

Rozdział 2

Różniczkowanie funkcji wielu zmiennych

2.1 Pochodne cząstkowe, kierunkowe i różniczka zupełna

Definicja 2.1 (pochodna cząstkowa). Mówimy, że funkcja $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$, gdzie zbiór $\Omega \subset \mathbb{R}^n$ jest otwarty, ma w punkcie $\mathbf{a} = (a_1, \dots, a_n) \in \Omega$ *pochodną cząstkową względem zmiennej x_i* wtedy i tylko wtedy, gdy funkcja *jednej* zmiennej rzeczywistej

$$F_i(t) = f(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_n), \quad F_i: (a_i - \delta, a_i + \delta) \rightarrow \mathbb{R}^m \quad (2.1)$$

ma pochodną w punkcie a_i , gdzie $\delta > 0$ wybieramy tak, aby odcinek $(\mathbf{a} - \delta \mathbf{e}_i, \mathbf{a} + \delta \mathbf{e}_i) \subset \Omega$. Przyjmujemy

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = F'_i(a_i) = \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h \mathbf{e}_i) - f(\mathbf{a})}{h}. \quad (2.2)$$

Używa się także innych oznaczeń:

$$D_i f(\mathbf{a}) = f_{x_i}(\mathbf{a}) = \frac{\partial f}{\partial x_i}(\mathbf{a}).$$

Jak widać, pochodną cząstkową $\frac{\partial f}{\partial x_i}$ obliczamy, traktując wszystkie zmienne oprócz x_i jako ustalone parametry i wykonując różniczkowanie względem x_i . Obowiązują przy tym oczywiście wszystkie reguły, które Czytelnik poznał, ucząc się rachunku różniczkowego funkcji jednej zmiennej rzeczywistej. Np. jeśli

$$f(x, y, z) = x^2 + x \cos(yz) + z \exp(x^2),$$

to

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y, z) &= 2x + \cos(yz) + 2zx \exp(x^2), & \frac{\partial f}{\partial y}(x, y, z) &= -xz \sin(yz), \\ \frac{\partial f}{\partial z}(x, y, z) &= -xy \sin(yz) + \exp(x^2). \end{aligned}$$

Uwaga 2.2. Ze wzoru (2.2) wynika, że gdy wartości f są liczbami rzeczywistymi, tzn. $m = 1$ w Definicji 2.1, to także $\frac{\partial f}{\partial x_i}(\mathbf{a})$ – tam, gdzie jest określona – jest liczbą rzeczywistą. Jeśli $m > 1$, to $\frac{\partial f}{\partial x_i}(\mathbf{a})$ jest wektorem z przestrzeni $\mathbb{R}^m$.

Podkreślmy od razu, że samo istnienie pochodnych cząstkowych funkcji f *nie gwarantuje* ciągłości funkcji.

Przykład 2.3. (i) Niech $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ będzie funkcją charakterystyczną zbioru $A = \{(x, y) \in \mathbb{R}^2 : y = x^2, y \neq 0\}$, tzn. niech $f \equiv 1$ na A i $f \equiv 0$ na $\mathbb{R}^2 \setminus A$. Na obu osiach układu współrzędnych mamy $f \equiv 0$ i dlatego, wprost z definicji,

$$\frac{\partial f}{\partial x}(\mathbf{0}) = 0 = \frac{\partial f}{\partial y}(\mathbf{0}).$$

Jednak f nie jest ciągła w zerze, gdyż $f(\mathbf{0}) = 0 \neq \lim_{j \rightarrow \infty} f(1/j, 1/j^2)$.

(ii) Rozpatrzmy ponownie funkcję z Przykładu 1.49:

$$f(x, y) = \begin{cases} \frac{yx^2}{y^2 + x^4}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases} \quad (2.3)$$

Jeśli $(x, y) \neq (0, 0) \in \mathbb{R}^2$, to

$$\frac{\partial f}{\partial x}(x, y) = \frac{2xy(y^2 + x^4) - 4x^3 \cdot yx^2}{(y^2 + x^4)^2} = \frac{2xy(y^2 - x^4)}{(y^2 + x^4)^2}, \quad (2.4)$$

$$\frac{\partial f}{\partial y}(x, y) = \frac{x^2(y^2 + x^4) - 2y \cdot yx^2}{(y^2 + x^4)^2} = \frac{-x^2(y^2 - x^4)}{(y^2 + x^4)^2}. \quad (2.5)$$

Podobnie jak w poprzednim przykładzie, $f \equiv 0$ na obu osiach układu współrzędnych mamy i dlatego

$$\frac{\partial f}{\partial x}(\mathbf{0}) = 0 = \frac{\partial f}{\partial y}(\mathbf{0}).$$

Jednak wiemy już, że f nie jest ciągła w punkcie $\mathbf{0} \in \mathbb{R}^2$.

Przykład 2.4. Niech $f: \mathbb{R} \rightarrow \mathbb{R}$ będzie funkcją różniczkowalną i niech $G(x) = (x, f(x))$, $G: \mathbb{R} \rightarrow \mathbb{R}^2$. Wówczas $\frac{\partial G}{\partial x} = G'(x) = (1, f'(x))$. Interpretacja geometryczna jest prosta: $G'(x)$ jest wektorem stycznym do wykresu f w punkcie $(x, f(x)) \in \mathbb{R}^2$.

Definicja 2.5 (pochodna kierunkowa). Mówimy, że funkcja $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$, gdzie zbiór $\Omega \subset \mathbb{R}^n$ jest otwarty, ma w punkcie $\mathbf{a} = (a_1, \dots, a_n) \in \Omega$ *pochodną kierunkową względem wektora* $\mathbf{v} \in \mathbb{R}^n \setminus \{0\}$ wtedy i tylko wtedy, gdy funkcja *jednej* zmiennej rzeczywistej

$$F_{\mathbf{v}}(t) = f(\mathbf{a} + t\mathbf{v}), \quad F_{\mathbf{v}}: (-\delta, \delta) \rightarrow \mathbb{R}^m \quad (2.6)$$

ma pochodną w zerze. (Liczbę $\delta > 0$ wybieramy tak, by odcinek $(\mathbf{a} - \delta\mathbf{v}, \mathbf{a} + \delta\mathbf{v}) \subset \Omega$.) Przyjmujemy

$$\frac{\partial f}{\partial \mathbf{v}}(\mathbf{a}) = (F_{\mathbf{v}})'(0) = \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h\mathbf{v}) - f(\mathbf{a})}{h}. \quad (2.7)$$

Używa się także innych oznaczeń:

$$D_{\mathbf{v}}f(\mathbf{a}) = f'_{\mathbf{v}}(\mathbf{a}) = \frac{\partial f}{\partial \mathbf{v}}(\mathbf{a}).$$

Uwaga 2.6. Zauważmy, że pochodna kierunkowa względem $\mathbf{e}_i$ jest tym samym, co pochodna cząstkowa względem $\mathbf{e}_i$:

$$D_{\mathbf{e}_i}f(\mathbf{a}) = \frac{\partial f}{\partial x_i}(\mathbf{a}).$$

Przykład 2.7. Raz jeszcze rozpatrzmy funkcję z Przykładu 1.49, daną wzorem (2.3). Dla $\mathbf{a} = \mathbf{0}$ i dowolnego wektora $\mathbf{v} = (\xi, \eta) \in \mathbb{R}^2$, gdzie $\xi, \eta \neq 0$, iloraz

$$\frac{f(\mathbf{a} + h\mathbf{v}) - f(\mathbf{a})}{h} = \frac{1}{h} \cdot \frac{h^3 \xi^2 \eta}{h^2 \eta^2 + h^4 \xi^4} = \frac{\xi^2 \eta}{\eta^2 + h^2 \xi^4} \rightarrow \frac{\xi^2}{\eta} \quad \text{dla } h \rightarrow 0.$$

Zatem f ma w zerze *wszystkie* pochodne kierunkowe (sprawdzaliśmy już istnienie pochodnych cząstkowych). Nietrudno stwierdzić, że w pozostałych punktach $\mathbb{R}^2$ funkcja f też ma wszystkie pochodne kierunkowe. Wynika stąd, że nawet istnienie wszystkich pochodnych kierunkowych w każdym punkcie dziedziny nie gwarantuje ciągłości funkcji wielu zmiennych rzeczywistych.

Właściwym odpowiednikiem pojęcia pochodnej jest, dla funkcji wielu zmiennych, pojęcie różniczki.

Definicja 2.8 (różniczkowalność funkcji wielu zmiennych). Mówimy, że funkcja $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$, gdzie $\Omega \subset \mathbb{R}^n$ jest zbiorem otwartym, jest różniczkowalna w punkcie $\mathbf{a} \in \Omega$ wtedy i tylko wtedy, gdy istnieje przekształcenie liniowe $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ takie, że

$$\lim_{\|\mathbf{h}\| \rightarrow 0} \frac{\|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}\|}{\|\mathbf{h}\|} = 0. \quad (2.8)$$

Przekształcenie A nazywamy *różniczką (lub pochodną, lub różniczką zupełną) f w punkcie $\mathbf{a}$* i oznaczamy $Df(\mathbf{a})$ lub $f'(\mathbf{a})$.

Stwierdzenie 2.9. Niech $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$, gdzie $\Omega \subset \mathbb{R}^n$ jest zbiorem otwartym. Następujące warunki są równoważne:

- (i) f jest różniczkowalna w punkcie $\mathbf{a} \in \Omega$;
- (ii) istnieją przekształcenie liniowe $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ i funkcja

$$r: \Omega_{\mathbf{a}} = \{\mathbf{h} \in \mathbb{R}^n: \mathbf{a} + \mathbf{h} \in \Omega\} \rightarrow \mathbb{R}^m$$

ciągła w punkcie $\mathbf{h} = \mathbf{0}$, $r(\mathbf{0}) = \mathbf{0}$, dla których zachodzi równość

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + A\mathbf{h} + \|\mathbf{h}\| \cdot r(\mathbf{h}) \quad \text{dla wszystkich } \mathbf{h} \in \Omega_{\mathbf{a}}. \quad (2.9)$$

Jeśli zachodzi warunek (ii), to $Df(\mathbf{a}) = A$.

Dowód. Jeśli różniczka $A = Df(\mathbf{a})$ istnieje, to wystarczy określić

$$r(\mathbf{h}) = \frac{f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}}{\|\mathbf{h}\|} \quad \text{dla } \mathbf{h} \neq \mathbf{0}, \quad r(\mathbf{0}) = \mathbf{0}.$$

Funkcja r jest określona, gdy $\mathbf{a} + \mathbf{h} \in \Omega$. Ponadto, dzięki warunkowi (2.8),

$$\|r(\mathbf{h})\| = \frac{\|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}\|}{\|\mathbf{h}\|} \rightarrow 0, \quad \text{dla } \|\mathbf{h}\| \rightarrow 0,$$

tzn. równoważnie $r(\mathbf{h}) \rightarrow 0$ dla $\mathbf{h} \rightarrow 0$. Na odwrót, jeśli zachodzi (ii), to warunek $r(\mathbf{h}) \rightarrow 0 = r(0)$ dla $\mathbf{h} \rightarrow 0$ implikuje, że granica we wzorze (2.8) jest równa zero, tzn. $A = Df(\mathbf{a})$. $\square$

Uwaga 2.10. Jeśli różniczka $Df(\mathbf{a})$ istnieje, to jest określona jednoznacznie. Istotnie, gdyby wzór (2.9) zachodził dla A_i, r_i , gdzie $i = 1, 2$, to mielibyśmy $A_1\mathbf{h} + \|\mathbf{h}\|r_1(\mathbf{h}) = A_2\mathbf{h} + \|\mathbf{h}\|r_2(\mathbf{h})$ dla wszystkich $\mathbf{h}$ z pewnej kuli $B(0, \delta) \subset \mathbb{R}^n$. Stąd, kładąc $\mathbf{h} = t \cdot \mathbf{v}$, gdzie $\|\mathbf{v}\| = 1$ i $t > 0$, a następnie dzieląc obie strony przez t , otrzymujemy

$$(A_1 - A_2)\mathbf{v} = r_2(t\mathbf{v}) - r_1(t\mathbf{v}) \rightarrow 0, \quad t \rightarrow 0.$$

Jednak lewa strona nie zależy od t . Zatem przekształcenia liniowe A_1 i A_2 pokrywają się na całej sferze jednostkowej $\mathbb{S}^{n-1} = \{\mathbf{v} \in \mathbb{R}^n : \|\mathbf{v}\| = 1\}$, a więc są równe.

Wniosek 2.11. Jeśli $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ jest różniczkowalna w punkcie $\mathbf{a} \in \Omega$, to dla każdego niezerowego wektora $\mathbf{v} \in \mathbb{R}^n$ jest

$$Df(\mathbf{a}) \cdot \mathbf{v} = \frac{\partial f}{\partial \mathbf{v}}(\mathbf{a}).$$

W szczególności, dla $\mathbf{v} = \mathbf{e}_i$ jest

$$Df(\mathbf{a}) \cdot \mathbf{e}_i = \frac{\partial f}{\partial x_i}(\mathbf{a}), \quad i = 1, \dots, n.$$

Dowód. Podstawiając w równości (2.9) wektor $\mathbf{h} = t\mathbf{v}$, gdzie $\mathbf{v}$ jest ustalony i $t \neq 0$, otrzymujemy

$$Df(\mathbf{a}) \cdot \mathbf{v} = \frac{1}{t} Df(\mathbf{a}) \cdot \mathbf{h} = \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t} \pm \|\mathbf{v}\| r(t\mathbf{v}) \rightarrow \frac{\partial f}{\partial \mathbf{v}}(\mathbf{a}), \quad t \rightarrow 0.$$

Lewa strona nie zależy od t ; dlatego zachodzi pierwsza równość z tezy wniosku. Druga równość to jej przypadek szczególny (wspominaliśmy już, że pochodna f w kierunku $\mathbf{e}_i$ i pochodna cząstkowa $\frac{\partial f}{\partial x_i}$ są równe). $\square$

Wniosek 2.12. Niech $\Omega \subset \mathbb{R}^n$ będzie zbiorem otwartym i niech

$$f = (f_1, \dots, f_m): \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m.$$

Następujące warunki są równoważne:

- (i) funkcja f jest różniczkowalna w punkcie $\mathbf{a} \in \Omega$;
- (ii) każda z funkcji f_i jest różniczkowalna w punkcie $\mathbf{a} \in \Omega$.

W standardowych bazach przestrzeni $\mathbb{R}^n$ i $\mathbb{R}^m$ macierz $Df_i(\mathbf{a}) \in M_{1 \times n}$ jest wtedy i -tym wierszem macierzy $Df(\mathbf{a}) \in M_{m \times n}$.

Dowód. Posługujemy się Stwierdzeniem 2.9. Obie strony równości (2.9) są wektorami z $\mathbb{R}^m$. Równość i -tych współrzędnych tych wektorów ($i = 1, \dots, m$) jest równoważna różniczkowalności f_i ($i = 1, \dots, m$) w punkcie $\mathbf{a} \in A$ oraz równości $Df_i(\mathbf{a})\mathbf{h} = (Df(\mathbf{a})\mathbf{h})_i$ dla $\mathbf{h} \in \mathbb{R}^n$, tzn. – po utożsamieniu przekształceń liniowych z ich macierzami w standardowych bazach – temu, że $Df_i(\mathbf{a})$ jest i -tym wierszem macierzy $Df(\mathbf{a})$. $\square$

Uwaga 2.13 (terminologia: macierz Jacobiego). Jak wiadomo z wykładów Algbrę Liniowej, przekształcenie liniowe $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ma w standardowych bazach macierz (oznaczaną zwykle tą samą literą) o m wierszach i n kolumnach, której j -tą kolumnę stanowi wektor $A\mathbf{e}_j \in \mathbb{R}^m$, gdzie $\mathbf{e}_j$ ($j = 1, \dots, n$) są wektorami standardowej bazy w $\mathbb{R}^n$. Z dwóch ostatnich wniosków wypływa zatem następująca obserwacja: jeśli

$$f = (f_1, \dots, f_m): \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m,$$

gdzie $f_i: \Omega \rightarrow \mathbb{R}$ dla $i = 1, \dots, m$, jest różniczkowalna w punkcie $\mathbf{a} \in \Omega$, to jej różniczka ma w standardowych bazach przestrzeni $\mathbb{R}^n$ i $\mathbb{R}^m$ macierz

$$Df(\mathbf{a}) = \left(\frac{\partial f_i}{\partial x_j}(\mathbf{a}) \right)_{1 \leq i \leq m, 1 \leq j \leq n} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}. \quad (2.10)$$

Kolumny tej macierzy to wektory

$$Df(\mathbf{a})\mathbf{e}_i = \frac{\partial f}{\partial \mathbf{e}_i}(\mathbf{a}) = \frac{\partial f}{\partial x_i}(\mathbf{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_i}(\mathbf{a}) \\ \vdots \\ \frac{\partial f_m}{\partial x_i}(\mathbf{a}) \end{pmatrix} \in \mathbb{R}^m, \quad i = 1, \dots, n.$$

Macierz (2.10) nazywamy *macierzą Jacobiego* przekształcenia różniczkowalnego f . Dla $n = m$ wyznacznik tej macierzy nazywamy *jacobianem* przekształcenia f w punkcie $\mathbf{a}$.

Wniosek 2.14. Jeśli $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ jest różniczkowalna w punkcie $\mathbf{a} \in \Omega$, to f jest ciągła w punkcie $\mathbf{a}$.

Dowód. Korzystamy ze Stwierdzenia 2.9 (ii) oraz ciągłości przekształceń liniowych: dla $\mathbf{h} \rightarrow 0$ jest

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = Df(\mathbf{a})\mathbf{h} + \|\mathbf{h}\| \cdot r(\mathbf{h}) \rightarrow \mathbf{0} \in \mathbb{R}^m,$$

to zaś oznacza, że f jest ciągła w punkcie $\mathbf{a}$. $\square$

Uwaga 2.15. Wiemy już zatem, że istnienie pochodnych cząstkowych funkcji f w danym punkcie jest warunkiem *koniecznym* różniczkowalności f w tym punkcie. Nie jest jednak warunkiem *dostatecznym*, gdyż z istnienia pochodnych cząstkowych (a nawet wszystkich pochodnych kierunkowych) nie wynika ciągłość! Poniżej podajemy warunek *dostateczny* różniczkowalności, wyrażony w języku pochodnych cząstkowych.

Twierdzenie 2.16. *Jeśli $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ i wszystkie pochodne cząstkowe $\frac{\partial f}{\partial x_i}$ istnieją na całej kuli $B(\mathbf{a}, r) \subset \Omega$ i są ciągłe w punkcie $\mathbf{a}$, gdzie $r > 0$, to f jest różniczkowalna w punkcie $\mathbf{a} \in \Omega$. Zachodzi wtedy wzór*

$$Df(\mathbf{a})\mathbf{h} = \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}(\mathbf{a}), \quad \mathbf{h} = (h_1, \dots, h_n) \in \mathbb{R}^n.$$

Dowód. Wobec Stwierdzenia 2.12, wystarczy przeprowadzić dowód dla $m = 1$.

Dla uproszczenia¹ przyjmijmy $n = 2$. Niech odąd $\mathbf{h} = (h_1, h_2)$, $\|\mathbf{h}\| < r$.

