{d}\alpha}$$

Hesjan – przykład

Wyznacz drugą pochodną funkcji złożonej $g(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{v})$

Pokazaliśmy już, że $g'(\alpha) = \nabla f(\boldsymbol{x} + \alpha \boldsymbol{v})^\top \boldsymbol{v}$

Możemy potraktować $g'(\alpha)$ jako funkcję złożoną $h(\boldsymbol{y}(\alpha))$, gdzie

$$\boldsymbol{y}(\alpha) \;=\; \boldsymbol{x} + \alpha \boldsymbol{v}, \qquad h(\boldsymbol{y}) \;=\; \nabla f(\boldsymbol{y})^\top \boldsymbol{v} \;=\; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot v_i$$

Z reguły łańcuchowej:

$$g''(\alpha) \;=\; \frac{\mathrm{d} g'(\alpha)}{\mathrm{d}\alpha} \;=\; \sum_{i=1}^{n} \frac{\partial h(\boldsymbol{y})}{\partial y_i} \cdot \frac{\mathrm{d} \overbrace{y_i}^{x_i + \alpha v_i}}{\mathrm{d}\alpha} \;=\; \sum_{i=1}^{n} \frac{\partial}{\partial y_i} \left(\sum_{j=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_j} v_j \right) v_i$$

Hesjan – przykład

Wyznacz drugą pochodną funkcji złożonej $g(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{v})$

Pokazaliśmy już, że $g'(\alpha) = \nabla f(\boldsymbol{x} + \alpha \boldsymbol{v})^\top \boldsymbol{v}$

Możemy potraktować $g'(\alpha)$ jako funkcję złożoną $h(\boldsymbol{y}(\alpha))$, gdzie

$$\boldsymbol{y}(\alpha) \;=\; \boldsymbol{x} + \alpha \boldsymbol{v}, \qquad h(\boldsymbol{y}) \;=\; \nabla f(\boldsymbol{y})^\top \boldsymbol{v} \;=\; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot v_i$$

Z reguły łańcuchowej:

$$\begin{aligned} g''(\alpha) \;&=\; \frac{\mathrm{d} g'(\alpha)}{\mathrm{d}\alpha} \;=\; \sum_{i=1}^{n} \frac{\partial h(\boldsymbol{y})}{\partial y_i} \cdot \frac{\mathrm{d}\overbrace{y_i}^{x_i + \alpha v_i}}{\mathrm{d}\alpha} \;=\; \sum_{i=1}^{n} \frac{\partial}{\partial y_i} \left(\sum_{j=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_j} v_j \right) v_i \\ &=\; \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f(\boldsymbol{y})}{\partial y_i \partial y_j} v_i v_j \end{aligned}$$

Hesjan – przykład

Wyznacz drugą pochodną funkcji złożonej $g(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{v})$

Pokazaliśmy już, że $g'(\alpha) = \nabla f(\boldsymbol{x} + \alpha \boldsymbol{v})^\top \boldsymbol{v}$

Możemy potraktować $g'(\alpha)$ jako funkcję złożoną $h(\boldsymbol{y}(\alpha))$, gdzie

$$\boldsymbol{y}(\alpha) \;=\; \boldsymbol{x} + \alpha \boldsymbol{v}, \qquad h(\boldsymbol{y}) \;=\; \nabla f(\boldsymbol{y})^\top \boldsymbol{v} \;=\; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot v_i$$

Z reguły łańcuchowej:

$$\begin{aligned} g''(\alpha) \;&=\; \frac{\mathrm{d} g'(\alpha)}{\mathrm{d}\alpha} \;=\; \sum_{i=1}^{n} \frac{\partial h(\boldsymbol{y})}{\partial y_i} \cdot \frac{\mathrm{d} \overbrace{y_i}^{x_i + \alpha v_i}}{\mathrm{d}\alpha} \;=\; \sum_{i=1}^{n} \frac{\partial}{\partial y_i} \left(\sum_{j=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_j} v_j \right) v_i \\ &=\; \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f(\boldsymbol{y})}{\partial y_i \partial y_j} v_i v_j \;=\; \boldsymbol{v}^\top \nabla^2 f(\boldsymbol{y}) \boldsymbol{v} \end{aligned}$$

