

Wykład 2

■ 2.1.2 Rozkład Taylora funkcji wielu zmiennych

Podobnie, jak w przypadku funkcji jednej zmiennej, rozkład Taylora zależy od pochodnych coraz wyższych rzędów. W dalszym wykładzie potrzebne nam będzie jedynie przybliżenie Taylora do rzędu 2.

Definicja. *Gradientem funkcji wielu zmiennych $x_1, x_2, \dots, x_n$ nazywamy wektor:*

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \partial_{x_1} f(\mathbf{x}) \\ \partial_{x_2} f(\mathbf{x}) \\ \dots \\ \partial_{x_n} f(\mathbf{x}) \end{pmatrix},$$

gdzie $\partial_{x_i} f(x)$ jest pochodną cząstkową funkcji f względem zmiennej x_i w punkcie x .

Definicja. *Hesjanem funkcji wielu zmiennych $x_1, x_2, \dots, x_n$ nazywamy macierz:*

$$\nabla^2 f(\mathbf{x}) = (\partial^2_{x_i, x_j} f(\mathbf{x})) ,$$

gdzie $\partial^2_{x_i, x_j} f(x)$ jest pochodną cząstkową funkcji f względem zmiennych x_i oraz x_j w punkcie x , (obliczanych w tym porządku).

Twierdzenie. *Hesjan funkcji klasy C^2 jest symetryczny*

Twierdzenie (rozkład Taylora). *Funkcja klasy C^2 z dokładnością do wyrazów drugiego rzędu ma w otoczeniu punktu $\bar{x}$ rozkład:*

$$f(\mathbf{x}) \approx f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^T (\mathbf{x} - \bar{\mathbf{x}}) + \frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}})^T \nabla^2 f(\bar{\mathbf{x}}) (\mathbf{x} - \bar{\mathbf{x}})$$

Znak $\approx$ należy rozumieć w ten sposób, że dla h dostatecznie małego zachodzi jeden z dwóch wzorów:

$$\begin{aligned} f(\bar{\mathbf{x}} + h) &= f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}} + \theta h)^T h, \\ f(\bar{\mathbf{x}} + h) &= f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^T h + \frac{1}{2} h^T \nabla^2 f(\bar{\mathbf{x}} + \theta h) h \end{aligned}$$

dla pewnego $\theta \in (0,1)$

■ Przykład 2.2 (Rozkład Taylora)

Rozwińmy w szereg Taylora w otoczeniu punktu $\bar{x} = (1, 2, 3)$ funkcję

$$f(x, y, z) = x^3 + 12xy^2 + 2y^2 + 5z^2 + xz^4$$

Łatwo obliczyć:

$$f(\bar{x}) = 183, \quad \nabla f(\bar{x}) = \begin{pmatrix} \partial_x f(x, y, z) \\ \partial_y f(x, y, z) \\ \partial_z f(x, y, z) \end{pmatrix} = \begin{pmatrix} 3x^2 + 12y^2 + z^4 \\ 4y + 24xy \\ 2z(5 + 2xz^2) \end{pmatrix}, \quad \nabla f(\bar{x}) = \begin{pmatrix} 132 \\ 56 \\ 138 \end{pmatrix},$$

$$\nabla^2 f(\bar{x}) = \begin{pmatrix} 6x & 24y & 4z^3 \\ 24y & 4 + 24x & 0 \\ 4z^3 & 0 & 2(5 + 6xz^2) \end{pmatrix}, \quad \nabla^2 f(\bar{x}) = \begin{pmatrix} 6 & 48 & 108 \\ 48 & 28 & 0 \\ 108 & 0 & 118 \end{pmatrix}$$

Po podstawieniu, otrzymamy rozwinięcie w szereg Taylora w otoczeniu punktu $\bar{x} = (1, 2, 3)$:

$$f(x, y, z) \approx 183 + (132, 56, 138) \begin{pmatrix} x-1 \\ y-2 \\ z-3 \end{pmatrix} + (x-1, y-2, z-3) \begin{pmatrix} 6 & 48 & 108 \\ 48 & 28 & 0 \\ 108 & 0 & 118 \end{pmatrix} \begin{pmatrix} x-1 \\ y-2 \\ z-3 \end{pmatrix}$$

co po wykonaniu działań da wielomian drugiego stopnia trzech zmiennych x, y, z :