Aby skorzystać z istnienia pochodnych cząstkowych, wyrazimy przyrost f na odcinku $[\mathbf{a}, \mathbf{a} + \mathbf{h}]$ jako sumę przyrostów wzdłuż dwóch odcinków równoległych do osi układu współrzędnych. Stosując twierdzenie Lagrange’a dla funkcji jednej zmiennej rzeczywistej o wartościach rzeczywistych do funkcji

$$F_1(t) = f(a_1 + t, a_2 + h_2), \quad t \in [0, h_1],$$

oraz

$$F_2(t) = f(a_1, a_2 + t), \quad t \in [0, h_2]$$

sprawdzamy, że dla pewnych punktów pośrednich $\theta_i = \theta_i(\mathbf{h}) \in [0, h_i]$ ($i = 1, 2$) jest

$$\begin{aligned} f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) &= (F_1(h_1) - F_1(0)) + (F_2(h_2) - F_2(0)) \\ &= h_1 F_1'(\theta_1) + h_2 F_2'(\theta_2) \\ &= h_1 \frac{\partial f}{\partial x_1}(\underbrace{a_1 + \theta_1, a_2 + h_2}_{\text{punkt } \mathbf{p}_1}) + h_2 \frac{\partial f}{\partial x_2}(\underbrace{a_1, a_2 + \theta_2}_{\text{punkt } \mathbf{p}_2}) \\ &= \underbrace{h_1 \frac{\partial f}{\partial x_1}(\mathbf{a}) + h_2 \frac{\partial f}{\partial x_2}(\mathbf{a})}_{\text{część liniowa przyrostu}} + R(\mathbf{h}), \end{aligned}$$

gdzie reszta

$$R(\mathbf{h}) = h_1 \left(\frac{\partial f}{\partial x_1}(\mathbf{p}_1) - \frac{\partial f}{\partial x_1}(\mathbf{a}) \right) + h_2 \left(\frac{\partial f}{\partial x_2}(\mathbf{p}_2) - \frac{\partial f}{\partial x_2}(\mathbf{a}) \right).$$

Z nierówności Schwarza wynika, że dla $\|\mathbf{h}\| < \delta < r$ jest

$$0 \leq |R(\mathbf{h})| \leq \sqrt{2} \|\mathbf{h}\| \cdot \max_{i=1,2} \left(\sup_{\mathbf{p} \in B(\mathbf{a}, \delta)} \left| \frac{\partial f}{\partial x_i}(\mathbf{p}) - \frac{\partial f}{\partial x_i}(\mathbf{a}) \right| \right).$$

Jednak

$$\sup_{\mathbf{p} \in B(\mathbf{a}, \delta)} \left| \frac{\partial f}{\partial x_i}(\mathbf{p}) - \frac{\partial f}{\partial x_i}(\mathbf{a}) \right| \quad \text{dla } \delta \rightarrow 0$$

dzięki ciągłości $\frac{\partial f}{\partial x_i}$ w punkcie $\mathbf{a}$. Zatem

$$R(\mathbf{h}) = \|\mathbf{h}\| \cdot r(\mathbf{h}),$$

¹Chodzi o uproszczenie zapisu, a nie istotnych trudności – Czytelnik zechce się nad tym zastanowić. W ogólnym przypadku mielibyśmy w dowodzie do czynienia z sumą n przyrostów, a nie dwóch.

gdzie $r(\mathbf{0}) = 0$ i $r(\mathbf{h}) \rightarrow 0$ dla $\mathbf{h} \rightarrow \mathbf{0}$. Ze Stwierdzenia 2.9 wynika teraz, że

$$Df(\mathbf{a}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}) \right).$$

Dowód został zakończony. $\square$

Na zakończenie tego podrozdziału podkreślmy jedno. Czytelnikowi może wydawać się, że pochodna cząstkowa i być może pochodna kierunkowa to pojęcia naturalniejsze od różniczki. Tak nie jest. Pochodne cząstkowe i kierunkowe określa się po to, żeby badać zachowanie funkcji na prostych. Z przytoczonych przykładów wynika jasno, że nie daje to dostatecznych informacji o zachowaniu funkcji w całym *otoczeniu* danego punktu. Naturalnym uogólnieniem pochodnej funkcji jednej zmiennej rzeczywistej jest właśnie różniczka. Jej istnienie oznacza, że odwzorowanie f można lokalnie przybliżać przekształceniami

$$\mathbf{x} \mapsto f(\mathbf{a}) + Df(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) = f(\mathbf{x}) + o(\|\mathbf{x} - \mathbf{a}\|) \approx f(\mathbf{x}).$$

Ponadto, przekształcenie liniowe $Df(\mathbf{a})$ koduje w sobie pełną informację o pochodnych kierunkowych i cząstkowych f .

2.2 Arytmetyczne własności różniczki

Stwierdzenie 2.17 (różniczka sumy funkcji). *Jeśli $f, g: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ są różniczkowalne w punkcie $\mathbf{a} \in \Omega$, to funkcja $f + g: \Omega \rightarrow \mathbb{R}^m$ jest różniczkowalna w $\mathbf{a}$ i zachodzi wzór*

$$D(f + g)(\mathbf{a}) = Df(\mathbf{a}) + Dg(\mathbf{a}).$$

Dowód. Stosujemy Stwierdzenie 2.9. Wzory 2.9 dla funkcji f, g dodajemy stronami; ponieważ $o(\|\mathbf{h}\|) + o(\|\mathbf{h}\|) = o(\|\mathbf{h}\|)$, więc uzyskujemy warunek (ii) Stwierdzenia 2.9 dla funkcji $f + g$. Szczegóły pozostawiamy Czytelnikowi jako proste ćwiczenie. $\square$

Uogólnimy teraz wzór $(fg)' = f'g + fg'$ na przypadek wielowymiarowy. Okazuje się, że jeśli można zdefiniować ‘iloczyn’ przekształceń różniczkowalnych f, g (to może być np. iloczyn funkcji o wartościach w $\mathbb{R}$ i $\mathbb{R}^m$, albo iloczyn skalarny wektorów z $\mathbb{R}^m$, albo iloczyn wektorowy wektorów z $\mathbb{R}^3$, albo iloczyn macierzy o odpowiednich rozmiarach, gdy wartości f, g są macierzami itp.), to ów iloczyn jest różniczkowalny, a jego pochodną oblicza się podobnie, jak dla funkcji z $\mathbb{R}$ w $\mathbb{R}$.

Twierdzenie 2.18 (różniczka ‘iloczynu’). *Jeśli $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ i $g: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^k$ są różniczkowalne w punkcie $\mathbf{a} \in \Omega$, a przekształcenie*

$$B: \mathbb{R}^m \times \mathbb{R}^k \ni (\mathbf{x}, \mathbf{y}) \mapsto B[\mathbf{x}, \mathbf{y}] \in \mathbb{R}^l$$

jest dwuliniowe², to wówczas funkcja

$$B[f, g]: \mathbb{R}^n \supset \Omega \ni \mathbf{x} \mapsto B[f(\mathbf{x}), g(\mathbf{x})] \in \mathbb{R}^l$$

jest różniczkowalna w punkcie $\mathbf{a} \in \Omega$ i zachodzi równość

$$DB[f, g](\mathbf{a})\mathbf{h} = B[Df(\mathbf{a})\mathbf{h}, g(\mathbf{a})] + B[f(\mathbf{a}), Dg(\mathbf{a})\mathbf{h}] \quad \text{dla wszystkich } \mathbf{h} \in \mathbb{R}^n. \quad (2.11)$$

²Oznaczenie $B[f, g]$ Czytelnik może zastąpić przez $f \cdot g$ – wtedy analogia z przypadkiem jednowymiarowym będzie widoczna jak na dłoni.

Uwaga 2.19. Nie zakładamy, że $B[f, g] = B[g, f]$ (bo np. mnożenie macierzy nie jest działaniem przemennym). Dlatego we wzorze (2.11) nie wolno zamieniać kolejności argumentów B w składnikach prawej strony.

DOWÓD TWIERDZENIA 2.18. Wobec Stwierdzenia 2.9,

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = Df(\mathbf{a})\mathbf{h} + R_f(\mathbf{h}), \quad (2.12)$$

$$g(\mathbf{a} + \mathbf{h}) - g(\mathbf{a}) = Dg(\mathbf{a})\mathbf{h} + R_g(\mathbf{h}), \quad (2.13)$$

gdzie $\|R_f(\mathbf{h})\| = \|R_g(\mathbf{h})\| = o(\|\mathbf{h}\|)$ dla $\mathbf{h} \rightarrow 0$. Ustalmy zatem liczbę $\delta > 0$ tak, aby dla wszystkich $\|\mathbf{h}\| < \delta$ mieć

$$\|R_f(\mathbf{h})\| + \|R_g(\mathbf{h})\| < \|\mathbf{h}\|. \quad (2.14)$$

Korzystając z dwuliniowości B , piszemy

$$\begin{aligned} & B[f(\mathbf{a} + \mathbf{h}), g(\mathbf{a} + \mathbf{h})] - B[f(\mathbf{a}), g(\mathbf{a})] \\ &= B[f(\mathbf{a} + \mathbf{h}), g(\mathbf{a} + \mathbf{h})] - B[f(\mathbf{a}), g(\mathbf{a} + \mathbf{h})] \\ &\quad + B[f(\mathbf{a}), g(\mathbf{a} + \mathbf{h})] - B[f(\mathbf{a}), g(\mathbf{a})] \\ &= B[f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}), g(\mathbf{a} + \mathbf{h})] + B[f(\mathbf{a}), g(\mathbf{a} + \mathbf{h}) - g(\mathbf{a})] \\ &= B[f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}), g(\mathbf{a})] + B[f(\mathbf{a}), g(\mathbf{a} + \mathbf{h}) - g(\mathbf{a})] \\ &\quad + B[f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}), g(\mathbf{a} + \mathbf{h}) - g(\mathbf{a})] \stackrel{\text{ozn.}}{=} S_1 + S_2 + S_3. \end{aligned}$$

Do prawej strony wstawiamy teraz równości (2.12) i (2.13). Składnik

$$S_1 = B[f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}), g(\mathbf{a})] = B[Df(\mathbf{a})\mathbf{h}, g(\mathbf{a})] + B[R_f(\mathbf{h}), g(\mathbf{a})],$$

gdzie $B[R_f(\mathbf{h}), g(\mathbf{a})] = o(\|\mathbf{h}\|)$ dla $\mathbf{h} \rightarrow 0$ (to łatwo wynika z dwuliniowości B). Podobnie,

$$S_2 = B[f(\mathbf{a}), Dg(\mathbf{a})\mathbf{h}] + B[f(\mathbf{a}), R_g(\mathbf{h})] = B[f(\mathbf{a}), Dg(\mathbf{a})\mathbf{h}] + o(\|\mathbf{h}\|), \quad \mathbf{h} \rightarrow 0.$$

Dlatego suma $S_1 + S_2$ daje prawą stronę wzoru (2.11) z tezy, z błędem $o(\|\mathbf{h}\|)$. Wreszcie, składnik $S_3 = o(\|\mathbf{h}\|)$ dla $\mathbf{h} \rightarrow 0$. Istotnie, każde przekształcenie dwuliniowe B spełnia nierówność

$$\|B[\mathbf{x}, \mathbf{y}]\| \leq C\|\mathbf{x}\| \cdot \|\mathbf{y}\|$$

z pewną stałą C zależną od B .³ Dlatego

$$\begin{aligned} \|S_3\| &\leq C\|Df(\mathbf{a})\mathbf{h} + R_f(\mathbf{h})\| \cdot \|Dg(\mathbf{a})\mathbf{h} + R_g(\mathbf{h})\| \\ &\leq C \cdot M^2\|\mathbf{h}\|^2 \quad \text{dla } \|\mathbf{h}\| < \delta, \end{aligned}$$

gdzie, wobec oszacowania (2.14), można wziąć np. stałą $M = \|Df(\mathbf{a})\| + \|Dg(\mathbf{a})\| + 1$. Ostatecznie więc

$$\begin{aligned} & B[f(\mathbf{a} + \mathbf{h}), g(\mathbf{a} + \mathbf{h})] - B[f(\mathbf{a}), g(\mathbf{a})] = \\ &= S_1 + S_2 + S_3 = \text{prawa strona wzoru (2.11)} + o(\|\mathbf{h}\|) \quad \text{dla } \mathbf{h} \rightarrow 0. \end{aligned}$$

Wobec Stwierdzenia 2.9, dowód jest zakończony. $\square$

³Czytelnik może to udowodnić samodzielnie, naśladowując dowód Stwierdzenia 1.41.

Twierdzenie 2.20 (różniczka złożenia funkcji). Niech $\Omega_1 \subset \mathbb{R}^n$ i $\Omega_2 \subset \mathbb{R}^m$ będą zbiorami otwartymi. Jeśli $f: \Omega_1 \rightarrow \mathbb{R}^m$ jest różniczkowalne w punkcie $\mathbf{a} \in \Omega_1$, a $\Omega_2 \supset f(\Omega_1)$ i $g: \Omega_2 \rightarrow \mathbb{R}^k$ jest różniczkowalne w punkcie $\mathbf{b} = f(\mathbf{a})$, to złożenie $g \circ f$ jest różniczkowalne w punkcie $\mathbf{a}$ i zachodzi wzór

$$D(g \circ f)(\mathbf{a}) = Dg(\mathbf{b}) \circ Df(\mathbf{a}) = Dg(f(\mathbf{a})) \circ Df(\mathbf{a}). \quad (2.15)$$

Dowód. Wobec Stwierdzenia 2.9,

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = Df(\mathbf{a})\mathbf{h} + R_f(\mathbf{h}), \quad (2.16)$$

$$g(\mathbf{b} + \mathbf{w}) - g(\mathbf{b}) = Dg(\mathbf{b})\mathbf{w} + R_g(\mathbf{w}), \quad (2.17)$$

gdzie $\|R_f(\mathbf{h})\| = o(\|\mathbf{h}\|)$ dla $\mathbf{h} \rightarrow 0$ i $\|R_g(\mathbf{w})\| = o(\|\mathbf{w}\|)$ dla $\mathbf{w} \rightarrow 0$. Do (2.17) podstawmy $\mathbf{b} = f(\mathbf{a})$ oraz $\mathbf{w} = \mathbf{w}(\mathbf{h}) = f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a})$. Korzystając z (2.16), otrzymujemy

$$\begin{aligned} g \circ f(\mathbf{a} + \mathbf{h}) - g \circ f(\mathbf{a}) &= g(\mathbf{b} + \mathbf{w}) - g(\mathbf{b}) \\ &= Dg(\mathbf{b})\mathbf{w} + R_g(\mathbf{w}) \\ &= Dg(\mathbf{b})(Df(\mathbf{a})\mathbf{h} + R_f(\mathbf{h})) + R_g(\mathbf{w}) \\ &= [Dg(\mathbf{b}) \circ Df(\mathbf{a})]\mathbf{h} + R, \end{aligned} \quad (2.18)$$

gdzie reszta

$$R = Dg(\mathbf{b})(R_f(\mathbf{h})) + R_g(\mathbf{w}).$$

Niech $M = \|Dg(\mathbf{b})\| + \|Df(\mathbf{a})\| + 1$. Dla małych $\|\mathbf{h}\|$ jest $\|R_f(\mathbf{h})\| < \|\mathbf{h}\|$ i dlatego

$$\|\mathbf{w}\| = \|Df(\mathbf{a})\mathbf{h} + R_f(\mathbf{h})\| \leq \|Df(\mathbf{a})\| \cdot \|\mathbf{h}\| + \|R_f(\mathbf{h})\| \leq M\|\mathbf{h}\|.$$

Zatem $\mathbf{w} = \mathbf{w}(\mathbf{h}) \rightarrow 0$, gdy $\mathbf{h} \rightarrow 0$ i mamy

$$\begin{aligned} \frac{\|R\|}{\|\mathbf{h}\|} &\leq \frac{M\|R_f(\mathbf{h})\| + \|R_g(\mathbf{w})\|}{\|\mathbf{h}\|} \\ &= M \frac{\|R_f(\mathbf{h})\|}{\|\mathbf{h}\|} + \frac{\|\mathbf{w}\|}{\|\mathbf{h}\|} \cdot \frac{\|R_g(\mathbf{w})\|}{\|\mathbf{w}\|} \\ &\leq M \left(\frac{\|R_f(\mathbf{h})\|}{\|\mathbf{h}\|} + \frac{\|R_g(\mathbf{w})\|}{\|\mathbf{w}\|} \right) \rightarrow 0 \quad \text{dla } \mathbf{h} \rightarrow 0, \end{aligned}$$

tzn. $R = o(\|\mathbf{h}\|)$ dla $\mathbf{h} \rightarrow 0$. Wobec równości (2.18) i Stwierdzenia 2.9, zachodzi równość $D(g \circ f)(\mathbf{a}) = Dg(\mathbf{b}) \circ Df(\mathbf{a})$. $\square$

Uwaga 2.21. Zgodnie z definicją różniczka $D(g \circ f)(\mathbf{a})$ złożenia $g \circ f: \mathbb{R}^n \supset \Omega_1 \rightarrow \mathbb{R}^k$ powinna być przekształceniem liniowym z $\mathbb{R}^n$ w $\mathbb{R}^k$. Istotnie tak jest: $Df(\mathbf{a}): \mathbb{R}^n \rightarrow \mathbb{R}^m$ i $Dg(\mathbf{b}): \mathbb{R}^m \rightarrow \mathbb{R}^k$, więc ich złożenie jest przekształceniem liniowym z $\mathbb{R}^n$ w $\mathbb{R}^k$.

Twierdzenie 2.20 ma następującą interpretację: macierz Jacobiego różniczki przekształcenia $g \circ f$ jest iloczynem macierzy Jacobiego przekształceń g i f , wziętych w odpowiednich punktach. Z wieloma zastosowaniami tej interpretacji Czytelnik spotka się w konkretnych przykładach.

2.3 Gradient. Płaszczyzna styczna do wykresu funkcji i punkty krytyczne

Definicja 2.22 (gradient funkcji wielu zmiennych). Gradientem funkcji różniczkowalnej $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}$ w punkcie $\mathbf{x} \in \Omega$ nazywamy wektor

$$\text{grad } f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right).$$

Z tej definicji oraz definicji macierzy Jacobiego wynika, że dla $f: \Omega \rightarrow \mathbb{R}$ mamy

$$\langle \text{grad } f(\mathbf{x}), \mathbf{h} \rangle = Df(\mathbf{x})\mathbf{h} = \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}(\mathbf{x}).$$

Czytelnik może więc uważać, że rozróżnianie gradientu i różniczki to niepotrzebna maniera. Podkreślmy jednak, że $\text{grad } f(\mathbf{x}) \in \mathbb{R}^n$, natomiast $Df(\mathbf{x}) \in L(\mathbb{R}^n, \mathbb{R}) = (\mathbb{R}^n)^*$. Przywykliśmy utożsamiać przestrzenie $(\mathbb{R}^n)^*$ i $\mathbb{R}^n$; to wymaga odwołania się do konkretnego układu współrzędnych. Na gładkich powierzchniach w $\mathbb{R}^n$ – powiedzmy na torusie czy na sferze – nie sposób jednak zwykle wskazać jakiegoś wyróżnionego układu współrzędnych. Dlatego odróżnianie gradientu i różniczki ma sens. Czytelnik zetknie się z tym później w bieżącym roku, a także na zajęciach z geometrii różniczkowej.

Stwierdzenie 2.23. Dla każdej funkcji różniczkowalnej $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}$ i każdego wektora $\mathbf{w} \in \mathbb{R}^n$ takiego, że

$$\|\mathbf{w}\| = \|\mathbf{v}\|, \quad \text{gdzie } \mathbf{v} = \text{grad } f(\mathbf{x}) \neq 0, \quad \mathbf{x} \in \Omega,$$

zachodzi nierówność

$$\frac{\partial f}{\partial \mathbf{w}}(\mathbf{x}) \leq \frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}). \quad (2.19)$$

Ponadto, równość w (2.19) zachodzi jedynie dla $\mathbf{w} = \mathbf{v}$.

Dowód. Wobec Wniosku 2.11 i nierówności Schwarza, mamy

$$\frac{\partial f}{\partial \mathbf{w}}(\mathbf{x}) = Df(\mathbf{x})\mathbf{w} = \langle \text{grad } f(\mathbf{x}), \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle \leq \|\mathbf{v}\| \|\mathbf{w}\| = \|\mathbf{v}\|^2,$$

natomiast

$$\frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}) = Df(\mathbf{x})\mathbf{v} = \langle \text{grad } f(\mathbf{x}), \mathbf{v} \rangle = \|\mathbf{v}\|^2.$$

Stąd już wynika nierówność (2.19).