Hesjan – przykład

Wyznacz drugą pochodną funkcji złożonej $g(\alpha) = f(\boldsymbol{x} + \alpha\boldsymbol{v})$

Pokazaliśmy już, że $g'(\alpha) = \nabla f(\boldsymbol{x} + \alpha\boldsymbol{v})^\top \boldsymbol{v}$

Możemy potraktować $g'(\alpha)$ jako funkcję złożoną $h(\boldsymbol{y}(\alpha))$, gdzie

$$\boldsymbol{y}(\alpha) \;=\; \boldsymbol{x} + \alpha\boldsymbol{v}, \qquad h(\boldsymbol{y}) \;=\; \nabla f(\boldsymbol{y})^\top \boldsymbol{v} \;=\; \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \cdot v_i$$

Z reguły łańcuchowej:

$$\begin{aligned} g''(\alpha) \;&=\; \frac{\mathrm{d}g'(\alpha)}{\mathrm{d}\alpha} \;=\; \sum_{i=1}^{n} \frac{\partial h(\boldsymbol{y})}{\partial y_i} \cdot \frac{\mathrm{d}\overbrace{y_i}^{x_i+\alpha v_i}}{\mathrm{d}\alpha} \;=\; \sum_{i=1}^{n} \frac{\partial}{\partial y_i}\left(\sum_{j=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_j} v_j\right) v_i \\ &=\; \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f(\boldsymbol{y})}{\partial y_i \partial y_j} v_i v_j \;=\; \boldsymbol{v}^\top \nabla^2 f(\boldsymbol{y}) \boldsymbol{v} \end{aligned}$$

$$\boxed{\frac{\mathrm{d}^2 f(\boldsymbol{x} + \alpha\boldsymbol{v})}{\mathrm{d}\alpha^2} \;=\; \boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}+\alpha\boldsymbol{v})\boldsymbol{v}}$$

Minimum lokalne

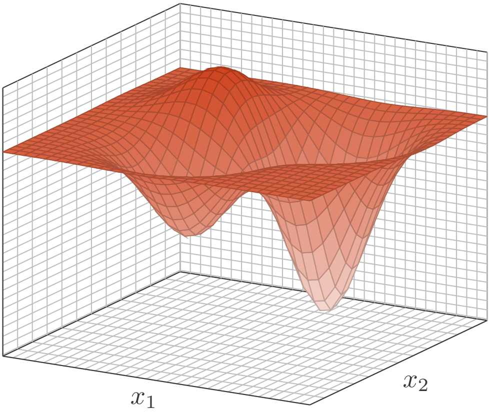
Minimum lokalne: Funkcja $f(\boldsymbol{x})$ ma w punkcie $\boldsymbol{x}_0$ minimum lokalne, jeśli istnieje takie ϵ, że dla dowolnych $\boldsymbol{x}$ spełniających $\|\boldsymbol{x} - \boldsymbol{x}_0\| \leqslant \epsilon$, mamy $f(\boldsymbol{x}_0) \leqslant f(\boldsymbol{x})$.

Analogicznie definiujemy **maksimum lokalne**

Minimum lokalne

Minimum lokalne: Funkcja $f(\boldsymbol{x})$ ma w punkcie $\boldsymbol{x}_0$ minimum lokalne, jeśli istnieje takie ϵ, że dla dowolnych $\boldsymbol{x}$ spełniających $\|\boldsymbol{x} - \boldsymbol{x}_0\| \leqslant \epsilon$, mamy $f(\boldsymbol{x}_0) \leqslant f(\boldsymbol{x})$.