$$535 - 294x + 3x^2 - 48y + 48xy + 14y^2 - 324z + 108xz + 59z^2$$

Ograniczając się jedynie do wyrazów pierwszego stopnia mamy:

$$-475 + 132x + 56y + 138z$$

Dokładna wartość funkcji $f(x)$ w punkcie $\bar{x} = (1.01, 2.01, 3.01)$ wynosi 186.2832791301. Z przybliżenia liniowego otrzymamy wartość 186.26; przybliżenie wielomianem stopnia 2 ma wartość 186.283200000000. Jak widać, w bliskim otoczeniu punktu $\bar{x}$ dokładność jest dobra i przybliżenie wielomianem stopnia 2 jest istotnie dokładniejsze.

Co innego w dalszej odległości od punktu $\bar{x}$. Np w punkcie (0,0,0) wartość rzeczywista wynosi 0, wartość z przybliżenia liniowego -475 a z przybliżenia kwadratowego 535. ■

■ 2.2 Warunek konieczny istnienia minimum

Niech funkcja $f(x)$ jest różniczkowalna w pewnym obszarze. Przypuśćmy, że w punkcie x^* tego obszaru jest minimum lokalne, to znaczy, że w pewnym otoczeniu $U(x^*)$ punktu x^* zachodzi nierówność $f(x^*) < f(x)$.

Wtedy funkcja

$$g(u) = f(x_1^*, \dots, u, \dots, x_r^*)$$

ma minimum w punkcie $u = x_i^*$. Warunkiem koniecznym istnienia minimum jest

$$0 = \frac{dg(u)}{du} \Big|_{(u=x_i^*)} = \partial_{x_i} f(x^*)$$

Udowodniliśmy więc

Twierdzenie. Warunkiem koniecznym istnienia minimum w punkcie x^* jest zachodzenie równości

$$\nabla f(x^*) = 0$$

Definicja. Punkty spełniające warunek (2.2.3) nazywamy **punktami stacjonarnymi funkcji**.

■ 2.3 Warunek dostateczny istnienia minimum

Niech $f(x)$ ma pierwsze i drugie pochodne cząstkowe ciągłe.

Twierdzenie. Warunkiem dostatecznym istnienia minimum w punkcie x^* jest zachodzenie nierówności

$$d^T (\nabla^2 f(x^*)) d > 0$$

dla każdego d .

Definicja. Macierz A spełniającą dla każdego d warunek

$$d^T A d > 0$$

nazywamy **macierzą dodatnio określoną**

Dowód twierdzenia.

Istnieje otoczenie U punktu 0 takie, że dla h z tego otoczenia zachodzą warunki:

$$\begin{aligned} f(x^* + h) &= f(x^*) + \nabla f(x^*)^T h + \frac{1}{2} h^T \nabla^2 f(x^* + \theta h) h, \\ 0 &< \theta < 1, \quad h \neq 0 \\ d^T (\nabla^2 f(x^*)) d &> 0 \end{aligned}$$

dla dowolnego d .

Z warunku koniecznego istnienia minimum wiemy, że w tym otoczeniu

$$\nabla f(x^*) = 0$$

Z ciągłości drugich pochodnych cząstkowych mamy, że dla dowolnego d

$$d^T \nabla^2 f(x^* + \theta h) d > 0$$

Stąd

$$f(x^* + h) - f(x^*) = \frac{1}{2} h^T \nabla^2 f(x^* + \theta h) h > 0$$

To kończy dowód ■

Wiadomo, że macierz A jest dodatnio określona wtedy i tylko wtedy, gdy wszystkie jej minory główne są dodatnie.