W nierówności Schwarza $|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\|$ dla wektorów $\mathbf{v}, \mathbf{w}$ o równych długościach równość zachodzi wtedy i tylko wtedy, gdy $\mathbf{v} = \pm \mathbf{w}$. Jednak dla $\mathbf{w} = -\mathbf{v}$ jest

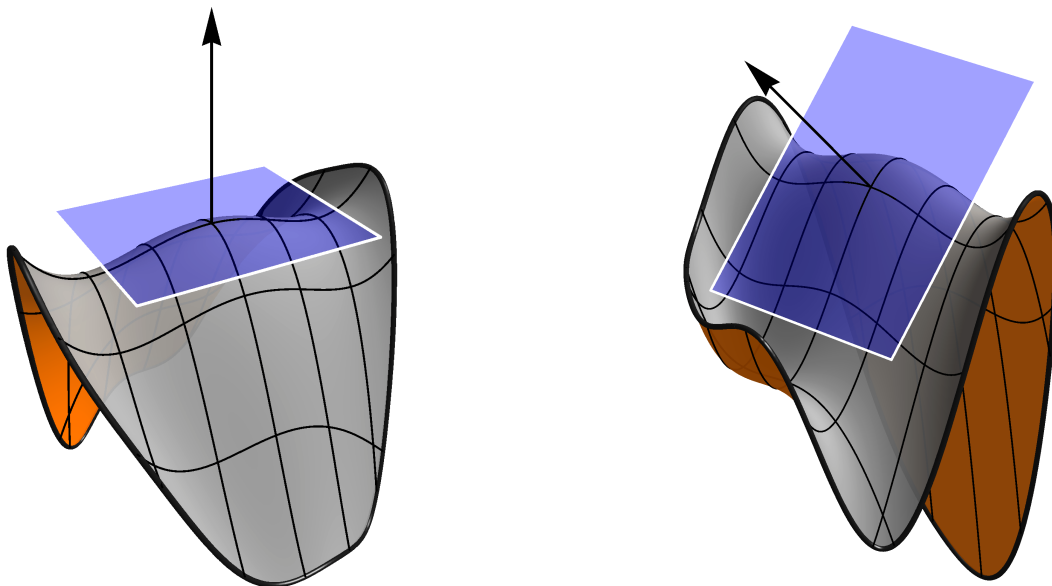
$$\frac{\partial f}{\partial \mathbf{w}}(\mathbf{x}) = \langle \mathbf{v}, \mathbf{w} \rangle = -\|\mathbf{v}\|^2 < 0 < \|\mathbf{v}\|^2 = \frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}).$$

Ta obserwacja kończy dowód. $\square$

Powyższe stwierdzenie ma następującą interpretację geometryczną: gradient funkcji w punkcie wyznacza kierunek najszybszego wzrostu funkcji w tym punkcie. Długość wektora gradientu odpowiada za tempo wzrostu w tym kierunku.

Definicja 2.24 (płaszczyzna styczna do wykresu funkcji). Jeśli $f: \Omega \rightarrow \mathbb{R}$ jest różniczkowalna w punkcie $\mathbf{a}$, to płaszczyzną styczną do wykresu f w punkcie $(\mathbf{a}, f(\mathbf{a})) \in \mathbb{R}^{n+1}$ nazywamy zbiór

$$T = \{(\mathbf{x}, x_{n+1}) \in \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1} : Df(\mathbf{a})(\mathbf{x} - \mathbf{a}) = x_{n+1} - f(\mathbf{a})\}$$



Płaszczyzna styczna do wykresu funkcji różniczkowalnej $f: \mathbb{R}^2 \supset \Omega \rightarrow \mathbb{R}$ w punkcie $(\mathbf{p}, f(\mathbf{p})) \in \mathbb{R}^3$ ma równanie

$$f_{x_1}(\mathbf{p})(x_1 - p_1) + f_{x_2}(\mathbf{p})(x_2 - p_2) = x_3 - f(\mathbf{p})$$

Wektor $(-f_{x_1}(\mathbf{p}), -f_{x_2}(\mathbf{p}), 1) \in \mathbb{R}^3$ jest prostopadły do płaszczyzny stycznej.

Po lewej: w punkcie ekstremum lokalnego $f_x = f_y = 0$ i tam płaszczyzna styczna jest pozioma.

Innymi słowy, zbiór T jest wykresem odwzorowania afinicznego

$$\mathbb{R}^n \ni \mathbf{x} \mapsto \phi(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x} - \mathbf{a}) \in \mathbb{R}.$$

Wprost z definicji różniczki wynika, że dla $\mathbf{x} \rightarrow \mathbf{a}$ jest $f(\mathbf{x}) - \phi(\mathbf{x}) = o(\|\mathbf{x} - \mathbf{a}\|)$. Widzieliśmy też, że ten warunek określa odwzorowanie ϕ jednoznacznie. To uzasadnia nazwę *płaszczyzna styczna*. Zauważmy, że jeśli $Df(\mathbf{a}) = \mathbf{0}$, to przekształcenie ϕ jest stałe, a więc jego wykresem jest hiperpłaszczyzna $x_{n+1} = \text{const}$.

Podamy teraz warunek konieczny istnienia ekstremum⁴ funkcji w punkcie wewnętrznym dziedziny.

Stwierdzenie 2.25 (lemat Fermata). Jeśli funkcja $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}$ ma ekstremum lokalne w punkcie $\mathbf{a} \in \Omega$ i jest w tym punkcie różniczkowalna, to

$$Df(\mathbf{a}) = \mathbf{0},$$

lub równoważnie $\frac{\partial f}{\partial x_i}(\mathbf{a}) = 0$ dla $i = 1, 2, \dots, n$.

⁴Definicja maksimum (minimum) lokalnego jest analogiczna, jak w wymiarze 1; trzeba tylko przedział wokół danego punktu w dziedzinie zastąpić kulą o środku w tym punkcie.

Dowód. Jeśli f ma ekstremum lokalne w $\mathbf{a}$, to dla każdego $i = 1, \dots, n$ funkcja

$$F_i(t) = f(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_n),$$

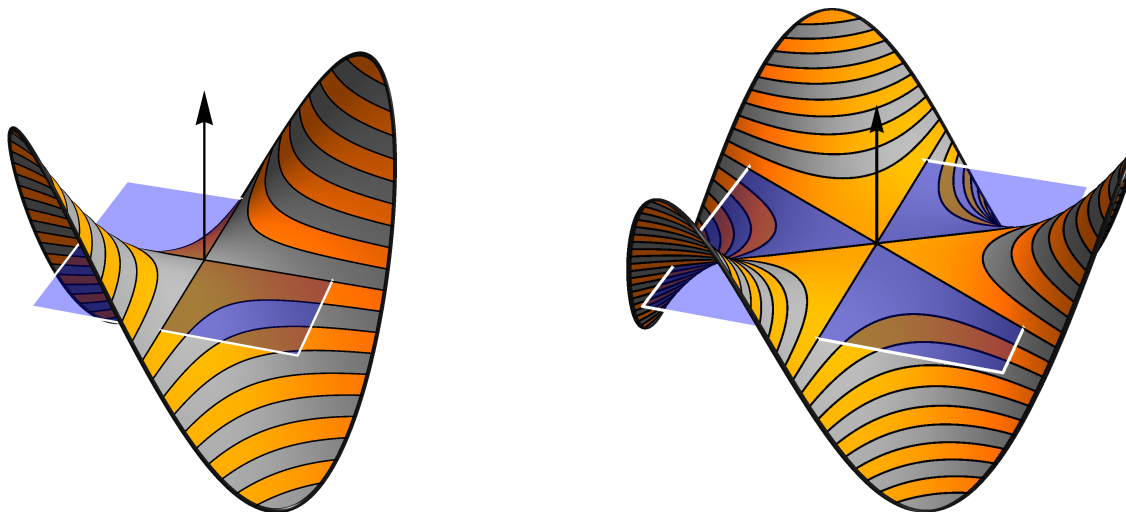
określona w pewnym przedziale $(a_i - \delta, a_i + \delta) \subset \mathbb{R}$, ma ekstremum lokalne w a_i . Dlatego $F'_i(a_i) = \frac{\partial f}{\partial x_i}(\mathbf{a}) = Df(\mathbf{a})\mathbf{e}_i = 0$. $\square$

W wielu sytuacjach wykorzystywane jest następujące ogólne pojęcie.

Definicja 2.26. Przypuśćmy, że odwzorowanie $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ jest różniczkowalne na Ω . Powiemy, że $\mathbf{a} \in \Omega$ jest *punktem krytycznym* f wtedy i tylko wtedy, gdy przekształcenie liniowe $Df(\mathbf{a}): \mathbb{R}^n \rightarrow \mathbb{R}^m$ ma rząd mniejszy, niż $\min(m, n)$.⁵

Gdy $m = 1$, to $\mathbf{a} \in \Omega$ jest punktem krytycznym funkcji $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}$ wtedy i tylko wtedy, gdy $Df(\mathbf{a}): \mathbb{R}^n \rightarrow \mathbb{R}$ ma rząd mniejszy, niż $\min(n, 1) = 1$, tzn. wtedy i tylko wtedy, gdy $Df(\mathbf{a})$ jest przekształceniem zerowym. Jest to równoważne temu, że wszystkie pochodne cząstkowe $\frac{\partial f}{\partial x_i}$ znikają w punkcie $\mathbf{a}$. Zachodzi zatem następujące:

Stwierdzenie 2.27. Wszystkie ekstrema lokalne funkcji różniczkowalnej $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}$ są jej punktami krytycznymi. $\square$

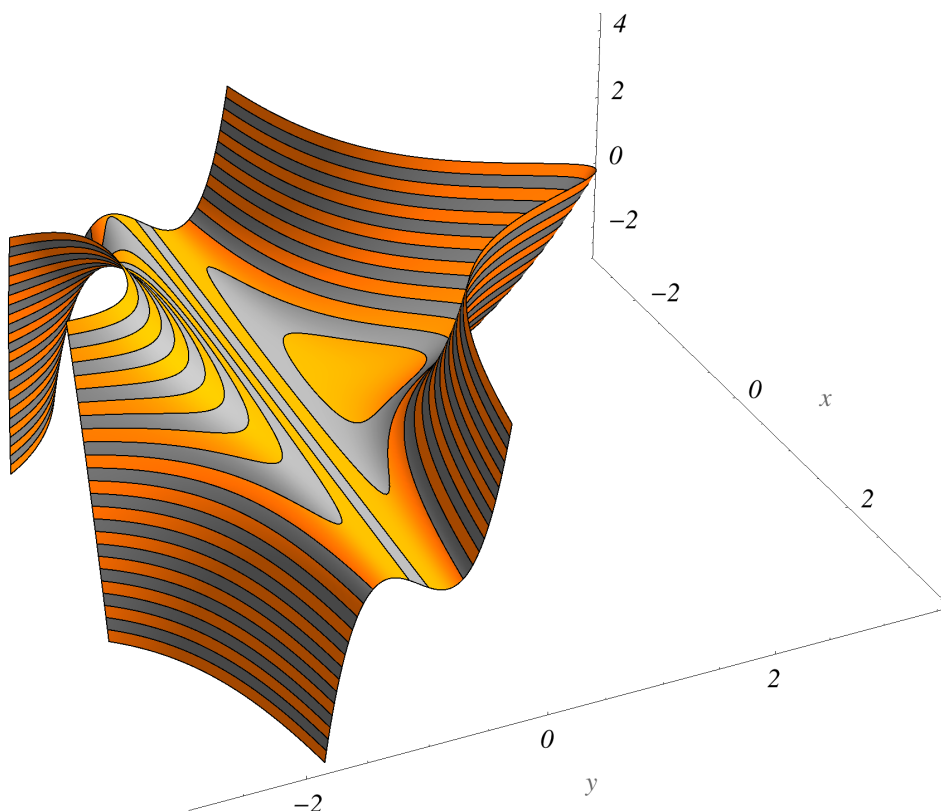


Po lewej: wykres funkcji $f(x, y) = xy$. Płaszczyzna styczna przecina wykres funkcji wzdłuż osi x i y . Po prawej: tzw. małpie siodło, wykres funkcji $g(x, y) = x^3 - 3y^2x$. Płaszczyzna styczna przecina wykres wzdłuż trzech prostych. Na obu rysunkach brzegi kolorowych pasów to poziomic (linie, na których funkcja ma stałą wartość).

Oczywiście nie zachodzi implikacja odwrotna. Funkcja $f(x, y) = xy$ ma pochodne cząstkowe $f_x(x, y) = y$ i $f_y(x, y) = x$, które znikają jednocześnie wtedy i tylko wtedy, gdy $x = y = 0$. Jednak w punkcie $(0, 0)$ funkcja f nie ma ani minimum, ani maksimum lokalnego (równego zero), gdyż w każdym otoczeniu tego punktu przyjmuje zarówno wartości dodatnie, jak i ujemne. Podobnie,

$$g(x, y) = x^3 - 3y^2x = x(x - y\sqrt{3})(x + y\sqrt{3}) \quad (2.20)$$

⁵Zauważmy: $k = \min(m, n)$ jest maksymalnym możliwym rzędem przekształcenia liniowego z $\mathbb{R}^n$ w $\mathbb{R}^m$.



Wykres funkcji $f(x, y) = x^2(1+y)^3 + y^2$. Zaznaczone linie to poziomice. W środku widoczne minimum lokalne. Dla $x = \text{const} > 0$ funkcja $f(\text{const}, y)$ jest wielomianem stopnia 3; proszę zwrócić uwagę na kształt przedniej krawędzi tego fragmentu wykresu f .

ma pochodne cząstkowe $g_x(x, y) = 3x^2 - 3y^2$ i $g_y(x, y) = 6xy$. Łatwo zauważyć, że jedynym punktem krytycznym tej funkcji jest $(0, 0)$, jednak w każdym otoczeniu tego punktu f przyjmuje zarówno wartości dodatnie, jak i ujemne.

Nawet dla $n = 2$ zachowanie funkcji różniczkowalnych $f: \mathbb{R}^n \rightarrow \mathbb{R}$ potrafi odbiegać od naiwnych oczekiwań, jakie Czytelnik mógłby mieć dzięki wcześniejszym jednowymiarowym intuicjom. Należy o tym pamiętać, szukając kresów zbioru wartości funkcji.

Przykład 2.28. Istnieją funkcje różniczkowalne $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ (o ciągłych pochodnych cząstkowych), które spełniają dwa warunki:

- f ma na $\mathbb{R}^2$ tylko jeden punkt krytyczny, w którym jest jej ekstremum lokalne;
- f nie jest ograniczona ani z góry, ani z dołu.

Taka funkcja w punkcie jedynego swego ekstremum lokalnego nie osiąga ani kresu dolnego, ani kresu górnego! Spójrzmy na dwa konkretne przykłady takich sytuacji.

Niech

$$f(x, y) = x^2(1 + y)^3 + y^2. \quad (2.21)$$

Funkcja f jest wielomianem, więc ma ciągłe pochodne cząstkowe. Wyznamy teraz

jej punkty krytyczne. Łatwo obliczamy

$$\frac{\partial f}{\partial x}(x, y) = 2x(1 + y)^3, \quad \frac{\partial f}{\partial y}(x, y) = 3x^2(1 + y)^2 + 2y.$$

Jeśli $\frac{\partial f}{\partial x} = 0$, to $x = 0$ lub $y = -1$. Jednak dla $y = -1$ jest $\frac{\partial f}{\partial y} \equiv -2 \neq 0$, tzn. f nie ma żadnych punktów krytycznych postaci $(a, -1)$. Natomiast $\frac{\partial f}{\partial y}(0, y) = 2y = 0$ dla $y = 0$. Dlatego jedynym punktem krytycznym f jest $(0, 0) \in \mathbb{R}^2$.

Nietrudno zauważyć, że dla $|y| < 1/2$ jest $2 \geq 1 + y \geq 1/2$ i dlatego

$$8(x^2 + y^2) \geq f(x, y) \geq \frac{1}{8}(x^2 + y^2), \quad |y| < \frac{1}{2}, \quad x \in \mathbb{R}.$$

Stąd wynika, że f ma w punkcie $(x, y) = (0, 0)$ minimum lokalne właściwe. Jednak f nie jest na $\mathbb{R}^2$ ograniczona ani z góry, ani z dołu, gdyż funkcja

$$h(y) = f(1, y) = (1 + y)^3 + y^2, \quad y \in \mathbb{R},$$

jest wielomianem trzeciego stopnia, a więc nie jest ograniczona ani z góry, ani z dołu.

Innego przykładu tego zjawiska dostarcza funkcja

$$g(x, y) = 3xe^y - x^3 - e^{3y}, \quad (2.22)$$

która jest nieograniczona z góry i z dołu na $\mathbb{R}^2$, gdyż $g(x, 0) = 3x - x^3 - 1$ jest wielomianem trzeciego stopnia zmiennej x . Mamy

$$\frac{\partial g}{\partial x}(x, y) = 3(e^y - x^2), \quad \frac{\partial g}{\partial y}(x, y) = 3e^y(x - e^{2y}).$$

Z równań $\frac{\partial g}{\partial x} = \frac{\partial g}{\partial y} = 0$ otrzymujemy $y = \ln x^2$ i $x = e^{2y} = x^4$, stąd zaś $x = 1$ i $y = 0$ (innych rozwiązań nie ma). Zachowanie funkcji g w otoczeniu punktu $(1, 0)$ można przeanalizować, korzystając ze wzoru Taylora.⁶ Wskażemy tylko kroki w rachunkach; Czytelnik zechce sprawdzić (nietrudne) szczegóły samodzielnie. Najpierw piszemy

$$g(x, y) = 3(x - 1)e^y - x^3 + 3e^y - e^{3y},$$

podstawiamy $x^3 = (1 + (x - 1))^3 = 1 + 3(x - 1) + 3(x - 1)^2 + (x - 1)^3$ i otrzymujemy

$$g(x, y) = 3e^y - e^{3y} - 1 + 3(x - 1)(e^y - 1) - 3(x - 1)^2 - (x - 1)^3.$$

Następnie wykorzystujemy rozwinięcie Taylora–Maclaurina funkcji wykładniczej. Prowadzi to do wyniku

$$g(x, y) = 1 - 3(x - 1)^2 - 3y^2 + 3y(x - 1) + \frac{3}{2}y^2(x - 1) + o(y^2) + o((x - 1)^2) \quad \text{dla } x \rightarrow 1, y \rightarrow 0.$$

⁶Nie znamy jeszcze wprowadzie wzoru Taylora dla funkcji wielu zmiennych, tu jednak nietrudno jest znaleźć najpierw rozwinięcie Taylora względem x (traktując y jako parametr), potem zaś skorzystać ze znanego rozwinięcia $e^y = 1 + y + y^2/2! + \dots$.

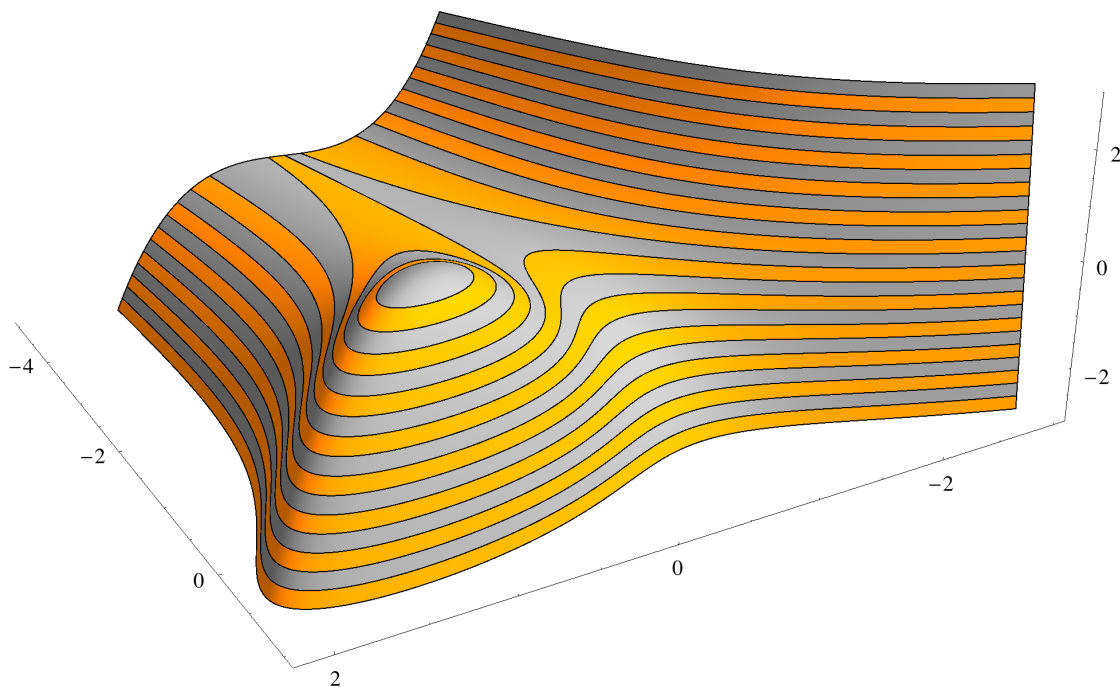
Jednak $|3y(x-1)| \leq \frac{3}{2}(y^2 + (x-1)^2)$, a z nierówności Younga $ab \leq a^p/p + b^q/q$ (gdzie $a, b \geq 0, p, q > 1$ i $\frac{1}{p} + \frac{1}{q} = 1$, patrz wykłady Analizy Matematycznej z I roku) zastosowanej dla $p = 3/2$ i $q = 3$ otrzymujemy

$$\frac{3}{2}|y^2(x-1)| \leq |y|^3 + \frac{1}{2}|x-1|^3 = o(y^2) + o((x-1)^2) \quad \text{dla } x \rightarrow 1, y \rightarrow 0.$$

Dlatego

$$g(x, y) \geq 1 - \frac{3}{2}(y^2 + (x-1)^2) + o(y^2) + o((x-1)^2) \quad \text{dla } x \rightarrow 1, y \rightarrow 0,$$

stąd zaś wynika, że g ma w punkcie $(1, 0)$ maksimum lokalne właściwe. Odpowiedni fragment wykresu funkcji g przedstawiony jest na rysunku. $\square$



Wykres funkcji $g(x, y) = 3xe^y - x^3 - e^{3y}$. Widoczny garb to jedyne maksimum lokalne tej funkcji. Innych punktów krytycznych g nie ma. Pomysł na prezentację wykresu zaczerpnięty z książki: Stan Wagon, *Mathematica in action*, wydanie 3, Springer Verlag 2010.