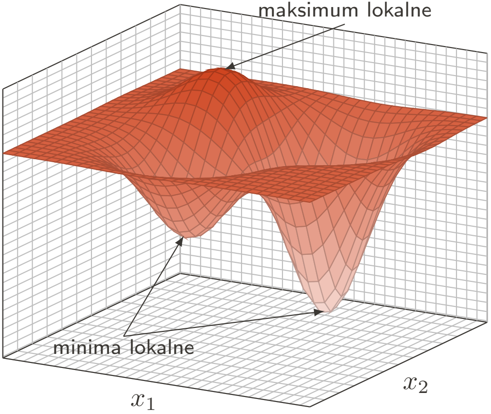
Analogicznie definiujemy **maksimum lokalne**



Minimum lokalne

Minimum lokalne: Funkcja $f(\boldsymbol{x})$ ma w punkcie $\boldsymbol{x}_0$ minimum lokalne, jeśli istnieje takie ϵ, że dla dowolnych $\boldsymbol{x}$ spełniających $\|\boldsymbol{x} - \boldsymbol{x}_0\| \leqslant \epsilon$, mamy $f(\boldsymbol{x}_0) \leqslant f(\boldsymbol{x})$.

Analogicznie definiujemy **maksimum lokalne**



Punkty stacjonarne

Niech $f(\boldsymbol{x})$ będzie różniczkowalna na swojej dziedzinie. Punkty, dla których $\nabla f(\boldsymbol{x}) = \mathbf{0}$ nazywamy **punktami stacjonarnymi** (**krytycznymi**)

Punkty stacjonarne

Niech $f(\boldsymbol{x})$ będzie różniczkowalna na swojej dziedzinie. Punkty, dla których $\nabla f(\boldsymbol{x}) = \mathbf{0}$ nazywamy **punktami stacjonarnymi** (**krytycznymi**)

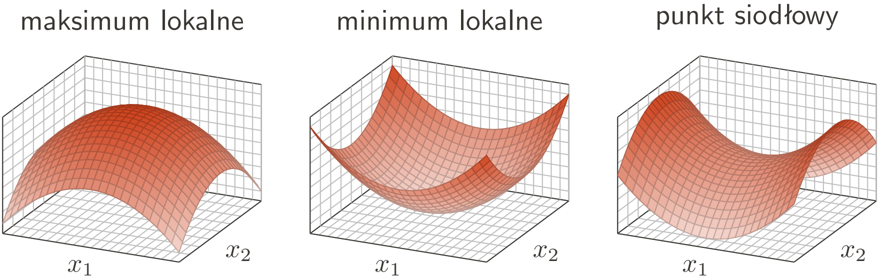
Warunek konieczny na ekstremum: Jeśli różniczkowalna funkcja $f(\boldsymbol{x})$ ma w $\boldsymbol{x}_0$ ekstremum, to $\boldsymbol{x}_0$ jest punktem stacjonarnym.

Punkty stacjonarne

Niech $f(\boldsymbol{x})$ będzie różniczkowalna na swojej dziedzinie. Punkty, dla których $\nabla f(\boldsymbol{x}) = \mathbf{0}$ nazywamy **punktami stacjonarnymi** (**krytycznymi**)

Warunek konieczny na ekstremum: Jeśli różniczkowalna funkcja $f(\boldsymbol{x})$ ma w $\boldsymbol{x}_0$ ekstremum, to $\boldsymbol{x}_0$ jest punktem stacjonarnym.

Uwaga: Nie wszystkie punkty stacjonarne to ekstrema.