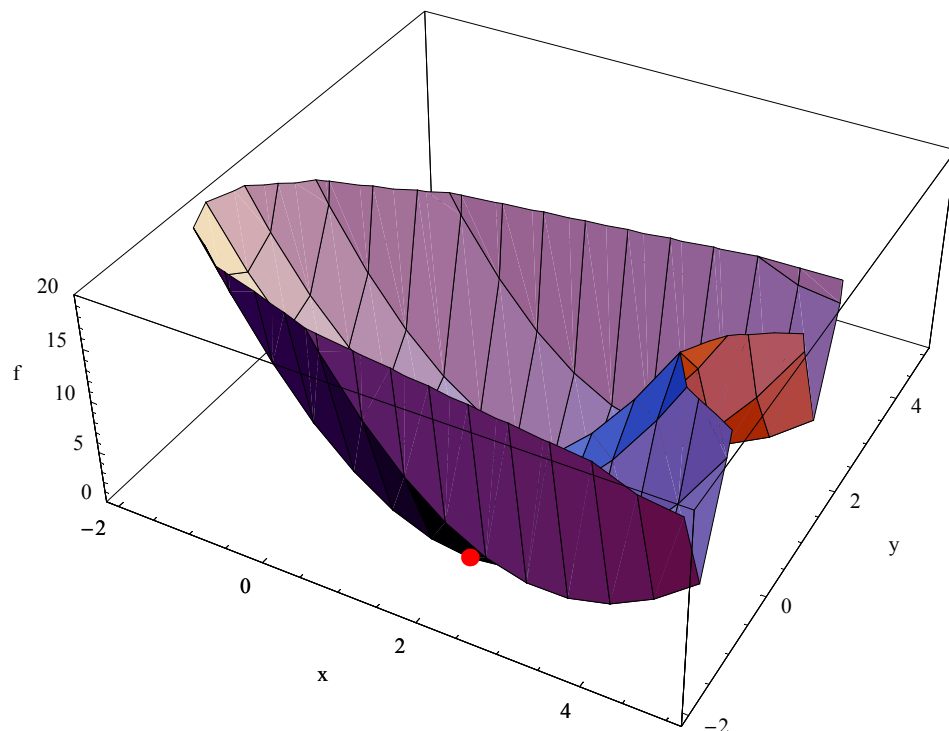
Definicja. Minorem głównym A_k macierzy $A = (a_{ij})$ nazywamy wyznacznik macierzy

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{pmatrix}$$

Uwaga. Aby zbadać, w których punktach funkcja $f(x)$ ma maksimum, wystarczy zbadać funkcję $-f(x)$. Punkty stacjonarne takiej funkcji pokrywają się z punktami stacjonarnymi funkcji $f(x)$. Hesjan funkcji $-f(x)$ powstaje z hesjanu funkcji $f(x)$ przez pomnożenie jej wszystkich elementów przez -1 . Punkty stacjonarne, w których nie ma ani minimum, ani maksimum nazywamy **punktami przegięcia**.

■ Przykład 2.3 (Klasyfikacja punktów stacjonarnych)

Zbadajmy punkty stacjonarne funkcji $f(x, y) = (x - 2)^2 + (x - 2y^2)^2$.



Oto wyniki, uzyskane za pomocą programu *Mathematica*.

```

Gradient vector →  $\begin{pmatrix} -4 + 4x - 4y^2 \\ -8xy + 16y^3 \end{pmatrix}$ 

Hessian matrix →  $\begin{pmatrix} 4 & -8y \\ -8y & -8x + 48y^2 \end{pmatrix}$ 

***** First-order optimality conditions *****

Necessary conditions →  $\begin{pmatrix} -4 + 4x - 4y^2 == 0 \\ -8xy + 16y^3 == 0 \end{pmatrix}$ 

Possible solutions (stationary points) →  $\begin{pmatrix} x \rightarrow 1. & y \rightarrow 0. \\ x \rightarrow 2. & y \rightarrow -1. \\ x \rightarrow 2. & y \rightarrow 1. \end{pmatrix}$ 

***** Second-order optimality conditions *****

----- Point → {x → 1., y → 0.}

Hessian →  $\begin{pmatrix} 4 & 0. \\ 0. & -8. \end{pmatrix}$  Principal minors →  $\begin{pmatrix} 4 \\ -32. \end{pmatrix}$ 

Status → InflectionPoint Function value → 2.

----- Point → {x → 2., y → -1.}

Hessian →  $\begin{pmatrix} 4 & 8. \\ 8. & 32. \end{pmatrix}$  Principal minors →  $\begin{pmatrix} 4 \\ 64. \end{pmatrix}$ 

Status → MinimumPoint Function value → 0.
```

----- Point $\rightarrow \{x \rightarrow 2., y \rightarrow 1.\}$

Hessian $\rightarrow \begin{pmatrix} 4 & -8. \\ -8. & 32. \end{pmatrix}$ Principal minors $\rightarrow \begin{pmatrix} 4 \\ 64. \end{pmatrix}$

Status $\rightarrow$ MinimumPoint Function value $\rightarrow 0.$