Przykład 2.29 (Nierówność między średnimi raz jeszcze). Udowodnimy ponownie nierówność między średnią arytmetyczną i geometryczną. Jak wcześniej w Przykładzie 1.48, założymy, że

$$x_1 + x_2 + \cdots + x_n = n, \quad x_i \geq 0.$$

Wykażemy, że $x_1 x_2 \cdots x_n \leq 1$ (przy czym równość zachodzi jedynie wtedy, gdy wszystkie x_i są równe).

Tym razem niech

$$f(x_1, x_2, \dots, x_{n-1}) = x_1 x_2 \dots x_{n-1} \underbrace{(n - (x_1 + x_2 + \dots + x_{n-1}))}_{=x_n \geq 0}$$

dla

$$\mathbf{x} = (x_1, \dots, x_{n-1}) \in K = \{\mathbf{x} \in \mathbb{R}^{n-1} : x_i \geq 0, \quad x_1 + \dots + x_{n-1} \leq n\}.$$

Zbiór K jest zwarty w $\mathbb{R}^{n-1}$, a funkcja f jest ciągła na K , zatem f osiąga swój kres górny. Na brzegu K jest $f \equiv 0$, a f przyjmuje wartości dodatnie, np. w punkcie $(1, \dots, 1)$, dlatego $\sup f = f(\mathbf{a})$ dla pewnego punktu $\mathbf{a}$ należącego do wnętrza zbioru K . W tym punkcie musi być $\text{grad } f(\mathbf{a}) = 0$.

Stosując wzór na pochodną iloczynu, łatwo sprawdzić, że wewnątrz K

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \frac{f(\mathbf{a})}{a_i} - a_1 \dots a_{n-1}, \quad i = 1, \dots, n-1.$$

Dlatego układ równań $\text{grad } f(\mathbf{a}) = 0$ jest równoważny innemu:

$$n - \sum_{j=1}^{n-1} a_j - a_i = 0, \quad i = 1, \dots, n-1.$$

Sumując te równania, otrzymujemy $n(n-1) = (n-1) \sum a_j + (a_1 + a_2 + \dots + a_{n-1})$, a stąd $\sum a_j = n-1$ i dlatego

$$n - \sum_{j=1}^{n-1} a_j - a_i = 1 - a_i,$$

ostatecznie więc $\text{grad } f(\mathbf{a}) = 0$ jedynie wtedy, gdy $a_i = 1$ dla wszystkich $i = 1, \dots, n-1$. Właśnie w punkcie $(1, \dots, 1)$ funkcja f przyjmuje więc swój kres górny, równy 1. $\square$

Opisaliśmy już geometryczną interpretację gradientu: jest to kierunek, w którym funkcja rośnie najszybciej. Okazuje się, że można powiedzieć więcej: przy nieznacznym dodatkowym założeniu funkcja „jest stała w kierunkach prostopadłych do gradientu”. Aby wyjaśnić to bliżej i ściślej, będziemy potrzebowali dwóch definicji.

Definicja 2.30 (poziomica funkcji). Poziomicą funkcji $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ nazywamy zbiór

$$M = \{\mathbf{x} \in \Omega : f(\mathbf{x}) = f(\mathbf{a})\},$$

gdzie $\mathbf{a} \in \Omega$ jest ustalonym punktem.

Innymi słowy, poziomica składa się z tych punktów, gdzie funkcja przybiera konkretną, ustaloną wartość (równą $f(\mathbf{a})$ dla danego punktu $\mathbf{a} \in \Omega$).

Definicja 2.31 (wektory styczne do zbioru w punkcie). Mówimy, że wektor $\mathbf{w} \in \mathbb{R}^n \setminus \{0\}$ jest styczny do zbioru $A \subset \mathbb{R}^n$ w punkcie $\mathbf{a} \in A$ (i piszemy $\mathbf{w} \in T_{\mathbf{a}}A$) wtedy i tylko wtedy, gdy istnieje ciąg punktów $\mathbf{x}_j \in A \setminus \{\mathbf{a}\}$ taki, że $\mathbf{x}_j \rightarrow \mathbf{a}$ dla $j \rightarrow \infty$, a ponadto

$$\frac{\mathbf{w}}{\|\mathbf{w}\|} = \lim_{j \rightarrow \infty} \frac{\mathbf{x}_j - \mathbf{a}}{\|\mathbf{x}_j - \mathbf{a}\|}. \quad (2.23)$$

Przyjmijmy także, że wektor $\mathbf{0} \in \mathbb{R}^n$ jest, dla dowolnych A i $\mathbf{a} \in A$, styczny do zbioru A w punkcie $\mathbf{a}$. Zbiór $T_{\mathbf{a}}A$ nazywamy przestrzenią styczną do A w punkcie $\mathbf{a}$ (lub, czasem, stożkiem stycznym do A w punkcie $\mathbf{a}$).

W literaturze matematycznej można spotkać różne (niekoniecznie równoważne) definicje wektora stycznego do zbioru. Przytoczona wyżej ma tę zaletę, że można się nią posługiwać, nie przyjmując *żadnych* dodatkowych założeń o zbiorze A .

Uwaga 2.32. Lewa strona wzoru (2.23) nie zmienia się, gdy wektor $\mathbf{w}$ mnożymy przez liczbę $t > 0$. Dlatego jeśli $\mathbf{w} \in T_{\mathbf{a}}A$, to $t \cdot \mathbf{w} \in T_{\mathbf{a}}A$ dla $t > 0$. To przestaje być prawdą dla $t < 0$: jeśli zbiór $A \subset \mathbb{R}^2$ jest wykresem funkcji $y = |x|^{1/2}$ i $\mathbf{a} = (0, 0)$, to nietrudno sprawdzić, że $\mathbf{w} = (0, 1) \in T_{\mathbf{a}}A$, natomiast $-\mathbf{w} \notin T_{\mathbf{a}}A$.

Uwaga 2.33. Przypuśćmy, że $\gamma: \mathbb{R} \supset I \rightarrow A \subset \mathbb{R}^n$ jest funkcją różniczkowalną zmiennej jednej zmiennej rzeczywistej $t \in I$, gdzie I jest jakimś przedziałem otwartym wokół zera. Wówczas wektor $\gamma'(0)$ jest styczny do zbioru A w punkcie $\mathbf{a} = \gamma(0)$. Jeśli $\gamma'(0) = 0$, to nie ma czego dowodzić. Przypuśćmy więc, że $\mathbf{w} \equiv \gamma'(0) \neq 0$. Z definicji pochodnej

$$\mathbf{w} = \gamma'(0) = \lim_{j \rightarrow \infty} \frac{\gamma(1/j) - \gamma(0)}{1/j}.$$

Ponieważ norma jest funkcją ciągłą, więc

$$0 \neq \|\mathbf{w}\| = \lim_{j \rightarrow \infty} \frac{\|\gamma(1/j) - \gamma(0)\|}{1/j}$$

(i wyrazy ciągu po prawej stronie są różne od zera dla dużych j). Dlatego

$$\frac{\mathbf{w}}{\|\mathbf{w}\|} = \lim_{j \rightarrow \infty} \left(\frac{\gamma(1/j) - \gamma(0)}{1/j} \cdot \frac{1/j}{\|\gamma(1/j) - \gamma(0)\|} \right) = \lim_{j \rightarrow \infty} \frac{\gamma(1/j) - \gamma(0)}{\|\gamma(1/j) - \gamma(0)\|},$$

a więc warunek (2.23) jest spełniony w punkcie $\mathbf{a} = \gamma(0)$ dla punktów $\mathbf{x}_j = \gamma(1/j)$.

Interpretacja fizyczna powyższego spostrzeżenia jest jasna: jeśli podróżujemy w zbiorze $A \subset \mathbb{R}^n$ (i położenie jest różniczkowalną funkcją czasu t), to wektor prędkości jest cały czas styczny do zbioru A .

Twierdzenie 2.34 (prostotałość gradientu do poziomicy). Załóżmy, że $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}$ jest różniczkowalna w punkcie $\mathbf{a} \in \Omega$ i ciągła na pewnej kuli $B(\mathbf{a}, r) \subset \Omega$, gdzie $r > 0$. Niech

$$A = \{x \in \Omega: f(\mathbf{x}) = f(\mathbf{a})\}.$$

Jeśli $\mathbf{v} := \text{grad } f(\mathbf{a}) \neq 0$, to następujące warunki są równoważne:

(i) $\mathbf{w} \in T_{\mathbf{a}}A$;

(ii) wektor $\mathbf{w}$ jest prostopadły do $\mathbf{v} = \text{grad } f(\mathbf{a})$, tzn. $\langle \mathbf{w}, \mathbf{v} \rangle = 0$.

Dowód. Jeśli $\mathbf{w} = 0$, to oba warunki są spełnione. Niech więc odtąd $\mathbf{w} \neq 0$. Załóżmy także, że $f(\mathbf{a}) = 0$; to nie zmniejsza ogólności rozważań, gdyż dodając do f stałą, nie zmieniamy gradientu.

Najpierw wykazemy, że (i) $\Rightarrow$ (ii). Niech $\mathbf{x}_j \in A \setminus \{\mathbf{a}\}$ będzie zbieżnym do $\mathbf{a}$ ciągiem punktów, dla którego zachodzi warunek (2.23). Ponieważ $\mathbf{x}_j \in A$, więc z definicji $f(\mathbf{x}_j) = 0$. Wobec różniczkowalności f w $\mathbf{a}$, mamy

$$\begin{aligned} 0 = f(\mathbf{x}_j) = f(\mathbf{a} + (\mathbf{x}_j - \mathbf{a})) &= f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x}_j - \mathbf{a}) + o(\|\mathbf{x}_j - \mathbf{a}\|) \\ &= \underbrace{f(\mathbf{a})}_{=0} + \langle \text{grad } f(\mathbf{a}), \mathbf{x}_j - \mathbf{a} \rangle + o(\|\mathbf{x}_j - \mathbf{a}\|) \\ &= \langle \text{grad } f(\mathbf{a}), \mathbf{x}_j - \mathbf{a} \rangle + o(\|\mathbf{x}_j - \mathbf{a}\|), \quad j \rightarrow \infty, \end{aligned}$$

a zatem

$$0 = \left\langle \operatorname{grad} f(\mathbf{a}), \frac{\mathbf{x}_j - \mathbf{a}}{\|\mathbf{x}_j - \mathbf{a}\|} \right\rangle + \frac{o(\|\mathbf{x}_j - \mathbf{a}\|)}{\|\mathbf{x}_j - \mathbf{a}\|}.$$

Zgodnie z (2.23), $(\mathbf{x}_j - \mathbf{a})/\|\mathbf{x}_j - \mathbf{a}\| \rightarrow \mathbf{w}/\|\mathbf{w}\|$, gdy $j \rightarrow \infty$. Zatem

$$0 = \lim_{j \rightarrow \infty} \left\langle \operatorname{grad} f(\mathbf{a}), \frac{\mathbf{x}_j - \mathbf{a}}{\|\mathbf{x}_j - \mathbf{a}\|} \right\rangle = \left\langle \operatorname{grad} f(\mathbf{a}), \frac{\mathbf{w}}{\|\mathbf{w}\|} \right\rangle = \langle \operatorname{grad} f(\mathbf{a}), \mathbf{w} \rangle.$$

Dowód implikacji (ii) $\Rightarrow$ (i) jest nieco trudniejszy. Niech $\mathbf{v} = \operatorname{grad} f(\mathbf{a}) \neq 0$. Ustalmy najpierw, posługując się wprost definicją różniczki i gradientu, liczbę $t > 0$ tak, aby mieć

$$f(\mathbf{a} + t \cdot \mathbf{v}) > 0 = f(\mathbf{a}) > f(\mathbf{a} - t \cdot \mathbf{v}) \quad \text{oraz} \quad t \cdot \|\mathbf{v}\| < \frac{r}{2};$$

można to zrobić, gdyż $Df(\mathbf{a})\mathbf{h} = \langle \operatorname{grad} f(\mathbf{a}), \mathbf{h} \rangle = \langle \mathbf{v}, \mathbf{h} \rangle$ i dla małych $|t|$ jest

$$f(\mathbf{a} + t\mathbf{v}) = f(\mathbf{a}) + Df(\mathbf{a})(t\mathbf{v}) + o(|t|) = Df(\mathbf{a})(t\mathbf{v}) + o(|t|) = t\|\mathbf{v}\|^2 + o(|t|),$$

a więc znak liczby $f(\mathbf{a} + t\mathbf{v})$ jest taki, jak znak t .

Dalszy ciąg dowodu polega na tym, by wybrać punkty $\mathbf{x}_j \in A$, dla których zachodzi warunek z definicji wektora stycznego. Kluczowy krok pod koniec rozumowania wykonujemy nie wprost; w dowodzie istotną rolę odgrywa ciągłość f w całym otoczeniu $\mathbf{a}$. (Czytelnik zechce wykonać rysunek, zakładając, że płaszczyzna kartki jest rozpięta na $\mathbf{v}$ i $\mathbf{w}$, i zaznaczać położenia kolejnych rozpatrywanych punktów).

Niech $j \in \mathbb{N}$, $0 < \frac{1}{j} < t$. Korzystając z ciągłości f w punktach $\mathbf{a} \pm \frac{1}{j}\mathbf{v}$, wybierzmy liczbę $\delta_j \in (0, 1/j)$ tak, aby

$$f(\mathbf{a} + \frac{1}{j}\mathbf{v} + s\mathbf{w}) > 0 \quad \text{dla wszystkich } |s| \leq \delta_j \quad (2.24)$$

i jednocześnie

$$f(\mathbf{a} - \frac{1}{j}\mathbf{v} + s\mathbf{w}) < 0 \quad \text{dla wszystkich } |s| \leq \delta_j. \quad (2.25)$$

Założymy też, że punkty $\mathbf{a} \pm \frac{1}{j}\mathbf{v} + \delta_j\mathbf{w} \in B(\mathbf{a}, r)$.

Dla każdego dostatecznie dużego $j \in \mathbb{N}$ funkcja

$$[-1/j, 1/j] \ni \theta \mapsto \phi_j(\theta) = f(\mathbf{a} + \theta\mathbf{v} + \delta_j\mathbf{w}) \in \mathbb{R}$$

jest ciągła i na końcach odcinka $[-1/j, 1/j]$ ma wartości różnych znaków, zatem istnieje punkt $\theta_j \in (-1/j, 1/j)$ taki, że $\phi_j(\theta_j) = 0$, lub równoważnie

$$\mathbf{x}_j \stackrel{\text{ozn.}}{=} \mathbf{a} + \theta_j\mathbf{v} + \delta_j\mathbf{w} \in A = \{f = 0\}.$$

Ponieważ $\delta_j \in (0, 1/j)$ i $|\theta_j| < 1/j$, więc $\mathbf{x}_j \neq \mathbf{a}$ i $\mathbf{x}_j \rightarrow \mathbf{a}$ dla $j \rightarrow \infty$. Korzystając ze zwartości sfery jednostkowej $\mathbb{S}^{n-1} = \{\mathbf{u} : \|\mathbf{u}\| = 1\}$, możemy założyć (przechodząc w razie potrzeby do odpowiedniego podciągu), że

$$\frac{\mathbf{x}_j - \mathbf{a}}{\|\mathbf{x}_j - \mathbf{a}\|} = \alpha_j\mathbf{v} + \beta_j\mathbf{w} \longrightarrow \mathbf{u}_0 = \alpha\mathbf{v} + \beta\mathbf{w} \in \mathbb{S}^{n-1}, \quad j \rightarrow \infty$$

gdzie współczynniki $\alpha_j \rightarrow \alpha$, $\beta_j \rightarrow \beta$ dla $j \rightarrow \infty$. Zauważmy, że $\beta_j = \delta_j/\|\mathbf{x}_j - \mathbf{a}\| > 0$. Dlatego $\beta = \lim \beta_j \geq 0$.

Z definicji, wektor $\mathbf{u}_0 \in T_{\mathbf{a}}A$. Gdyby $\alpha \neq 0$, to mielibyśmy

$$\langle \mathbf{u}_0, \mathbf{v} \rangle = \alpha \|\mathbf{v}\|^2 + \beta \langle \mathbf{w}, \mathbf{v} \rangle \stackrel{(i)}{=} \alpha \|\mathbf{v}\|^2 \neq 0.$$

Byłoby więc $\mathbf{u}_0 \in T_{\mathbf{a}}A$ i nie zachodziłby warunek (ii), co przeczyłoby udowodnionej już implikacji (i) $\Rightarrow$ (ii).

Dlatego musi być $\alpha = 0$. Ponieważ $\mathbf{u}_0 \neq 0$, więc $\beta \neq 0$, czyli $\beta > 0$. Wykazaliśmy zatem, że

$$\mathbf{u}_0 = \beta \mathbf{w} \in T_{\mathbf{a}}A, \quad \beta > 0.$$

Zgodnie z Uwagą 2.32, $\mathbf{w} \in T_{\mathbf{a}}A$. $\square$

Zadanie 2.35. Wskazać przykład funkcji $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ różniczkowalnej w $\mathbf{a} = (0, 0) \in \mathbb{R}^2$, $\text{grad } f(0, 0) = (1, 0)$, dla której nie zachodzi implikacja (ii) $\Rightarrow$ (i) w ostatnim twierdzeniu.

Wskazówka. Podzielić płaszczyznę na trzy obszary (dolną półpłaszczyznę $y < 0$ i dwie ćwiartki górnej półpłaszczyzny); na jednym z nich przyjąć $f(x, y) = x$, a na dwóch pozostałych $f(x, y) = x \pm y^2$, tak, aby zbiór $A = \{f = 0\}$ był półprostą domkniętą o końcu w punkcie $\mathbf{a}$.

Przykład 2.36 (styczna do okręgu). Niech $f(x, y) = x^2 + y^2 - R^2$, gdzie $R > 0$. Zbiór $\{f = 0\}$ to okrąg γ_R o promieniu $R > 0$. Gradient funkcji f w punkcie $(x, y) \in \gamma_R$ to wektor $2(x, y)$, współliniowy z promieniem okręgu γ_R , prowadzącym do punktu (x, y) . Twierdzenie 2.34 implikuje więc, że styczna do okręgu jest prostopadła do promienia, poprowadzonego w punkcie styczności.

Zadanie 2.37. Proszę wykazać, że jeśli P jest płaszczyzną styczną (w sensie Definicji 2.24) do wykresu funkcji różniczkowalnej $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}$ w punkcie $(\mathbf{a}, f(\mathbf{a}))$, to dla każdego punktu $(\mathbf{x}, x_{n+1}) \in P$ wektor

$$\mathbf{v} = (\mathbf{x}, x_{n+1}) - (\mathbf{a}, f(\mathbf{a})) \in \mathbb{R}^{n+1}$$

jest styczny (w sensie Definicji 2.31) do wykresu funkcji f . Można posłużyć się Twierdzeniem 2.34, tzn. przedstawić wykres funkcji n zmiennych jako poziomice pewnej funkcji $n + 1$ zmiennych.