Warunek wystarczający na minimum

Jeśli $\boldsymbol{x}_0$ jest punktem stacjonarnym dwukrotnie różniczkowalnej funkcji $f(\boldsymbol{x})$, tzn. $\nabla f(\boldsymbol{x}_0) = \mathbf{0}$, oraz hesjan $\nabla^2 f(\boldsymbol{x}_0)$ jest **dodatnio określony**, to $f(\boldsymbol{x})$ ma **minimum lokalne w** $\boldsymbol{x}_0$

Warunek wystarczający na minimum

Jeśli $\boldsymbol{x}_0$ jest punktem stacjonarnym dwukrotnie różniczkowalnej funkcji $f(\boldsymbol{x})$, tzn. $\nabla f(\boldsymbol{x}_0) = \mathbf{0}$, oraz hesjan $\nabla^2 f(\boldsymbol{x}_0)$ jest **dodatnio określony**, to $f(\boldsymbol{x})$ ma **minimum lokalne w** $\boldsymbol{x}_0$

Przypomnienie: symetryczna macierz $\boldsymbol{A}$ jest dodatnio określona jeśli $\boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{v} > 0$ dla dowolnego niezerowego wektora $\boldsymbol{v}$

Równoważnie, wszystkie wartości własne $\boldsymbol{A}$ są dodatnie

Minimum funkcji kwadratowej

Rozważmy funkcję kwadratową o ogólnej postaci:

$$f(\boldsymbol{x}) \ = \ \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$$

gdzie $\boldsymbol{A}$ jest symetryczna i odwracalna.

Minimum funkcji kwadratowej

Rozważmy funkcję kwadratową o ogólnej postaci:

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$$

gdzie $\boldsymbol{A}$ jest symetryczna i odwracalna.

Wyznaczamy gradient i hesjan:

$$\nabla f(\boldsymbol{x}) \;=\; 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \qquad \nabla^2 f(\boldsymbol{x}) \;=\; 2\boldsymbol{A}$$

Minimum funkcji kwadratowej

Rozważmy funkcję kwadratową o ogólnej postaci:

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$$

gdzie $\boldsymbol{A}$ jest symetryczna i odwracalna.

Wyznaczamy gradient i hesjan:

$$\nabla f(\boldsymbol{x}) \;=\; 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \qquad \nabla^2 f(\boldsymbol{x}) \;=\; 2\boldsymbol{A}$$

Znajdujemy punkt stacjonarny:

$$\nabla f(\boldsymbol{x}) \;=\; \boldsymbol{0} \quad \Longleftrightarrow \quad 2\boldsymbol{A}\boldsymbol{x} \;=\; -\,\boldsymbol{b} \quad \Longleftrightarrow \quad \boldsymbol{x} \;=\; -\,\frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$$

Jest to **minimum**, gdy $\boldsymbol{A}$ jest **dodatnio określona**

Minimum funkcji kwadratowej

Rozważmy funkcję kwadratową o ogólnej postaci:

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$$

gdzie $\boldsymbol{A}$ jest symetryczna i odwracalna.

Wyznaczamy gradient i hesjan:

$$\nabla f(\boldsymbol{x}) \;=\; 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \qquad \nabla^2 f(\boldsymbol{x}) \;=\; 2\boldsymbol{A}$$

Znajdujemy punkt stacjonarny:

$$\nabla f(\boldsymbol{x}) \;=\; \boldsymbol{0} \quad \iff \quad 2\boldsymbol{A}\boldsymbol{x} \;=\; -\,\boldsymbol{b} \quad \iff \quad \boldsymbol{x} \;=\; -\,\frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$$

Jest to **minimum**, gdy $\boldsymbol{A}$ jest **dodatnio określona**

Wniosek: Minimum funkcji kwadratowej:

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c, \qquad \boldsymbol{A} \text{ – dodatnio określona,}$$

ma postać $\boldsymbol{x}^* = -\frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$

Funkcja kwadratowa – przykład

$$f(x_1, x_2) \;=\; 3x_1^2 + 3x_2^2 - 2x_1x_2 + x_1 - 2x_2 + 1$$

Funkcja kwadratowa – przykład

$$f(x_1, x_2) \;=\; 3x_1^2 + 3x_2^2 - 2x_1x_2 + x_1 - 2x_2 + 1$$

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c \quad \text{gdzie} \;\; c = 1, \boldsymbol{b} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$$