2.4 Twierdzenie o wartości średniej

Definicja 2.38 (Funkcje klasy C^1). Niech Ω będzie zbiorem otwartym w $\mathbb{R}^n$. Mówimy, że $f \in C^1(\Omega, \mathbb{R}^m)$, jeśli $f: \Omega \rightarrow \mathbb{R}^m$ ma na Ω ciągłe pochodne cząstkowe $\frac{\partial f}{\partial x_j}$, $j = 1, 2, \dots, n$.

Uwaga 2.39. Jeśli $f \in C^1(\Omega, \mathbb{R}^m)$, to wobec twierdzenia 2.16 f jest różniczkowalna w każdym punkcie Ω , a ponadto odwzorowanie

$$\Omega \ni \mathbf{x} \longmapsto Df(\mathbf{x}) \in L(\mathbb{R}^n, \mathbb{R}^m) \simeq M_{m \times n}$$

jest ciągłe (tu korzystamy ze Stwierdzenia 1.35). Ponieważ z istnienia różniczki wynika ciągłość funkcji, więc funkcje klasy C^1 są ciągłe.

Na odwrót, jeśli założymy, że $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ jest ciągła i ma ciągłą różniczkę $Df: \mathbb{R}^n \supset \Omega \rightarrow L(\mathbb{R}^n, \mathbb{R}^m)$, to oczywiście spełnione są warunki Definicji 2.38. To wynika ze Stwierdzenia 1.35.

Podamy teraz odpowiednik twierdzenia Lagrange’a o wartości średniej dla funkcji wielu zmiennych.

Twierdzenie 2.40 (o wartości średniej). Niech $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$. Załóżmy, że dla pewnych $\mathbf{x}, \mathbf{y} \in \Omega$ odcinek $[\mathbf{x}, \mathbf{y}] \subset \Omega$ i f jest różniczkowalna we wszystkich punktach tego odcinka. Wówczas

$$\|f(\mathbf{y}) - f(\mathbf{x})\| \leq \|\mathbf{y} - \mathbf{x}\| \cdot \sup_{\theta \in [0,1]} \|Df(\mathbf{x} + \theta(\mathbf{y} - \mathbf{x}))\|. \quad (2.26)$$

Geometryczny sens tego twierdzenia jest następujący: jeśli różniczka $Df(\mathbf{p})$ zwiększa długość wektorów co najwyżej k -krotnie, to norma przyrostu funkcji wzdłuż odcinka $[\mathbf{x}, \mathbf{y}]$, zawartego w dziedzinie, nie przekracza iloczynu liczby k i długości tego odcinka.

Dowód. Wprowadźmy dwie funkcje pomocnicze,

$$g(t) = f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f(\mathbf{x}), \quad t \in [0, 1], \quad (2.27)$$

$$\Phi(t) = \langle g(1), g(t) \rangle, \quad t \in [0, 1]. \quad (2.28)$$

Dla $t \in [0, 1]$ punkt $\mathbf{x} + t(\mathbf{y} - \mathbf{x}) \in [\mathbf{x}, \mathbf{y}]$, zatem g jest funkcją różniczkowalną zmiennej t (jako złożenie funkcji różniczkowalnych). Podobnie, Φ jest funkcją różniczkowalną. Ponadto, $\Phi(1) = \|g(1)\|^2 = \|f(\mathbf{y}) - f(\mathbf{x})\|^2$ i $\Phi(0) = 0$. Wyraźmy przyrost funkcji Φ , stosując twierdzenia Lagrange’a dla funkcji jednej zmiennej, a następnie obliczmy pochodną $\Phi'(\theta)$, korzystając z twierdzeń o pochodnej iloczynu i różniczce złożenia. Otrzymamy

$$\begin{aligned} \|f(\mathbf{y}) - f(\mathbf{x})\|^2 &= \Phi(1) - \Phi(0) \\ &= \Phi'(\theta) \quad \text{dla pewnego } \theta \in (0, 1) \\ &= \langle g(1), g'(\theta) \rangle \\ &= \langle f(\mathbf{y}) - f(\mathbf{x}), Df(\mathbf{x} + \theta(\mathbf{y} - \mathbf{x})) \cdot (\mathbf{y} - \mathbf{x}) \rangle \\ &\leq \|f(\mathbf{y}) - f(\mathbf{x})\| \cdot \|Df(\mathbf{x} + \theta(\mathbf{y} - \mathbf{x})) \cdot (\mathbf{y} - \mathbf{x})\| \\ &\leq \|f(\mathbf{y}) - f(\mathbf{x})\| \cdot \|Df(\mathbf{x} + \theta(\mathbf{y} - \mathbf{x}))\| \cdot \|\mathbf{y} - \mathbf{x}\|. \end{aligned}$$

(Pierwsza nierówność to nierówność Schwarz’a; druga wynika z definicji normy przekształcenia liniowego). Jeśli $\|f(\mathbf{y}) - f(\mathbf{x})\| = 0$, to teza jest oczywista. W przeciwnym przypadku dzielimy otrzymaną nierówność przez $\|f(\mathbf{y}) - f(\mathbf{x})\| > 0$ i biorąc kres górny prawej strony względem $\theta \in [0, 1]$ otrzymujemy (2.26). $\square$

Podamy jeszcze drugi dowód tego ważnego twierdzenia. Wymaga on wprowadzić nieco mocniejszych założeń, jednak użyty w nim sposób postępowania jest bardzo naturalny i często wykorzystywany w wielu działach analizy.

DRUGI DOWÓD TWIERDZENIA O WARTOŚCI ŚREDNIEJ. Niech g nadal oznacza funkcję pomocniczą, określoną wzorem (2.27). Założymy dodatkowo, że $f \in C^1(\Omega, \mathbb{R}^m)$. Wtedy $g'(t)$ jest funkcją ciągłą. Wyrazimy przyrost f , tzn. przyrost g na odcinku $[0, 1]$, całkując g' .

Uwaga. Wartościami g' są wektory z przestrzeni $\mathbb{R}^m$. Przyjmujemy naturalną umowę: całka oznaczona $\int_a^b h(t) dt$ z (ciągłej) funkcji wektorowej $h = (h_1, \dots, h_m): [a, b] \rightarrow \mathbb{R}^m$ jest wektorem o współrzędnych $\int_a^b h_j(t) dt$. Zachodzi wtedy nierówność

$$\left\| \int_a^b h(t) dt \right\| \leq \int_a^b \|h(t)\| dt, \quad (2.29)$$

którą Czytelnik może udowodnić samodzielnie.⁷

Mamy

$$f(\mathbf{y}) - f(\mathbf{x}) = g(1) - g(0) = \int_0^1 g'(t) dt.$$

Korzystając z nierówności 2.29 i obliczając g' (jak w pierwszym dowodzie), otrzymujemy

$$\begin{aligned} \|f(\mathbf{y}) - f(\mathbf{x})\| &= \left\| \int_0^1 g'(t) dt \right\| \\ &\leq \int_0^1 \|g'(t)\| dt \\ &= \int_0^1 \|Df(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) \cdot (\mathbf{y} - \mathbf{x})\| dt \\ &\leq \int_0^1 \|Df(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))\| \cdot \|\mathbf{y} - \mathbf{x}\| dt \\ &= \|\mathbf{y} - \mathbf{x}\| \cdot \int_0^1 \|Df(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))\| dt \\ &\leq \|\mathbf{y} - \mathbf{x}\| \cdot \sup_{t \in [0,1]} \|Df(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))\|. \end{aligned}$$

(Pisząc ostatnią nierówność, oszacowaliśmy całkę przez iloczyn kresu górnego funkcji i długości odcinka). $\square$

2.5 Pochodne cząstkowe wyższych rzędów i wzór Taylora

Zajmiemy się teraz określeniem pochodnych cząstkowych rzędu wyższego niż pierwszy, różniczek wyższych rzędów, oraz uogólnieniem wzoru Taylora na funkcje wielu zmiennych. Podobnie jak w przypadku funkcji jednej zmiennej, wzór Taylora pozwala znajdować najlepsze przybliżenia wielomianowe funkcji, a także prowadzi do warunków dostatecznych, gwarantujących, że w punkcie krytycznym funkcja ma ekstremum lokalne.

Jak można się spodziewać, pochodne cząstkowe wyższych rzędów definiuje się indukcyjnie.

Definicja 2.41 (pochodne cząstkowe drugiego rzędu). Załóżmy, że funkcja $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ ma na Ω pochodną cząstkową $\frac{\partial f}{\partial x_i}: \Omega \rightarrow \mathbb{R}^m$. Jeśli funkcja $D_i f = \frac{\partial f}{\partial x_i}$ ma w punkcie $\mathbf{a} \in \Omega$ pochodną cząstkową względem x_j , to przyjmujemy

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) = \frac{\partial}{\partial x_j} \left[\frac{\partial f}{\partial x_i} \right](\mathbf{a}).$$

Będziemy też używać innych oznaczeń:

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) = D_j D_i f(\mathbf{a}) = f_{x_i x_j}(\mathbf{a}).$$

⁷Wskazówka: Całkę można przybliżać sumami Riemanna, a dla sum Riemanna mamy tu do czynienia po prostu z nierównością trójkąta dla normy.

Uwaga. Posługując się oznaczeniami $f_{x_i x_j}$, przestrzegamy naturalnej mnemotechnicznej konwencji: *najpierw* różniczkujemy względem tej zmiennej, która jest zapisana *bliżej* f .

Pochodne cząstkowe wyższych rzędów definiuje się analogicznie, np.

$$\frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i}(\mathbf{a}) = \frac{\partial}{\partial x_k} \left[\frac{\partial^2 f}{\partial x_j \partial x_i} \right](\mathbf{a});$$

stosując inne oznaczenia, napisalibyśmy $D_k D_j D_i f(\mathbf{a}) = D_k(D_j D_i f)(\mathbf{a})$ oraz

$$(f_{x_i x_j})_{x_k}(\mathbf{a}) = f_{x_i x_j x_k}(\mathbf{a}).$$

Stosowanie wszelkich oznaczeń tego typu dla pochodnych cząstkowych rzędu wyższego niż drugi jest w praktyce dość niewygodne. Dlatego później poznamy jeszcze inną, wygodną i bardzo skrótową konwencję notacyjną. Najpierw jednak omówimy najważniejsze własności pochodnych cząstkowych drugiego rzędu.

2.5.1 Przykład Peano i twierdzenie Schwarz'a o równości pochodnych mieszanych

Bardzo naturalne jest pytanie: czy, wprowadzając oznaczenia pochodnych cząstkowych drugiego rzędu, trzeba rzeczywiście koniecznie odróżniać $f_{x_i x_j}$ od $f_{x_j x_i}$? Czytelnik, oswojony już nieco z przykładami patologicznego zachowania funkcji wielu zmiennych, może spodziewać się, że odpowiedź jest twierdząca.

Przykład 2.42 (G. Peano, 1884). Niech

$$f(x, y) = \begin{cases} xy \frac{x^2 - y^2}{x^2 + y^2}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$$

W punktach $\mathbb{R}^2 \setminus \{(0, 0)\}$ funkcja f jest różniczkowalna. Ponadto,

$$|f(x, y)| \leq |xy| \leq \frac{1}{2}(x^2 + y^2);$$

korzystając z tej nierówności, sprawdza się łatwo (wprost z definicji różniczki), że $Df(0, 0) = (0, 0) \in L(\mathbb{R}^2, \mathbb{R})$. Obliczymy teraz pochodne mieszane $f_{xy}(0, 0)$ i $f_{yx}(0, 0)$.

Mamy

$$f_{xy}(0, 0) = (f_x)_y(0, 0) = \lim_{y \rightarrow 0} \frac{f_x(0, y) - f_x(0, 0)}{y} = \lim_{y \rightarrow 0} \frac{f_x(0, y)}{y},$$

gdyż $f_y(0, 0) = 0$ (pamiętajmy: różniczka f znika w zerze). Wartość $f_x(0, y)$ obliczamy, posługując się definicją f ; aby nie wykonywać długich rachunków, zauważmy, że $f(x) = x^2$ ma w zerze pochodną 0 i dlatego

$$f_x(0, y) = y \cdot \left(\frac{x^2 - y^2}{x^2 + y^2} \right)_{|x=0} = -y.$$

Zatem $f_{xy}(0, 0) = -1$. Zamieniając x, y rolami, otrzymujemy w ten sam sposób $f_y(x, 0) = x$ i $f_{yx}(0, 0) = 1$. Jest więc $f_{xy}(0, 0) \neq f_{yx}(0, 0)$.

Podobny (bardziej skomplikowany) przykład podał H.A. Schwarz nieco wcześniej. Okazuje się jednak, że takie zachowanie jest wykluczone wówczas, gdy pochodne mieszane są ciągłe.

Twierdzenie 2.43 (Schwarza o równości pochodnych mieszanych). *Ustalmy $i, j \in 1, \dots, n$. Jeśli funkcja $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ jest klasy C^1 i ma na Ω ciągłą pochodną $f_{x_i x_j}$, to pochodna $f_{x_j x_i}$ istnieje we wszystkich punktach Ω i $f_{x_j x_i} = f_{x_i x_j}$.*

Dowód poprzedzimy kilkoma lematami.

Lemat 2.44. *Niech $Q = [a, b] \times [c, d] \subset \mathbb{R}^2$. Jeśli funkcja $f: Q \rightarrow \mathbb{R}^m$ zmiennych $(x, y) \in Q$ jest ciągła i ma ciągłą pochodną cząstkową $f_y: Q \rightarrow \mathbb{R}^m$, to*

$$\Phi(y) = \int_a^b f(x, y) dx, \quad y \in [c, d],$$

jest funkcją różniczkowalną i zachodzi wzór

$$\Phi'(y) = \int_a^b f_y(x, y) dx. \quad (2.30)$$

Dowód. Z definicji całki oznaczonej,

$$f(x, y + h) - f(x, y) = \int_0^1 \frac{d}{ds} f(x, y + sh) ds = h \int_0^1 f_y(x, y + sh) ds.$$

Dlatego iloraz różnicowy funkcji Φ jest równy

$$\begin{aligned} \Delta_h \Phi(y) &\stackrel{\text{ozn.}}{=} \frac{\Phi(y + h) - \Phi(y)}{h} \\ &= \frac{1}{h} \int_a^b (f(x, y + h) - f(x, y)) dx = \int_a^b \left(\int_0^1 f_y(x, y + sh) ds \right) dx. \end{aligned}$$

Oznaczmy literą I całkę po prawej stronie wzoru (2.30). Ponieważ $f_y(x, y) = \int_0^1 f_y(x, y) ds$, więc

$$\begin{aligned} \|\Delta_h \Phi(y) - I\| &= \left\| \int_a^b \left(\int_0^1 f_y(x, y + sh) ds \right) dx - \int_a^b \left(\int_0^1 f_y(x, y) ds \right) dx \right\| \\ &= \left\| \int_a^b \left(\int_0^1 (f_y(x, y + sh) - f_y(x, y)) ds \right) dx \right\| \\ &\leq \int_a^b \left(\int_0^1 \|f_y(x, y + sh) - f_y(x, y)\| ds \right) dx. \end{aligned} \quad (2.31)$$

Funkcja f_y jest ciągła na zbiorze zwartym $Q = [a, b] \times [c, d]$, a więc jest jednostajnie ciągła na Q . Ustalmy $\varepsilon > 0$ i dobierzmy $\delta > 0$ tak, aby $\|f_y(\mathbf{p}_1) - f_y(\mathbf{p}_2)\| < \varepsilon/(b - a)$ dla $\|\mathbf{p}_1 - \mathbf{p}_2\| < \delta$. Wówczas, dla $|h| < \delta$, funkcja podcałkowa w (2.31) jest w każdym punkcie mniejsza od $\varepsilon/(b - a)$ i otrzymujemy

$$\|\Delta_h \Phi(y) - I\| < \int_a^b \left(\int_0^1 \frac{\varepsilon}{b - a} ds \right) dx = (b - a) \cdot \frac{\varepsilon}{b - a} = \varepsilon.$$

Wprost z definicji granicy, $\Delta_h \Phi(y) \rightarrow I$ dla $h \rightarrow 0$, tzn. istotnie zachodzi wzór (2.30). $\square$

Uwaga 2.45. Oczywiście, wzór analogiczny do (2.30) zachodzi także wtedy, gdy zmienne zamienimy rolami.

Lemat 2.46. Niech $\Omega \subset \mathbb{R}^2$ będzie zbiorem otwartym i niech $Q = [a, b] \times [c, d] \subset \Omega$. Jeśli $f \in C^1(\Omega, \mathbb{R}^m)$ ma pochodną cząstkową f_{yx} ciągłą na Q , to wówczas f_{xy} istnieje w punktach prostokąta Q i $f_{xy}(x, y) = f_{yx}(x, y)$ dla $(x, y) \in Q$.

Dowód. Załóżmy najpierw, że $m = 1$. Niech $(x, y), (x, y_0) \in Q$. Napiszmy

$$f(x, y) = f(x, y_0) + \int_{y_0}^y f_y(x, z) dz = f(x, y_0) + \int_{y_0}^y g(x, z) dz,$$

gdzie funkcja $g = f_y$ jest ciągła na Ω i ma pochodną $g_x = f_{yx}$ ciągłą na Q . Różniczkując powyższy wzór względem x i stosując poprzedni lemat do g , otrzymujemy

$$f_x(x, y) = f_x(x, y_0) + \int_{y_0}^y g_x(x, z) dz = f_x(x, y_0) + \int_{y_0}^y f_{yx}(x, z) dz.$$

Zatem

$$\frac{f_x(x, y) - f_x(x, y_0)}{y - y_0} = \frac{1}{y - y_0} \int_{y_0}^y f_{yx}(x, z) dz.$$

Dla $y \rightarrow y_0$ lewa strona dąży do $f_{xy}(x, y_0)$, prawa zaś do wartości funkcji podcałkowej w punkcie y_0 , tzn. do $f_{yx}(x, y_0)$ (tu ponownie korzystamy z ciągłości f_{yx} i z twierdzenia o wartości średniej dla całki). Dowód jest zakończony w przypadku $m = 1$.

Gdy $m > 1$, to z pierwszej części dowodu wynika, że wszystkie współrzędne pochodnej f_{xy} są dobrze określone i równe odpowiednim współrzędnym f_{yx} . $\square$

DOWÓD TWIERDZENIA 2.43. Dla zbiorów $\Omega \subset \mathbb{R}^2$ twierdzenie wynika natychmiast z ostatniego lematu. Jeśli $n > 2$, to zauważmy, że aby określić pochodne $f_{x_i x_j}$ oraz $f_{x_j x_i}$ w punkcie $\mathbf{a} \in \Omega \subset \mathbb{R}^n$, wystarczy znać wartości f jedynie na dwuwymiarowej płaszczyźnie afinicznej $\mathbf{a} + \text{span}(\mathbf{e}_i, \mathbf{e}_j)$. Stosując Lemat 2.46 na przecięciach zbioru Ω z takimi płaszczyznami, łatwo otrzymujemy tezę. $\square$

2.5.2 Druga różniczka

Zacznijmy od objaśnienia, jakim obiektem matematycznym miałyby być druga różniczka $D^2 f$ funkcji wielu zmiennych. Przypuśćmy, że $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ jest różniczkowalna na Ω . Dla ustalonego $\mathbf{x} \in \Omega$ jej różniczka $Df(\mathbf{x})$ jest elementem przestrzeni $L(\mathbb{R}^n, \mathbb{R}^m)$, którą, ustaliwszy bazy w $\mathbb{R}^n$ i $\mathbb{R}^m$, można utożsamiać z $M_{m \times n}$ lub $\mathbb{R}^{mn}$. Inaczej mówiąc,

$$Df: \mathbb{R}^n \supset \Omega \ni \mathbf{x} \mapsto Df(\mathbf{x}) \in L(\mathbb{R}^n, \mathbb{R}^m) \simeq \mathbb{R}^{mn}.$$

Naturalnie byłoby określić drugą różniczkę $D^2 f$ jako $D^2 f = D(Df)$ (wszędzie tam, gdzie Df sama jest funkcją różniczkowalną). Zgodnie z definicją różniczki, powinno wtedy być

$$D^2 f = D(Df): \mathbb{R}^n \supset \Omega \ni \mathbf{x} \mapsto D^2 f(\mathbf{x}) \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)).$$

tzn. $D^2 f(\mathbf{x}) = D(Df)(\mathbf{x})$ powinna być, dla ustalonego $\mathbf{x}$, przekształceniem liniowym z $\mathbb{R}^n$ w przestrzeń, do której należą wartości różniczkowanej funkcji Df , tzn. $L(\mathbb{R}^n, \mathbb{R}^m)$.