Funkcja kwadratowa – przykład

$$f(x_1, x_2) \;=\; 3x_1^2 + 3x_2^2 - 2x_1x_2 + x_1 - 2x_2 + 1$$

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c \quad \text{gdzie} \;\; c = 1, \boldsymbol{b} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$$

$$\boldsymbol{A}^{-1} = \frac{1}{8}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

Funkcja kwadratowa – przykład

$$f(x_1, x_2) \ = \ 3x_1^2 + 3x_2^2 - 2x_1x_2 + x_1 - 2x_2 + 1$$

$$f(\boldsymbol{x}) \ = \ \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c \quad \text{gdzie} \ \ c = 1, \boldsymbol{b} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$$

$$\boldsymbol{A}^{-1} = \frac{1}{8}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \qquad \Longrightarrow \qquad \boldsymbol{x}^* = -\boldsymbol{A}^{-1}\boldsymbol{b} = \begin{bmatrix} -\frac{1}{16} \\ \frac{5}{16} \end{bmatrix}$$

Punkt $\boldsymbol{x}^*$ jest punktem stacjonarnym

Funkcja kwadratowa – przykład

$$f(x_1, x_2) \;=\; 3x_1^2 + 3x_2^2 - 2x_1x_2 + x_1 - 2x_2 + 1$$

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c \quad \text{gdzie} \;\; c = 1, \boldsymbol{b} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$$

$$\boldsymbol{A}^{-1} = \frac{1}{8}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \qquad \Longrightarrow \qquad \boldsymbol{x}^* = -\boldsymbol{A}^{-1}\boldsymbol{b} = \begin{bmatrix} -\frac{1}{16} \\ \frac{5}{16} \end{bmatrix}$$

Punkt $\boldsymbol{x}^*$ jest punktem stacjonarnym

Ponieważ $\det \boldsymbol{A} = 8$ i $A_{11} = 3$, $\boldsymbol{A}$ jest **dodatnio określona**, stąd funkcja ma **minimum w $\boldsymbol{x}^*$**

Wielomiany Taylora dla funkcji wielu zmiennych

Dla danej funkcji $f(\boldsymbol{x})$ i punktu $\boldsymbol{x}_0$, rozważmy funkcję **jednej zmiennej**:

$$g(\alpha) \; = \; f\big(\boldsymbol{x}_0 + \alpha(\boldsymbol{x} - \boldsymbol{x}_0)\big)$$

Zachodzi $g(0) = f(\boldsymbol{x}_0)$ i $g(1) = f(\boldsymbol{x})$.

Wielomiany Taylora dla funkcji wielu zmiennych

Dla danej funkcji $f(\boldsymbol{x})$ i punktu $\boldsymbol{x}_0$, rozważmy funkcję **jednej zmiennej**:

$$g(\alpha) \ = \ f\big(\boldsymbol{x}_0 + \alpha(\boldsymbol{x} - \boldsymbol{x}_0)\big)$$

Zachodzi $g(0) = f(\boldsymbol{x}_0)$ i $g(1) = f(\boldsymbol{x})$.

Konstrukcja wielomianu Taylora dla funkcji $f(x)$ wokół punktu x_0:

- Wyznaczamy wielomian Taylora $P_k(\alpha; 0)$ funkcji $g(\alpha)$ wokół $\alpha = 0$
- Ponieważ $P_k(\alpha; 0)$ przybliża $g(\alpha)$ wokół $\alpha = 0$, oraz $g(0) = f(\boldsymbol{x}_0)$ i $g(1) = f(\boldsymbol{x})$, to $P_k(1; 0)$ przybliża funkcję $f(\boldsymbol{x})$ wokół $\boldsymbol{x} = \boldsymbol{x}_0$

Wielomiany Taylora dla funkcji wielu zmiennych

Dla danej funkcji $f(\boldsymbol{x})$ i punktu $\boldsymbol{x}_0$, rozważmy funkcję **jednej zmiennej**:

$$g(\alpha) \;=\; f\big(\boldsymbol{x}_0 + \alpha(\boldsymbol{x} - \boldsymbol{x}_0)\big)$$

Zachodzi $g(0) = f(\boldsymbol{x}_0)$ i $g(1) = f(\boldsymbol{x})$.