Brzmi to zawile i widać od razu, że definiowanie różniczek wyższych rzędów prowadziłoby do coraz dłuższych napisów (i coraz bardziej skomplikowanych przestrzeni liniowych). Dlatego korzysta się z naturalnego izomorfizmu

$$L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)) \equiv L(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^m)$$

między przestrzenią $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$ i przestrzenią $L(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^m)$ przekształceń *dwuliniowych* $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^m$. Jest to izomorfizm *kanoniczny*, tzn. zdefiniowany bez odwoływania się do układu współrzędnych, za pomocą wzoru

$$L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)) \ni F \longmapsto B_F \in L(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^m) \quad (2.32)$$

gdzie

$$B_F(\mathbf{u}, \mathbf{v}) = F(\mathbf{u})\mathbf{v} \quad \text{dla } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n. \quad (2.33)$$

Sprawdzenie, że to rzeczywiście izomorfizm, jest łatwym ćwiczeniem.

Definicja 2.47. Przypuśćmy, że różniczka $Df: \Omega \rightarrow L(\mathbb{R}^n, \mathbb{R}^m)$ funkcji $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ jest określona w każdym punkcie zbioru otwartego $\Omega \subset \mathbb{R}^n$. Jeśli funkcja

$$g = Df: \Omega \rightarrow L(\mathbb{R}^n, \mathbb{R}^m) \simeq M_{m \times n}$$

jest różniczkowalna w punkcie $\mathbf{a} \in \Omega$, to przekształcenie dwuliniowe

$$D^2f(\mathbf{a}) = D(Df)(\mathbf{a}) = Dg(\mathbf{a}) \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)) \equiv L(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^m)$$

nazywamy drugą różniczką funkcji f w punkcie $\mathbf{a}$.

Uwaga 2.48. Dla $m = 1$ różniczka $Df(\mathbf{x}) \in L(\mathbb{R}^n, \mathbb{R}) = (\mathbb{R}^n)^* \simeq \mathbb{R}^n$ ma jako współrzędne pochodne cząstkowe $f_{x_i}(\mathbf{x})$. Dlatego przekształcenie dwuliniowe $D^2f(\mathbf{x})$ ma, w standardowej bazie $\mathbb{R}^n$, macierz, której wyrazami są pochodne cząstkowe drugiego rzędu funkcji f ; ponadto,

$$D^2f(\mathbf{x})(\mathbf{v}, \mathbf{w}) = \mathbf{v}^T D^2f(\mathbf{x})\mathbf{w},$$

gdzie lewą stronę interpretujemy jako wartość przekształcenia dwuliniowego dla pary wektorów $\mathbf{v}, \mathbf{w}$, prawą zaś jako wynik mnożenia trzech macierzy, o rozmiarach (odpowiednio) $1 \times n$, $n \times n$ i $n \times 1$.

Z twierdzenia Schwarza o równości pochodnych mieszanych wynika natychmiast, że jeśli $f: \Omega \rightarrow \mathbb{R}$ ma ciągłe pochodne cząstkowe do rzędu 2 włącznie, to $Df: \Omega \rightarrow (\mathbb{R}^n)^*$ jest funkcją różniczkowalną i dla każdego $\mathbf{a} \in \Omega$ macierz przekształcenia dwuliniowego $D^2f(\mathbf{a})$ jest macierzą symetryczną, gdyż $f_{x_i x_j} = f_{x_j x_i}$. Okazuje się jednak, że tak jest również wtedy, gdy $D^2f(\mathbf{a})$ po prostu istnieje; nie trzeba zakładać ciągłości pochodnych mieszanych w pewnym otoczeniu punktu $\mathbf{a}$.

Twierdzenie 2.49 (Schwarza o symetrii drugiej różniczki). Załóżmy, że różniczka $Df: \Omega \rightarrow L(\mathbb{R}^n, \mathbb{R}^m)$ funkcji $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ jest określona w każdym punkcie zbioru otwartego Ω . Jeśli $D^2f(\mathbf{a}) \in L(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^m)$ istnieje dla pewnego $\mathbf{a} \in \Omega$, to jest przekształceniem dwuliniowym symetrycznym, tzn.

$$D^2f(\mathbf{a})(\mathbf{v}, \mathbf{w}) = D^2f(\mathbf{a})(\mathbf{w}, \mathbf{v}) \quad \text{dla wszystkich } \mathbf{v}, \mathbf{w} \in \mathbb{R}^n.$$

Dowód. Ustalmy $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. Niech $\max(\|\mathbf{v}\|, \|\mathbf{w}\|) = M$. Rozpatrzmy funkcję pomocniczą

$$\phi(s, t) = f(\mathbf{a} + t\mathbf{w} + s\mathbf{v}) - f(\mathbf{a} + t\mathbf{w}) - f(\mathbf{a} + s\mathbf{v}) + f(\mathbf{a}) - ts D^2 f(\mathbf{a})(\mathbf{w}, \mathbf{v}),$$

określoną dla s, t w pewnym otoczeniu zera. Mamy $\phi(0, t) = 0$; z twierdzenia o wartości średniej wynika, że

$$\begin{aligned} \|\phi(s, t)\| &= \|\phi(s, t) - \phi(0, t)\| \leq |s| \sup_{\sigma \in [0, s]} \|\phi_s(\sigma, t)\| \\ &= |s| \sup_{\sigma \in [0, s]} \left\| \left(Df(\mathbf{a} + t\mathbf{w} + \sigma\mathbf{v}) - Df(\mathbf{a} + \sigma\mathbf{v}) \right) \cdot \mathbf{v} - t D^2 f(\mathbf{a})(\mathbf{w}, \mathbf{v}) \right\|. \end{aligned} \quad (2.34)$$

Ponieważ $D^2 f(\mathbf{a})$ istnieje, więc $Df(\mathbf{a} + \mathbf{h}) = Df(\mathbf{a}) + D^2 f(\mathbf{a})\mathbf{h} + \|\mathbf{h}\|r(\mathbf{h})$, gdzie $r(\mathbf{h}) \rightarrow 0$ dla $\mathbf{h} \rightarrow 0$ (patrz Stwierdzenie 2.9). Podstawiając w tej równości wektory $\mathbf{h}_1 = t\mathbf{w} + \sigma\mathbf{v}$ i $\mathbf{h}_2 = \sigma\mathbf{v}$, otrzymujemy

$$\begin{aligned} Df(\mathbf{a} + t\mathbf{w} + \sigma\mathbf{v}) &= Df(\mathbf{a}) + D^2 f(\mathbf{a}) \cdot (t\mathbf{w} + \sigma\mathbf{v}) + \|\mathbf{h}_1\|r(\mathbf{h}_1), \\ Df(\mathbf{a} + \sigma\mathbf{v}) &= Df(\mathbf{a}) + D^2 f(\mathbf{a}) \cdot \sigma\mathbf{v} + \|\mathbf{h}_2\|r(\mathbf{h}_2). \end{aligned}$$

Odejmując oba wzory stronami i pamiętając o izomorfizmie przestrzeni $L(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^m)$ oraz $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$, sprawdzamy, że

$$\begin{aligned} \left(Df(\mathbf{a} + t\mathbf{w} + \sigma\mathbf{v}) - Df(\mathbf{a} + \sigma\mathbf{v}) \right) \cdot \mathbf{v} - t D^2 f(\mathbf{a})(\mathbf{w}, \mathbf{v}) \\ = (\|\mathbf{h}_1\|r(\mathbf{h}_1) - \|\mathbf{h}_2\|r(\mathbf{h}_2)) \cdot \mathbf{v}. \end{aligned} \quad (2.35)$$

Niech odtąd $s = t$. Wtedy $|\sigma| \leq |s| = |t|$, co daje oszacowania $\|\mathbf{h}_i\| \leq 2M|t|$ dla $i = 1, 2$ oraz

$$\|r(\mathbf{h}_i)\| \leq \sup_{\|\mathbf{h}\| \leq 2Mt} \|r(\mathbf{h})\| = o(1), \quad t \rightarrow 0. \quad (2.36)$$

Korzystając ze wzorów (2.35)–(2.36), przepisujemy dla $s = t$ nierówność (2.34) w postaci

$$\phi(t, t) \leq |t| \cdot \left(2Mt \cdot 2 \sup_{\|\mathbf{h}\| \leq 2Mt} \|r(\mathbf{h})\| \right) \cdot \|\mathbf{v}\| = t^2 o(1), \quad t \rightarrow 0.$$

Innymi słowy, $\phi(t, t)/t^2 \rightarrow 0$ dla $t \rightarrow 0$, lub równoważnie

$$D^2 f(\mathbf{w}, \mathbf{v}) = \lim_{t \rightarrow 0} \frac{f(\mathbf{a} + t\mathbf{w} + t\mathbf{v}) - f(\mathbf{a} + t\mathbf{w}) - f(\mathbf{a} + t\mathbf{v}) + f(\mathbf{a})}{t^2}. \quad (2.37)$$

Prawa strona wzoru (2.37) nie zmienia się, gdy zamienimy wektory $\mathbf{w}, \mathbf{v}$ rolami. Dlatego lewa strona też musi być symetryczną funkcją $\mathbf{w}$ i $\mathbf{v}$, tzn. $D^2 f(\mathbf{w}, \mathbf{v}) = D^2 f(\mathbf{v}, \mathbf{w})$. $\square$

Przykład 2.50. a) Jeśli $f(\mathbf{x}) = A\mathbf{x}$, gdzie $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ jest ustalonym przekształceniem liniowym, to $Df(\mathbf{x}) = A$ jest przekształceniem stałym i dlatego $D^2 f(\mathbf{x}) = \mathbf{0} \in L^2(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^m)$.

b) Jeśli

$$f(\mathbf{x}) = \langle A\mathbf{x}, \mathbf{x} \rangle \quad \text{dla } \mathbf{x} \in \mathbb{R}^n,$$

gdzie A jest macierzą $n \times n$, to ze wzoru na pochodną ‘iloczynu’ otrzymujemy

$$Df(\mathbf{x})\mathbf{h} = \langle A\mathbf{h}, \mathbf{x} \rangle + \langle A\mathbf{x}, \mathbf{h} \rangle = \langle A^T \mathbf{x}, \mathbf{h} \rangle + \langle A\mathbf{x}, \mathbf{h} \rangle = \langle (A + A^T)\mathbf{x}, \mathbf{h} \rangle,$$

co oznacza, że $Df(\mathbf{x}) = (A + A^T)\mathbf{x}$ dla wszystkich $\mathbf{x} \in \mathbb{R}^n$.⁸ Zatem, Df zależy od $\mathbf{x}$ liniowo i mamy $D^2 f(\mathbf{x}) = A + A^T$. Jeśli $A = A^T$, tzn. macierz A jest symetryczna, to $D^2 f(\mathbf{x}) = 2A$.

⁸Pisząc wzór $Df(\mathbf{x}) = (A + A^T)\mathbf{x}$, utożsamiamy funkcjonal liniowy $Df(\mathbf{x})$ z wektorem $(A + A^T)\mathbf{x}$.

2.5.3 Różniczki wyższych rzędów

Różniczki wyższych rzędów definiuje się indukcyjnie, wzorem

$$D^k(f)(\mathbf{a}) = D(D^{k-1}f)(\mathbf{a}), \quad \mathbf{a} \in \Omega, \quad f: \Omega \rightarrow \mathbb{R}^m.$$

Aby definicja $D^k(f)(\mathbf{a})$ miała sens, różniczka rzędu $(k-1)$ powinna być funkcją określoną w otoczeniu punktu $\mathbf{a}$ i różniczkowalną w $\mathbf{a}$. Różniczka k -tego rzędu, $D^k(f)(\mathbf{a})$ jest przekształceniem k -liniowym z $\mathbb{R}^n \times \dots \times \mathbb{R}^n$ w $\mathbb{R}^m$, tzn.

$$D^k f(\mathbf{a}): \underbrace{\mathbb{R}^n \times \dots \times \mathbb{R}^n}_{k \text{ razy}} \ni (\mathbf{v}_1, \dots, \mathbf{v}_k) \longmapsto D^k f(\mathbf{a})(\mathbf{v}_1, \dots, \mathbf{v}_k) \in \mathbb{R}^m.$$

Wartość $D^k f(\mathbf{a})(\mathbf{v}_1, \dots, \mathbf{v}_k)$, która jest wektorem z $\mathbb{R}^m$, zależy liniowo od każdego z wektorów $\mathbf{v}_i$ ($i = 1, \dots, k$) z osobna. Taka interpretacja różniczki k -tego rzędu jest rzeczą naturalną: jeśli

$$D^{k-1}f: \Omega \ni \mathbf{x} \longmapsto D^{k-1}f(\mathbf{x}) \in L(\underbrace{\mathbb{R}^n, \dots, \mathbb{R}^n}_{k-1 \text{ razy}}, \mathbb{R}^m),$$

gdzie

$$L(\underbrace{\mathbb{R}^n, \dots, \mathbb{R}^n}_{k-1 \text{ razy}}, \mathbb{R}^m)$$

oznacza przestrzeń przekształceń $(k-1)$ -liniowych z $\mathbb{R}^n \times \mathbb{R}^n$ w $\mathbb{R}^m$, to zgodnie z definicją różniczki

$$D^k f: \Omega \ni \mathbf{x} \longmapsto D^k f(\mathbf{x}) = D(D^{k-1}f)(\mathbf{x}) \in L(\mathbb{R}^n, L(\underbrace{\mathbb{R}^n, \dots, \mathbb{R}^n}_{k-1 \text{ razy}}, \mathbb{R}^m)).$$

Jednak przestrzenie

$$V_1 := L(\mathbb{R}^n, L(\underbrace{\mathbb{R}^n, \dots, \mathbb{R}^n}_{k-1 \text{ razy}}, \mathbb{R}^m)) \quad \text{oraz} \quad V_2 := L(\underbrace{\mathbb{R}^n, \dots, \mathbb{R}^n}_{k \text{ razy}}, \mathbb{R}^m)$$

można utożsamić; ich naturalnym izomorfizmem jest przekształcenie $V_1 \ni F \mapsto B_F \in V_2$, gdzie F i B_F powiązane są zależnością

$$F(\mathbf{v}_1)(\mathbf{v}_2, \dots, \mathbf{v}_k) = B_F(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k).$$

Uwaga 2.51. Jeśli $D^k f(\mathbf{a})$ istnieje, to jest przekształceniem wieloliniowym symetrycznym, tzn.

$$D^k f(\mathbf{a})(\mathbf{v}_1, \dots, \mathbf{v}_k) = D^k f(\mathbf{a})(\mathbf{v}_{\sigma(1)}, \dots, \mathbf{v}_{\sigma(k)})$$

dla każdej permutacji $\sigma \in S_k$ zbioru k -elementowego. Można to udowodnić przez indukcję względem k , posługując się twierdzeniem Schwarz'a o symetrii drugiej różniczki.

Uwaga 2.52. Będziemy odtąd używać oznaczenia

$$Df(\mathbf{a})\mathbf{h}^k = Df(\mathbf{a})(\underbrace{\mathbf{h}, \dots, \mathbf{h}}_{k \text{ razy}}) \tag{2.38}$$

dla oznaczenia wartości k -tej różniczki (która jest przekształceniem k -liniowym) na układzie k identycznych wektorów.

Aby wszystkie rachunki w następnym podrozdziale Czytelnik mógł prześledzić ze zrozumieniem, wprowadzimy jeszcze jedną definicję.

Definicja 2.53 (norma przekształcenia wieloliniowego). Normą przekształcenia k -liniowego

$$B \in L(\underbrace{\mathbb{R}^n, \dots, \mathbb{R}^n}_k \text{ razy}, \mathbb{R}^m)$$

nazywamy najmniejszą stałą $C = \|B\| \geq 0$ taką, że

$$\|B(\mathbf{v}_1, \dots, \mathbf{v}_k)\| \leq \|B\| \cdot \|\mathbf{v}_1\| \cdot \dots \cdot \|\mathbf{v}_k\| \quad \text{dla wszystkich } \mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n.$$

Zapisywanie różniczek wyższych rzędów we współrzędnych jest zajęciem niewdzięcznym i nie będziemy tego robić. Wygodną metodą oznaczania pochodnych cząstkowych wyższych rzędów podamy w następnym podrozdziale.

2.5.4 Wzór Taylora. Funkcje klasy C^k i notacja wielowskaźnikowa.

Okazuje się, że przy odpowiedniej notacji, wprowadzonej wyżej, wzór Taylora w najprostszej wersji, z resztą w postaci Peano, wygląda zupełnie tak samo, jak dla funkcji jednej zmiennej.

Twierdzenie 2.54 (wzór Taylora z resztą w postaci Peano). Załóżmy, że funkcja $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}^m$ jest $(k-1)$ -krotnie różniczkowalna na Ω , kula $B(\mathbf{a}, r) \subset \Omega$ dla pewnego $r > 0$ i $D^k f(\mathbf{a})$ istnieje. Wówczas, dla $\|\mathbf{h}\| < r$,

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2!}D^2f(\mathbf{a})\mathbf{h}^2 + \dots + \frac{1}{k!}D^k f(\mathbf{a})\mathbf{h}^k + R(\mathbf{h}), \quad (2.39)$$

gdzie $R(\mathbf{h})/\|\mathbf{h}\|^k \rightarrow 0$ dla $\mathbf{h} \rightarrow 0$.

Dowód. Oszacujemy resztę

$$R(\mathbf{h}) = f(\mathbf{a} + \mathbf{h}) - \left(f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2!}D^2f(\mathbf{a})\mathbf{h}^2 + \dots + \frac{1}{k!}D^k f(\mathbf{a})\mathbf{h}^k \right),$$

stosując $k-1$ razy twierdzenie o wartości średniej. Zauważmy, że $R(0) = 0$, a ponadto

$$\begin{aligned} DR(\mathbf{h}) &= Df(\mathbf{a} + \mathbf{h}) - Df(\mathbf{a}) - \frac{1}{1!}D^2f(\mathbf{a})\mathbf{h} - \dots - \frac{1}{(k-1)!}D^k f(\mathbf{a})\mathbf{h}^{k-1}, \\ D^2R(\mathbf{h}) &= D^2f(\mathbf{a} + \mathbf{h}) - D^2f(\mathbf{a}) - \dots - \frac{1}{(k-2)!}D^k f(\mathbf{a})\mathbf{h}^{k-2}, \\ &\vdots \\ D^{k-1}R(\mathbf{h}) &= D^{k-1}f(\mathbf{a} + \mathbf{h}) - D^{k-1}f(\mathbf{a}) - D^k f(\mathbf{a})\mathbf{h}. \end{aligned}$$

Mamy więc $DR(\mathbf{0}) = \mathbf{0}, \dots, D^{k-1}R(\mathbf{0}) = \mathbf{0}$. Korzystając z Twierdzenia 2.40, otrzymujemy nierówności

$$\begin{aligned}
 \|R(\mathbf{h})\| &= \|R(\mathbf{h}) - R(\mathbf{0})\| \leq \|\mathbf{h}\| \cdot \sup_{\theta \in [0,1]} \|DR(\theta\mathbf{h})\| \\
 &\leq \|\mathbf{h}\|^2 \sup_{\theta \in [0,1]} \|D^2R(\theta\mathbf{h})\| \\
 &\vdots \\
 &\leq \|\mathbf{h}\|^{k-1} \sup_{\theta \in [0,1]} \|D^{k-1}R(\theta\mathbf{h})\|. \tag{2.40}
 \end{aligned}$$

Normę $\|D^{k-1}R(\theta\mathbf{h})\|$ trzeba oszacować inaczej, gdyż $D^k f$ istnieje *tylko* w punkcie $\mathbf{a}$. Można jednak skorzystać po prostu z definicji różniczki; wobec wzoru na $D^{k-1}R$ mamy

$$\begin{aligned}
 \sup_{\theta \in [0,1]} \|D^{k-1}R(\theta\mathbf{h})\| &= \sup_{\theta \in [0,1]} \left(|\theta| \cdot \|\mathbf{h}\| \frac{\|D^{k-1}f(\mathbf{a} + \theta\mathbf{h}) - D^{k-1}f(\mathbf{a}) - D^k f(\mathbf{a})\theta\mathbf{h}\|}{\|\theta\mathbf{h}\|} \right) \\
 &\leq \|\mathbf{h}\| \sup_{\theta \in [0,1]} \frac{\|D^{k-1}f(\mathbf{a} + \theta\mathbf{h}) - D^{k-1}f(\mathbf{a}) - D^k f(\mathbf{a})\theta\mathbf{h}\|}{\|\theta\mathbf{h}\|} = \|\mathbf{h}\| \cdot o(1) \quad \text{dla } \mathbf{h} \rightarrow \mathbf{0}.
 \end{aligned}$$

Łącząc tę nierówność z (2.40), otrzymujemy $\|R(\mathbf{h})\| = \|\mathbf{h}\|^k o(1)$ dla $\mathbf{h} \rightarrow \mathbf{0}$. $\square$

W praktyce wygodnie jest znać także inne postacie wzoru Taylora. Jedną z nich, używającą tzw. notacji wielowskaźnikowej, podajemy niżej.

Notacja wielowskaźnikowa. Funkcje klasy C^k .

Definicja 2.55. Wielowskaźnik $\alpha = (\alpha_1, \dots, \alpha_n)$ to wektor o współrzędnych α_i całkowitych nieujemnych, lub równoważnie element zbioru $(\mathbb{N} \cup \{0\})^n$. Dla wielowskaźników α, β i każdego punktu $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ piszemy:

$$\alpha! = \alpha_1! \cdot \dots \cdot \alpha_n!, \quad |\alpha| = \alpha_1 + \dots + \alpha_n, \tag{2.41}$$

$$\mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdot \dots \cdot x_n^{\alpha_n}, \tag{2.42}$$

$$\beta \leq \alpha \Leftrightarrow \beta_i \leq \alpha_i \quad \text{dla wszystkich } i = 1, \dots, n, \tag{2.43}$$

$$\binom{\alpha}{\beta} = \frac{\alpha!}{\beta!} (\alpha - \beta)! \quad \text{dla } \beta \leq \alpha, \tag{2.44}$$

gdzie $\alpha - \beta = (\alpha_1 - \beta_1, \dots, \alpha_n - \beta_n)$.

Za pomocą wielowskaźników wygodnie jest oznaczać pochodne cząstkowe wyższych rzędów w takich sytuacjach, gdy kolejność wykonywania poszczególnych różniczkowań nie ma znaczenia.

Definicja 2.56. Niech Ω będzie zbiorem otwartym w $\mathbb{R}^n$. Mówimy, że $f \in C^k(\Omega, \mathbb{R}^m)$ wtedy i tylko wtedy, gdy f ma wszystkie pochodne cząstkowe rzędu k ciągle na zbiorze Ω .

Uwaga 2.57. Podobnie jak w przypadku funkcji klasy C^1 , powyższa definicja jest równoważna temu, że f jest ciągła na Ω i wszystkie przekształcenia j -liniowe

$$D^j f: \Omega \ni \mathbf{x} \longmapsto D^j f(\mathbf{x}) \in L(\underbrace{\mathbb{R}^n, \dots, \mathbb{R}^n}_j \text{ razy}, \mathbb{R}^m) \quad (j = 1, 2, \dots, k)$$

są dobrze określone i ciągłe na zbiorze Ω . Można to udowodnić, posługując się Twierdzeniem 2.16. Jest to dość łatwe: rozumowanie wymaga tylko znajomości pojęć i nie są potrzebne żadne rachunki.

Umowa. Dla funkcji $f \in C^k(\Omega, \mathbb{R}^m)$ symbol

$$D^\alpha f(\mathbf{x}) \stackrel{\text{ozn.}}{=} (D_1)^{\alpha_1} (D_2)^{\alpha_2} \dots (D_n)^{\alpha_n} f(\mathbf{x}), \quad \mathbf{x} \in \Omega, \quad (2.45)$$

oznacza pochodną cząstkową tej funkcji, rzędu $|\alpha| = \alpha_1 + \dots + \alpha_n$, przy czym różniczkowanie względem zmiennej x_i wykonujemy α_i -krotnie ($i = 1, 2, \dots, n$). Z twierdzenia Schwarz'a o równości pochodnych mieszanych wynika, że dla funkcji klasy C^k kolejność wykonywania różniczkowań nie odgrywa roli; można więc w ten sposób oznaczać wszystkie pochodne cząstkowe takiej funkcji, bez obawy, że nie wiadomo, o jaką pochodną chodzi. Przyjmujemy także

$$D^{(0,0,\dots,0)} f \equiv f.$$

Twierdzenie 2.58 (wzór Taylora z resztą całkową). Załóżmy, że $\Omega \subset \mathbb{R}^n$ jest zbiorem otwartym i $f \in C^k(\Omega, \mathbb{R}^m)$. Jeśli kula $B(\mathbf{a}, \delta) \subset \Omega$ dla pewnego $\delta > 0$, to wówczas

$$f(\mathbf{a} + \mathbf{h}) = \sum_{|\alpha| \leq k} \frac{1}{\alpha!} D^\alpha f(\mathbf{a}) \mathbf{h}^\alpha + R_k(\mathbf{a}, \mathbf{h}), \quad (2.46)$$

dla $\|\mathbf{h}\| < \delta$, gdzie reszta

$$R_k(\mathbf{x}, \mathbf{h}) = k \int_0^1 (1-t)^{k-1} \sum_{|\alpha|=k} \frac{1}{\alpha!} \left(D^\alpha f(\mathbf{a} + t\mathbf{h}) - D^\alpha f(\mathbf{a}) \right) \mathbf{h}^\alpha dt. \quad (2.47)$$

Uwaga. We wzorze (2.46) sumowanie po prawej stronie odbywa się względem *wszystkich* wielowskazyńników długości $|\alpha| \leq k$.

Dowód. Skorzystamy ze wzoru Taylora z resztą całkową dla funkcji jednej zmiennej rzeczywistej (patrz Skrypt z Analizy Matematycznej I). Ustalmy $\mathbf{a} \in \Omega$ i $\mathbf{h} = (h_1, \dots, h_n)$, $\|\mathbf{h}\| < \delta$. Niech $g(t) = f(\mathbf{a} + t\mathbf{h})$ dla $t \in [0, 1]$. Funkcja g jest klasy C^k na pewnym odcinku otwartym $I \supset [0, 1]$. Dlatego

$$\begin{aligned} g(1) &= \sum_{j=0}^{k-1} \frac{g^{(j)}(0)}{j!} + \int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} g^{(k)}(t) dt \\ &= \sum_{j=0}^k \frac{g^{(j)}(0)}{j!} + \int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} (g^{(k)}(t) - g^{(k)}(0)) dt. \end{aligned} \quad (2.48)$$

Aby zakończyć pracę, wyrazimy pochodne funkcji g przez pochodne cząstkowe funkcji f . Posługując się wzorem na pochodną złożenia, dowodzimy przez indukcję, że

$$g'(t) = \sum_{i=1}^n D_i f(\mathbf{a} + t\mathbf{h}) h_i, \quad g''(t) = \sum_{i_1, i_2=1}^n D_{i_2} D_{i_1} f(\mathbf{a} + t\mathbf{h}) h_{i_1} h_{i_2}$$

itd.; ogólnie,

$$g^{(s)}(t) = \sum_{i_1, i_2, \dots, i_s=1}^n D_{i_s} \dots D_{i_2} D_{i_1} f(\mathbf{a} + t\mathbf{h}) h_{i_1} h_{i_2} \dots h_{i_s}, \quad s = 1, \dots, k. \quad (2.49)$$

Ostatnią sumę zapiszemy, używając notacji wielowskazanikowej. Ustalmy wielowskazanik $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ taki, że $|\alpha| = s \in [1, k]$. Liczba takich ciągów $(i_1, i_2, \dots, i_s)$ o wyrazach ze zbioru $\{1, 2, \dots, n\}$, w których 1, 2, ..., n występują (odpowiednio) α_1 -krotnie, α_2 -krotnie, ..., α_n krotnie, wynosi, zgodnie ze znanym wzorem kombinatorycznym,

$$\binom{s}{\alpha_1} \cdot \binom{s - \alpha_1}{\alpha_2} \cdot \dots \cdot \binom{s - (\alpha_1 + \dots + \alpha_{n-1})}{\alpha_n} = \frac{s!}{\alpha_1! \cdot \alpha_2! \cdot \dots \cdot \alpha_n!} = \frac{s!}{\alpha!}.$$

Dla każdego z tych ciągów mamy

$$D_{i_s} \dots D_{i_2} D_{i_1} f(\mathbf{a} + t\mathbf{h}) h_{i_1} h_{i_2} \dots h_{i_s} = D^\alpha f(\mathbf{a} + t\mathbf{h}) \mathbf{h}^\alpha,$$

a więc wzór (2.49) można przepisać w postaci

$$\frac{g^{(s)}(t)}{s!} = \sum_{|\alpha|=s} \frac{1}{\alpha!} D^\alpha f(\mathbf{a} + t\mathbf{h}) \mathbf{h}^\alpha, \quad s = 1, 2, \dots, k. \quad (2.50)$$

Podstawiając (2.50) do wzoru (2.48), otrzymujemy tezę twierdzenia. $\square$

W szczególnym przypadku $k = 2$, dla funkcji $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}$ różniczkowalnej dwukrotnie w sposób ciągły, można podobnie (stosując wzór Taylora z resztą Lagrange'a dla funkcji $g(t) = f(\mathbf{a} + t\mathbf{h})$ jednej zmiennej t) uzyskać następujący fakt.

Wniosek 2.59. *Jeśli $\Omega \subset \mathbb{R}^n$, $f \in C^2(\Omega, \mathbb{R})$ i odcinek $[\mathbf{a}, \mathbf{a} + \mathbf{h}] \subset \Omega$, gdzie $\mathbf{h} = (h_1, \dots, h_n)$, to istnieje wówczas punkt $\theta \in (0, 1)$ taki, że*

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \sum_{i=1}^n f_{x_i}(\mathbf{a}) h_i + \frac{1}{2} \sum_{i,j=1}^n f_{x_i x_j}(\mathbf{a} + \theta \mathbf{h}) h_i h_j. \quad (2.51)$$

Dowód. Ćwiczenie dla Czytelnika.

Zadanie 2.60. Wykazać, że dla każdego $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ i dla każdego $k \in \mathbb{N}$ jest

$$(x_1 + x_2 + \dots + x_n)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} \mathbf{x}^\alpha.$$

Wskazówka. Oznaczyć lewą stronę $f(\mathbf{x})$ i zastosować wzór Taylora.

2.5.5 Ekstrema lokalne.

Wiemy już, że warunkiem koniecznym, by funkcja różniczkowalna $f: \Omega \rightarrow \mathbb{R}$ miała ekstremum w punkcie $\mathbf{a} \in \Omega$, jest znikanie jej gradientu w tym punkcie. Zajmiemy się teraz sformułowaniem warunków dostatecznych istnienia ekstremum lokalnego funkcji klasy C^2 . Wyrazimy je za pomocą własności drugiej różniczki.

Definicja 2.61. Niech $f \in C^2(\Omega, \mathbb{R})$. Macierz

$$H_f(\mathbf{a}) = D^2 f(\mathbf{a}) = (f_{x_i x_j}(\mathbf{a}))_{i,j=1,2,\dots,n} \in M_{n \times n}$$

nazywamy *hesjanem* funkcji f w punkcie $\mathbf{a} \in \Omega$.

Z twierdzenia Schwarza o równości pochodnych mieszanych wynika natychmiast, że $H_f(\mathbf{a})$ jest macierzą symetryczną. Jak wiadomo z wykładów Algebry Liniowej, wszystkie wartości własne macierzy symetrycznej A są rzeczywiste, a w $\mathbb{R}^n$ istnieje baza ortonormalna, złożona z wektorów własnych A .

Przypomnijmy, że macierz symetryczna $A \in M_{n \times n}$ nazywa się dodatnia (ujemna) wtedy i tylko wtedy, gdy $\langle A\mathbf{v}, \mathbf{v} \rangle > 0$ dla $\mathbf{v} \in \mathbb{R}^n \setminus \{0\}$ (odpowiednio $\langle A\mathbf{v}, \mathbf{v} \rangle < 0$ dla $\mathbf{v} \in \mathbb{R}^n \setminus \{0\}$). Macierze nieujemne i niedodatnie definiuje się analogicznie, za pomocą nierówności nieostrych. Jeśli A jest dodatnia (ujemna, nieujemna, niedodatnia), to piszemy $A > 0$ (odpowiednio: $A < 0$, $A \geq 0$, $A \leq 0$).

Stwierdzenie 2.62. Załóżmy, że $\mathbf{a} \in \Omega$ jest punktem krytycznym funkcji $f \in C^2(\Omega, \mathbb{R})$. Jeśli f ma w $\mathbf{a}$ minimum (odpowiednio: maksimum) lokalne, to $H_f(\mathbf{a}) \geq 0$ (odpowiednio: $H_f(\mathbf{a}) \leq 0$).

Dowód. Dla ustalenia uwagi załóżmy, że f ma w $\mathbf{a}$ minimum lokalne. W punkcie krytycznym $f_{x_i}(\mathbf{a}) = 0$ dla $i = 1, 2, \dots, n$. Dlatego ze wzoru Taylora (2.51) (patrz Wniosek 2.59) otrzymujemy

$$\frac{1}{2} \langle H_f(\mathbf{a} + \theta \mathbf{h}) \mathbf{h}, \mathbf{h} \rangle = \frac{1}{2} \sum_{1 \leq i, j \leq n} f_{x_i x_j}(\mathbf{a} + \theta \mathbf{h}) h_i h_j = f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) \geq 0$$

dla wszystkich $\|\mathbf{h}\|$ dostatecznie małych. Ustalmy $\mathbf{v} \in \mathbb{R}^n \setminus \{0\}$ i podstawmy w tej nierówności $\mathbf{h} = t\mathbf{v}$, gdzie $t \in \mathbb{R}$ i $|t| \ll 1$. Dzieląc obie strony przez $\frac{1}{2}t^2 > 0$, otrzymujemy

$$\langle H_f(\mathbf{a} + \theta t \mathbf{v}) \mathbf{v}, \mathbf{v} \rangle \geq 0.$$

Przechodząc do granicy $t \rightarrow 0$ i korzystając z ciągłości drugich pochodnych cząstkowych f otrzymujemy $\langle H_f(\mathbf{a}) \mathbf{v}, \mathbf{v} \rangle \geq 0$.

Jeśli f ma w $\mathbf{a}$ maksimum lokalne, to rozpatrujemy funkcję $-f$, która ma w tym punkcie minimum lokalne. $\square$

Przydatna w praktyce jest oczywiście implikacja odwrotna.

Twierdzenie 2.63 (warunki dostateczne ekstremów lokalnych). Niech $\Omega \subset \mathbb{R}^n$ będzie zbiorem otwartym. Przypuśćmy, że $f \in C^2(\Omega, \mathbb{R})$ ma w $\mathbf{a} \in \Omega$ punkt krytyczny, tzn. $\text{grad } f(\mathbf{a}) = 0$. Wówczas:

- (i) Jeśli $H_f(\mathbf{x}) \geq 0$ w pewnym otoczeniu punktu $\mathbf{a}$, to f ma w $\mathbf{a}$ minimum lokalne.
- (ii) Jeśli $H_f(\mathbf{a}) > 0$, to f ma w $\mathbf{a}$ minimum lokalne właściwe.
- (iii) Jeśli $H_f(\mathbf{x}) \leq 0$ w pewnym otoczeniu punktu $\mathbf{a}$, to f ma w $\mathbf{a}$ maksimum lokalne.
- (iv) Jeśli $H_f(\mathbf{a}) < 0$, to f ma w $\mathbf{a}$ maksimum lokalne właściwe.

Dowód. Ponieważ $f_{x_i}(\mathbf{a}) = 0$ dla $i = 1, 2, \dots, n$, więc ze wzoru Taylora (2.51) otrzymujemy

$$\frac{1}{2} \langle H_f(\mathbf{a} + \theta \mathbf{h}) \mathbf{h}, \mathbf{h} \rangle = f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}),$$

gdzie $\theta = \theta(\mathbf{h}) \in (0, 1)$. Z tej równości natychmiast wynikają podpunkty (i) oraz (iii) Twierdzenia 2.63.

Założmy teraz, że $A := H_f(\mathbf{a}) > 0$. Funkcja $\mathbb{S}^{n-1} \ni \mathbf{v} \mapsto \phi(\mathbf{v}) = \langle A\mathbf{v}, \mathbf{v} \rangle$ jest wtedy dodatnia i ciągła na sferze jednostkowej $\mathbb{S}^{n-1}$, która jest zbiorem zwartym. Wobec twierdzenia Weierstrassa o przyjmowaniu kresów, istnieją stałe $\alpha, \beta > 0$ takie, że

$$\beta \geq \phi(\mathbf{v}) = \langle A\mathbf{v}, \mathbf{v} \rangle \geq \alpha > 0 \quad \text{dla wszystkich } \mathbf{v} \in \mathbb{S}^{n-1}.$$

Podstawiając w tej nierówności $\mathbf{v} = \mathbf{w}/\|\mathbf{w}\|$, gdzie $\mathbf{w} \in \mathbb{R}^n$ jest dowolnym wektorem różnym od 0, otrzymujemy

$$\beta \|\mathbf{w}\|^2 \geq \langle A\mathbf{w}, \mathbf{w} \rangle \geq \alpha \|\mathbf{w}\|^2 > 0 \quad \text{dla wszystkich } \mathbf{w} \in \mathbb{R}^n \setminus \{0\}.$$

Dlatego, z nierówności Schwarza i definicji normy macierzy,

$$\begin{aligned} \langle H_f(\mathbf{a} + \theta \mathbf{h}) \mathbf{h}, \mathbf{h} \rangle &= \langle A\mathbf{h}, \mathbf{h} \rangle + \langle (H_f(\mathbf{a} + \theta \mathbf{h}) - A) \mathbf{h}, \mathbf{h} \rangle \\ &\geq \alpha \|\mathbf{h}\|^2 - \|(H_f(\mathbf{a} + \theta \mathbf{h}) - A) \mathbf{h}\| \cdot \|\mathbf{h}\| \\ &\geq \alpha \|\mathbf{h}\|^2 - \|(H_f(\mathbf{a} + \theta \mathbf{h}) - A)\| \cdot \|\mathbf{h}\|^2 \end{aligned} \quad (2.52)$$

Ponieważ $f \in C^2$, więc wszystkie współrzędne macierzy $H_f(\mathbf{x})$ zależą od $\mathbf{x}$ w sposób ciągły. Istnieje zatem liczba $\delta > 0$ taka, że jeśli $0 < \|\mathbf{h}\| < \delta$ i $\theta \in (0, 1)$, to

$$\|(H_f(\mathbf{a} + \theta \mathbf{h}) - H_f(\mathbf{a}))\| = \|(H_f(\mathbf{a} + \theta \mathbf{h}) - A)\| < \frac{\alpha}{2}.$$

Wtedy jednak, wobec (2.52),

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = \frac{1}{2} \langle H_f(\mathbf{a} + \theta \mathbf{h}) \mathbf{h}, \mathbf{h} \rangle > \frac{\alpha}{4} \|\mathbf{h}\|^2 > 0.$$

To dowodzi punktu (ii). Dowód (iv) jest taki sam. $\square$

Uwaga 2.64. W dowodach podpunktów (ii) oraz (iv) w Twierdzeniu 2.63 nie trzeba zakładać, że $f \in C^2$. Wystarczy po prostu, żeby f była różniczkowalna na zbiorze Ω i jej druga różniczka $D^2f(\mathbf{a})$ istniała w punkcie krytycznym $\mathbf{a}$ i była w nim dodatnia (wtedy f ma w $\mathbf{a}$ minimum lokalne właściwe) bądź ujemna (wtedy f ma w $\mathbf{a}$ maksimum lokalne właściwe). W dowodzie wykorzystuje się wzór Taylora z resztą Peano. Zainteresowany Czytelnik zdoła sam uzupełnić szczegóły rozumowania.

Zanim przejdziemy do przykładów, przytoczymy jeszcze twierdzenie, które pozwala wnioskować, kiedy f z pewnością *nie* ma ekstremum lokalnego w punkcie krytycznym.