Konstrukcja wielomianu Taylora dla funkcji $f(x)$ wokół punktu x_0:

- Wyznaczamy wielomian Taylora $P_k(\alpha; 0)$ funkcji $g(\alpha)$ wokół $\alpha = 0$
- Ponieważ $P_k(\alpha; 0)$ przybliża $g(\alpha)$ wokół $\alpha = 0$, oraz $g(0) = f(\boldsymbol{x}_0)$ i $g(1) = f(\boldsymbol{x})$, to $P_k(1; 0)$ przybliża funkcję $f(\boldsymbol{x})$ wokół $\boldsymbol{x} = \boldsymbol{x}_0$

Pokazaliśmy już wcześniej (dla $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{x}_0$), że:

$$\begin{aligned}
g'(\alpha) &= \nabla f\big(\boldsymbol{x}_0 + \alpha(\boldsymbol{x} - \boldsymbol{x}_0)\big)^\top (\boldsymbol{x} - \boldsymbol{x}_0) \\
g''(\alpha) &= (\boldsymbol{x} - \boldsymbol{x}_0)^\top \nabla^2 f\big(\boldsymbol{x}_0 + \alpha(\boldsymbol{x} - \boldsymbol{x}_0)\big)(\boldsymbol{x} - \boldsymbol{x}_0)
\end{aligned}$$

Wielomiany Taylora dla funkcji wielu zmiennych

Dla danej funkcji $f(\boldsymbol{x})$ i punktu $\boldsymbol{x}_0$, rozważmy funkcję **jednej zmiennej**:

$$g(\alpha) \;=\; f\big(\boldsymbol{x}_0 + \alpha(\boldsymbol{x} - \boldsymbol{x}_0)\big)$$

Zachodzi $g(0) = f(\boldsymbol{x}_0)$ i $g(1) = f(\boldsymbol{x})$.

Konstrukcja wielomianu Taylora dla funkcji $f(x)$ wokół punktu x_0:

- Wyznaczamy wielomian Taylora $P_k(\alpha; 0)$ funkcji $g(\alpha)$ wokół $\alpha = 0$
- Ponieważ $P_k(\alpha; 0)$ przybliża $g(\alpha)$ wokół $\alpha = 0$, oraz $g(0) = f(\boldsymbol{x}_0)$ i $g(1) = f(\boldsymbol{x})$, to $P_k(1; 0)$ przybliża funkcję $f(\boldsymbol{x})$ wokół $\boldsymbol{x} = \boldsymbol{x}_0$

Pokazaliśmy już wcześniej (dla $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{x}_0$), że:

$$\begin{aligned} g'(0) &= \nabla f(\boldsymbol{x}_0)^\top (\boldsymbol{x} - \boldsymbol{x}_0) \\ g''(0) &= (\boldsymbol{x} - \boldsymbol{x}_0)^\top \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \end{aligned}$$

Wielomiany Taylora dla funkcji wielu zmiennych

Dla danej funkcji $f(\boldsymbol{x})$ i punktu $\boldsymbol{x}_0$, rozważmy funkcję **jednej zmiennej**:

$$g(\alpha) = f(\boldsymbol{x}_0 + \alpha(\boldsymbol{x} - \boldsymbol{x}_0))$$

Zachodzi $g(0) = f(\boldsymbol{x}_0)$ i $g(1) = f(\boldsymbol{x})$.

Konstrukcja wielomianu Taylora dla funkcji $f(x)$ wokół punktu x_0:

- Wyznaczamy wielomian Taylora $P_k(\alpha; 0)$ funkcji $g(\alpha)$ wokół $\alpha = 0$
- Ponieważ $P_k(\alpha; 0)$ przybliża $g(\alpha)$ wokół $\alpha = 0$, oraz $g(0) = f(\boldsymbol{x}_0)$ i $g(1) = f(\boldsymbol{x})$, to $P_k(1; 0)$ przybliża funkcję $f(\boldsymbol{x})$ wokół $\boldsymbol{x} = \boldsymbol{x}_0$

Pokazaliśmy już wcześniej (dla $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{x}_0$), że:

$$\begin{aligned} g'(0) &= \nabla f(\boldsymbol{x}_0)^\top (\boldsymbol{x} - \boldsymbol{x}_0) \\ g''(0) &= (\boldsymbol{x} - \boldsymbol{x}_0)^\top \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \end{aligned}$$

Stąd:

$$\begin{aligned} P_0(\alpha; 0) &= g(0) \\ P_1(\alpha; 0) &= g(0) + g'(0)\alpha \\ P_2(\alpha; 0) &= g(0) + g'(0)\alpha + \frac{g''(0)}{2}\alpha^2 \end{aligned}$$

Wielomiany Taylora dla funkcji wielu zmiennych

Dla danej funkcji $f(\boldsymbol{x})$ i punktu $\boldsymbol{x}_0$, rozważmy funkcję **jednej zmiennej**:

$$g(\alpha) \;=\; f\big(\boldsymbol{x}_0 + \alpha(\boldsymbol{x} - \boldsymbol{x}_0)\big)$$

Zachodzi $g(0) = f(\boldsymbol{x}_0)$ i $g(1) = f(\boldsymbol{x})$.

Konstrukcja wielomianu Taylora dla funkcji $f(x)$ wokół punktu x_0:

- Wyznaczamy wielomian Taylora $P_k(\alpha; 0)$ funkcji $g(\alpha)$ wokół $\alpha = 0$
- Ponieważ $P_k(\alpha; 0)$ przybliża $g(\alpha)$ wokół $\alpha = 0$, oraz $g(0) = f(\boldsymbol{x}_0)$ i $g(1) = f(\boldsymbol{x})$, to $P_k(1; 0)$ przybliża funkcję $f(\boldsymbol{x})$ wokół $\boldsymbol{x} = \boldsymbol{x}_0$

Pokazaliśmy już wcześniej (dla $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{x}_0$), że:

$$\begin{aligned} g'(0) &= \nabla f(\boldsymbol{x}_0)^\top (\boldsymbol{x} - \boldsymbol{x}_0) \\ g''(0) &= (\boldsymbol{x} - \boldsymbol{x}_0)^\top \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \end{aligned}$$

Stąd:

$$\begin{aligned} P_0(\alpha; 0) &= f(\boldsymbol{x}) \\ P_1(\alpha; 0) &= f(\boldsymbol{x}_0) + \alpha \nabla f(\boldsymbol{x}_0)^\top (\boldsymbol{x} - \boldsymbol{x}_0) \\ P_2(\alpha; 0) &= f(\boldsymbol{x}_0) + \alpha \nabla f(\boldsymbol{x}_0)^\top (\boldsymbol{x} - \boldsymbol{x}_0) + \frac{\alpha^2}{2} (\boldsymbol{x} - \boldsymbol{x}_0)^\top \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \end{aligned}$$

Wielomiany Taylora dla funkcji wielu zmiennych

Wielomiany Taylora dla funkcji $f(\boldsymbol{x})$ wokół punktu $\boldsymbol{x}_0$:

$$
\begin{aligned}
P_0(\boldsymbol{x};\boldsymbol{x}_0) &= f(\boldsymbol{x}_0) \\
P_1(\boldsymbol{x};\boldsymbol{x}_0) &= f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^\top(\boldsymbol{x}-\boldsymbol{x}_0) \\
P_2(\boldsymbol{x};\boldsymbol{x}_0) &= f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^\top(\boldsymbol{x}-\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^\top \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0) \\
\ldots &= \ldots
\end{aligned}
$$

Wielomiany Taylora dla funkcji wielu zmiennych

Wielomiany Taylora dla funkcji $f(\boldsymbol{x})$ wokół punktu $\boldsymbol{x}_0$:

$$
\begin{aligned}
P_0(\boldsymbol{x};\boldsymbol{x}_0) &= f(\boldsymbol{x}_0) \\
P_1(\boldsymbol{x};\boldsymbol{x}_0) &= f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^\top(\boldsymbol{x}-\boldsymbol{x}_0) \\
P_2(\boldsymbol{x};\boldsymbol{x}_0) &= f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^\top(\boldsymbol{x}-\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^\top\nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0) \\
\ldots &= \ldots
\end{aligned}
$$

Wzór Taylora (tylko dla drugiego rzędu):

$$
f(\boldsymbol{x}) \;=\; f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^\top(\boldsymbol{x}-\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^\top\nabla^2 f(\boldsymbol{\xi})(\boldsymbol{x}-\boldsymbol{x}_0),
$$

gdzie $\boldsymbol{\xi}$ jest punktem między $\boldsymbol{x}_0$ a $\boldsymbol{x}$

Wielomian Taylora dla funkcji kwadratowej

Rozważmy funkcję kwadratową z dodatnio określoną macierzą $\boldsymbol{A}$

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c, \qquad \nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \quad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}$$

Wielomian Taylora dla funkcji kwadratowej

Rozważmy funkcję kwadratową z dodatnio określoną macierzą $\boldsymbol{A}$

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c, \qquad \nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \quad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}$$

Przybliżenie funkcji kwadratowej wielomianem drugiego stopnia (funkcją kwadratową!) jest **dokładne**, a ze wzoru Taylora dla $\boldsymbol{x}_0 = \boldsymbol{x}^*$:

Wielomian Taylora dla funkcji kwadratowej

Rozważmy funkcję kwadratową z dodatnio określoną macierzą $\boldsymbol{A}$

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c, \qquad \nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \quad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}$$

Przybliżenie funkcji kwadratowej wielomianem drugiego stopnia (funkcją kwadratową!) jest **dokładne**, a ze wzoru Taylora dla $\boldsymbol{x}_0 = \boldsymbol{x}^*$:

$$f(\boldsymbol{x}) \;=\; f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^\top \nabla^2 f(\boldsymbol{\xi})(\boldsymbol{x} - \boldsymbol{x}^*)$$

Wielomian Taylora dla funkcji kwadratowej

Rozważmy funkcję kwadratową z dodatnio określoną macierzą $\boldsymbol{A}$

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c, \qquad \nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \quad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}$$

Przybliżenie funkcji kwadratowej wielomianem drugiego stopnia (funkcją kwadratową!) jest **dokładne**, a ze wzoru Taylora dla $\boldsymbol{x}_0 = \boldsymbol{x}^*$:

$$f(\boldsymbol{x}) \;=\; f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^\top \nabla^2 f(\boldsymbol{\xi})(\boldsymbol{x} - \boldsymbol{x}^*)$$

Ponieważ $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ (bo $\boldsymbol{x}^*$ jest punktem minimum), a $\nabla^2 f(\boldsymbol{\xi}) = 2\boldsymbol{A}$:

Dla funkcji kwadratowej z dodatnio określoną macierzą $\boldsymbol{A}$:

$$f(\boldsymbol{x}) \;=\; f(\boldsymbol{x}^*) + (\boldsymbol{x} - \boldsymbol{x}^*)^\top \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*)$$