Twierdzenie 2.65. *Założmy, że $f \in C^1(\Omega, \mathbb{R})$ ma w $\mathbf{a} \in \Omega$ punkt krytyczny i $D^2f(\mathbf{a})$ istnieje. Jeśli $H_f(\mathbf{a}) = D^2f(\mathbf{a})$ ma wartość własną $\lambda_1 > 0$ i wartość własną $\lambda_2 < 0$, to f nie ma ekstremum lokalnego w punkcie $\mathbf{a}$.*

Dowód. Niech $\mathbf{v}_i$ będzie unormowanym wektorem własnym macierzy $H_f(\mathbf{a})$, odpowiadającym wartości własnej λ_i , gdzie $i = 1, 2$. Dla dostatecznie małej liczby $\delta > 0$ rozpatrzmy dwie funkcje pomocnicze,

$$g_i(t) = f(\mathbf{a} + t\mathbf{v}_i), \quad |t| < \delta, \quad i = 1, 2.$$

Mamy $g'_i(t) = Df(\mathbf{a} + t\mathbf{v}_i)\mathbf{v}_i$, tzn. $g'_i(0) = 0$, oraz

$$g''(t) = \langle D^2f(\mathbf{a} + t\mathbf{v}_i)\mathbf{v}_i, \mathbf{v}_i \rangle.$$

Zatem $g''_1(0) = \langle D^2f(\mathbf{a})\mathbf{v}_1, \mathbf{v}_1 \rangle = \lambda_1 \|\mathbf{v}_1\|^2 = \lambda_1 > 0$. Podobnie, $g''_2(0) = \lambda_2 < 0$. Dlatego g_1 ma minimum lokalne właściwe w zerze, a g_2 ma maksimum lokalne właściwe w zerze. Wynika stąd, że f nie ma ekstremum lokalnego w punkcie $\mathbf{a}$ (gdyby miała, to każda z funkcji $g_v = f(\mathbf{a} + t\mathbf{v})$ miałaby w zerze ekstremum lokalne tego samego typu, co f). $\square$

To, czy macierz $H_f(\mathbf{a}) = D^2f(\mathbf{a})$ jest dodatnia (ujemna), można rozpoznawać za pomocą kryterium Sylwestera.

Twierdzenie 2.66 (kryterium Sylwestera). Niech $A = (a_{ij}) \in M_{n \times n}(\mathbb{R})$ i $a_{ij} = a_{ji}$ dla wszystkich $i, j = 1, 2, \dots, n$. Oznaczmy

$$d_\ell = \det \left(a_{ij} \right)_{i,j=1,\dots,\ell}, \quad \ell = 1, 2, \dots, n.$$

- (i) Jeśli $d_\ell > 0$ dla każdego $\ell = 1, 2, \dots, n$, to $A > 0$.
- (ii) Jeśli $(-1)^\ell d_\ell > 0$ dla każdego $\ell = 1, 2, \dots, n$, to $A < 0$.
- (iii) Jeśli $d_\ell \neq 0$ dla każdego $\ell = 1, \dots, n$, ale nie zachodzi ani założenie (i), ani założenie (ii), to macierz A ma wartości własne różnych znaków.

Dowód Czytelnik miał okazję poznać na wykładach z Algebry Liniowej. Zainteresowanym polecam książkę A. Mostowskiego i M. Starka *Elementy algebry wyższej*.

Uwaga 2.67. Jeśli $f: \mathbb{R}^n \supset \Omega \rightarrow \mathbb{R}$ jest klasy C^2 , ma punkt krytyczny $\mathbf{a} \in \Omega$ i wszystkie wartości własne macierzy $D^2f(\mathbf{a})$ są różne od zera, to mówimy, że $\mathbf{a}$ jest *niezdegenerowanym punktem krytycznym*. Z Twierdzeń 2.65 i 2.63 wynika, że o tym, czy funkcja f ma w niezdegenerowanym punkcie krytycznym ekstremum lokalne, można jednoznacznie przesądzić, badając znaki wartości własnych macierzy $D^2f(\mathbf{a})$.

Uwaga 2.68. Podkreślmy wyraźnie: założenie ostrych nierówności w punktach (ii) i (iv) Twierdzenia 2.63 jest istotne. Każda z funkcji

$$f_1(x, y) = x^4 + y^4, \quad f_2(x, y) = -x^4 - y^4, \quad f_3(x, y) = x^4 - y^4, \quad (x, y) \in \mathbb{R}^2$$

ma (jedyny) punkt krytyczny w $(0, 0)$. Jest oczywiste, że dla funkcji f_1 ten punkt jest minimum lokalnym właściwym, dla f_2 – maksimum lokalnym właściwym, natomiast f_3 w ogóle nie ma tym punkcie ekstremum lokalnego. Mamy jednak

$$D^2f_i(0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad i = 1, 2, 3.$$

Biorąc

$$f_4(x, y) = x^2 + y^4, \quad f_5(x, y) = x^2, \quad f_6(x, y) = x^2 - y^4, \quad (x, y) \in \mathbb{R}^2$$

otrzymamy

$$\text{grad } f_i(0,0) = (0,0) \quad \text{oraz} \quad D^2 f_i(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \quad i = 4, 5, 6.$$

Łatwo zauważyć, że w punkcie $(0,0) \in \mathbb{R}^2$ funkcja f_4 ma minimum lokalne właściwe, f_5 – minimum lokalne (które nie jest właściwe), natomiast f_6 w ogóle nie ma ekstremum.

Przykład 2.69. Niech $h(x,y) = ay(e^x - 1) + x \sin x + 1 - \cos y$ dla $x, y \in \mathbb{R}$. Wykażemy, że h ma ekstremum lokalne w punkcie $(0,0)$ wtedy i tylko wtedy, gdy $a \in (-\sqrt{2}, \sqrt{2})$.

Pochodne cząstkowe funkcji h są równe

$$h_x(x,y) = aye^x + \sin x + x \cos x, \quad h_y(x,y) = a(e^x - 1) + \sin y,$$

więc (niezależnie od wartości parametru $a \in \mathbb{R}$) jest $h_x(0,0) = h_y(0,0) = 0$ i h ma w zerze punkt krytyczny. Dalej, obliczamy

$$\begin{aligned} h_{xx}(x,y) &= aye^x + 2 \cos x - x \sin x, \\ h_{xy}(x,y) &= ae^x, \quad h_{yy} = \cos y. \end{aligned}$$

Podstawiając $x = y = 0$ otrzymujemy

$$D^2 h(0,0) = \begin{pmatrix} 2 & a \\ a & 1 \end{pmatrix}, \quad \det D^2 h(0,0) = 2 - a^2.$$

Jeśli $a \in (-\sqrt{2}, \sqrt{2})$, to $\det D^2 h(0,0) = 2 - a^2 > 0$ i z kryterium Sylwestera wynika, że macierz $D^2 h(0,0)$ jest dodatnio określona, a więc h ma minimum lokalne właściwe w punkcie $(0,0)$ (patrz Twierdzenie 2.63 (ii)).

Jeśli $a \notin [-\sqrt{2}, \sqrt{2}]$, to $\det D^2 h(0,0) = 2 - a^2 < 0$. Macierz $D^2 h(0,0)$ ma więc wartości własne różnych znaków i wobec Twierdzenia 2.65 h nie ma w zerze ekstremum lokalnego.

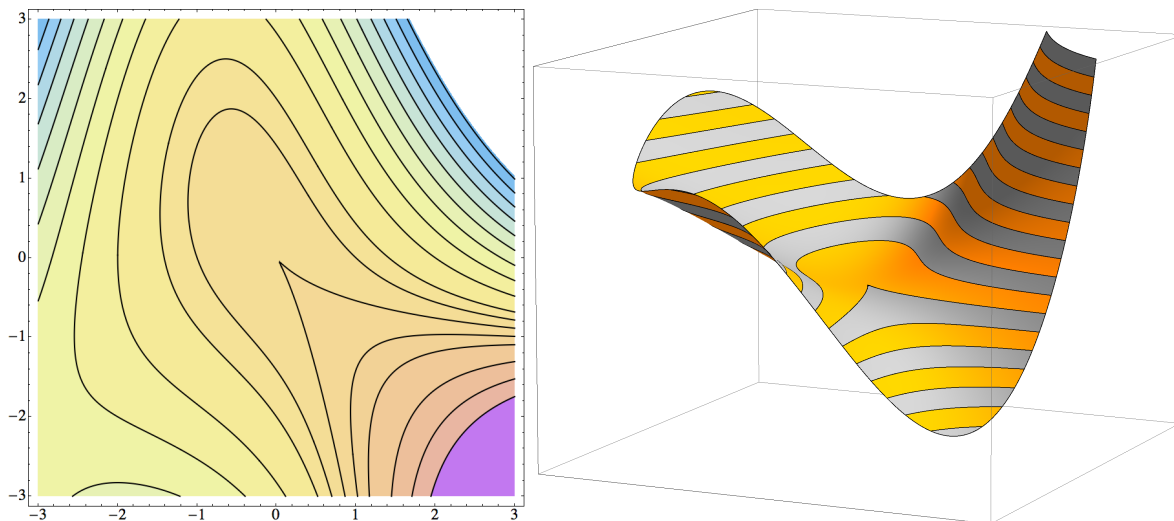
Przypadek $a = \pm\sqrt{2}$ trzeba rozpatrzyć osobno. Macierz $D^2 h(0,0)$ ma wtedy wartości własne 3 i 0, więc nie jest dodatnia i nie wolno stosować Twierdzenia 2.63 (ii); jak wynika z wcześniej przytoczonych przykładów, w takiej sytuacji funkcja może zarówno mieć ekstremum lokalne, jak i go nie mieć.

Dla ustalenia uwagi, niech $a = \sqrt{2}$. Użyjemy wzoru Taylora (najprościej jest w tym przypadku wykorzystać znane rozwinięcia funkcji elementarnych) i napiszemy

$$\begin{aligned} h(x,y) &= y\sqrt{2}\left(x + \frac{x^2}{2} + \frac{x^3}{6} + o(x^3)\right) + x\left(x - \frac{x^3}{6} + o(x^3)\right) + \frac{y^2}{2} - \frac{y^4}{4!} + o(y^4) \\ &= x^2 + xy\sqrt{2} + \frac{y^2}{2} + \frac{yx^2}{\sqrt{2}} + o(x^3) + o(y^3) \\ &= \left(x + \frac{y}{\sqrt{2}}\right)^2 + \frac{yx^2}{\sqrt{2}} + o(x^3) + o(y^3). \end{aligned}$$

Na prostej $y = -x\sqrt{2}$ mamy więc $h(x, -x\sqrt{2}) = -x^3 + o(x^3)$, $x \rightarrow 0$. Zatem, h nie ma ekstremum w zerze: wyrazy trzeciego rzędu we wzorze Taylora powodują, że $h(x, -x\sqrt{2})$ zmienia znak w każdym otoczeniu $0 \in \mathbb{R}$, a przecież $h(0,0) = 0$. Przypadek $a = -\sqrt{2}$ sprawdza się tak samo; Czytelnik łatwo uzupełni szczegóły obliczeń.

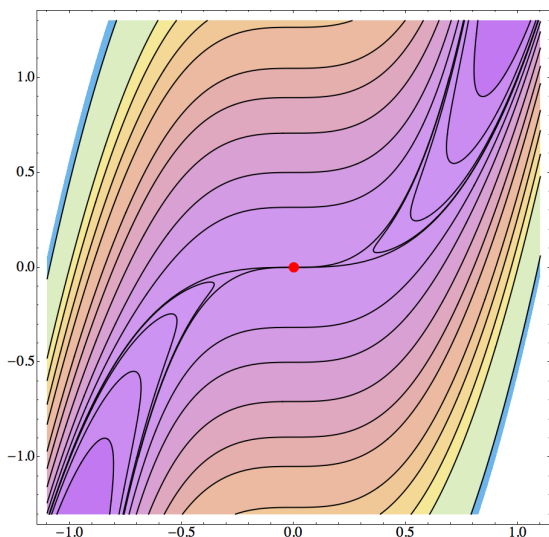
Czytelnik może sprawdzić, że kierunek prostej $y = -x\sqrt{2}$ jest wyznaczony przez wektor $\mathbf{v} \in \mathbb{S}^1$ taki, że $D^2 h(0,0)(\mathbf{v}, \mathbf{v}) = 0$. W innych kierunkach hesjan ma dodatnie wartości. Sprawdzaliśmy więc w istocie, jak zachowuje się funkcja h wokół zera “w podejrzanym kierunku” – i to wystarczyło, by stwierdzić brak ekstremum lokalnego. $\square$



Ilustracja do Przykładu 2.69. Parametr $a = \sqrt{2}$. Mamy wówczas $h(x, y) = P(x, y) + o(x^3) + o(y^3)$ dla $x, y \rightarrow 0$, gdzie

$$P(x, y) = \left(x + \frac{y}{\sqrt{2}}\right)^2 + \frac{yx^2}{\sqrt{2}}$$

jest wielomianem Taylora rzędu 3 funkcji h wokół zera. Po lewej: poziomice funkcji P , narysowane na płaszczyźnie $\mathbb{R}^2$ (w dziedzinie funkcji). Zbiór punktów w $\mathbb{R}^2$, opisany równaniem $P(x, y) = 0$, to krzywa z wyraźnym dziobkiem. Po prawej: fragment wykresu funkcji P , tzn. powierzchnia w $\mathbb{R}^3$ o równaniu $z = P(x, y)$.



Ilustracja do Przykładu 2.70. Krzywe $g(x, y) = \text{const}$ na płaszczyźnie $\mathbb{R}^2$. W punkcie krytycznym $(0, 0)$ spotykają się dwa szerokie grzbiety i dwie wąskie, wygięte doliny.

W Przykładzie 2.69 wystarczyło użyć twierdzeń 2.63 i 2.65 (dających automatyczne kryteria badania funkcji wokół punktu krytycznego), a w wątpliwym przypadku zbadać zachowanie funkcji na prostych, przechodzących przez punkt krytyczny. Podkreślimy jednak, że z zachowania funkcji na poszczególnych takich prostych *nie wolno* wnioskować, że ma ona ekstremum lokalne!

Przykład 2.70. Niech

$$g(x, y) = (y - x^3)(y - 3x^3)$$

dla $(x, y) \in \mathbb{R}^2$. Wtedy

$$g_x(x, y) = 18x^5 - 12x^2y,$$

$$g_y(x, y) = -4x^3 + 2y,$$

$$g_{xx}(x, y) = 90x^4 - 24xy,$$

$$g_{xy}(x, y) = -12x^2,$$

$$g_{yy}(x, y) = 2.$$

Zatem funkcja g ma w zerze (jedyne) punkt krytyczny;

$$D^2g(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \geq 0.$$

Na każdej prostej $y = kx$ jest $g(kx, x) = k^2x^2 + o(x^3)$ dla $x \rightarrow 0$, a więc obcięcie funkcji g do takiej prostej ma w zerze *minimum* lokalne (właściwe). Na prostej $x = 0$ jest $g(x, y) = g(0, y) = y^2$ (tzn. znów mamy funkcję jednej zmiennej, która ma minimum w zerze). Jednak na krzywej $y = 2x^3$ jest $g(x, y) = g(x, 2x^3) = -x^6$, a więc w dowolnie małym otoczeniu zera funkcja g przyjmuje nie tylko wartości dodatnie, ale także ujemne.

2.6 Funkcje gładkie

Definicja 2.71. Jeśli $\Omega \subset \mathbb{R}^n$ jest zbiorem otwartym, to przyjmujemy

$$C^\infty(\Omega, \mathbb{R}^m) = \bigcap_{k=1}^{\infty} C^k(\Omega, \mathbb{R}^m)$$

Funkcje $f \in C^\infty(\Omega, \mathbb{R}^m)$ nazywamy funkcjami klasy C^∞ lub *funkcjami gładkimi*. Są to funkcje, które mają ciągłe pochodne cząstkowe wszystkich rzędów (a zatem, mają ciągłe różniczki wszystkich rzędów).

Dla krótkości, pisze się $C^\infty(\Omega)$ zamiast $C^\infty(\Omega, \mathbb{R})$.

Definicja 2.72. Nośnikiem $\text{supp } f$ funkcji $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ nazywamy domknięcie zbioru tych punktów, w których f ma wartości różne od zera:

$$\text{supp } f = \overline{\{x \in \mathbb{R}^n: f(\mathbf{x}) \neq \mathbf{0}\}}.$$

Twierdzenie 2.73. Istnieją funkcje klasy $C^\infty(\mathbb{R}^n)$, których nośnik jest niepustym zbiorem zwartym. Ściślej mówiąc, dla każdego punktu $\mathbf{a} \in \mathbb{R}^n$ i każdych liczb $0 < r < R$ istnieje funkcja $f \in C^\infty(\mathbb{R}^n)$ taka, że $f \equiv 1$ na kuli $B(\mathbf{a}, r)$ i $f \equiv 0$ na $\mathbb{R}^n \setminus B(\mathbf{a}, R)$.

SZKIC DOWODU. *Krok 1.* Niech $n = 1$. Nietrudno wykazać, że istnieje funkcja $\varphi_1: \mathbb{R} \rightarrow \mathbb{R}$, która jest klasy C^∞ i znika poza przedziałem $[-1, 1]$, ale $\varphi_1(0) = 1$. Taką funkcją jest np.

$$\varphi_1(x) = \begin{cases} \exp(-\text{tg}^2(\pi x/2)), & |x| < 1, \\ 0, & |x| \geq 1. \end{cases}$$

Sprawdzenie, że φ_1 istotnie spełnia podane warunki, pozostawiamy jako ćwiczenie dla Czytelnika.

Krok 2. Funkcja

$$\varphi_2(x) = \int_{-\infty}^x \varphi_1(t) dt$$

jest dobrze określona (całkujemy tylko po przedziale skończonym), nieujemna i gładka. Mamy $\varphi_2 \equiv 0$ na $(-\infty, -1]$ i $\varphi_2(x) \equiv c := \int_{-1}^1 \varphi_1$ dla $x \geq 1$. Na przedziale $[-1, 1]$ funkcja φ_2 jest rosnąca.

Teraz wykorzystamy przesuwanie, skalowanie i mnożenie funkcji gładkich.

Krok 3. Ustalmy $R > r > 0$. Dobierzmy $a > 0$ tak, żeby $\frac{a+2}{a} = \frac{R}{r}$. Funkcja

$$\varphi_3(x) = \frac{1}{c^2} \varphi_2(1+a+x) \varphi_2(1+a-x), \quad x \in \mathbb{R},$$

jest gładka, znika poza przedziałem $[-a-2, a+2]$ i jest równa 1 na przedziale $[-a, a]$ (patrz załączony rysunek). Funkcja

$$\varphi_4(x) = \varphi_3\left(\frac{(a+2)x}{R}\right)$$

też jest gładka. Wobec doboru a , $\varphi_4 \equiv 1$ na $[-r, r]$ i $\varphi_4 \equiv 0$ poza przedziałem $[-R, R]$.

Krok 4. Funkcja $f_1(\mathbf{x}) = \varphi_4(\|\mathbf{x}\|)$ spełnia warunki twierdzenia dla $\mathbf{a} = 0$. (Zauważmy, że dla $\|\mathbf{x}\| < r$ funkcja f ma stałą wartość 1, więc jej pochodne cząstkowe znikają w punktach kuli otwartej $B(0, r)$. Norma $\|\mathbf{x}\| = (x_1^2 + \dots + x_n^2)^{1/2}$ jest funkcją gładką na zbiorze $\{\mathbf{x} : \|\mathbf{x}\| > r/2\}$, i dlatego f jest gładka na całej przestrzeni $\mathbb{R}^n$.) Przesuwając f_1 , tzn. biorąc $f(\mathbf{x}) = f_1(\mathbf{x} - \mathbf{a})$, kończymy dowód w ogólnym przypadku. $\square$

Uwaga 2.74. Zbiór wszystkich funkcji gładkich o zwartym nośniku w $\mathbb{R}^n$ oznacza się symbolem $C_0^\infty(\mathbb{R}^n)$.

Samodzielne rozwiązanie poniższych zadań pozwoli Czytelnikowi lepiej oswoić się z pojęciem funkcji gładkiej.

Zadanie 2.75. Niech $K \subset \Omega \subset \mathbb{R}^n$. Załóżmy, że zbiór K jest zwarty, a zbiór Ω jest otwarty. Wykazać, że istnieje funkcja $f \in C_0^\infty(\mathbb{R}^n)$ taka, że $f \equiv 1$ na K i $\text{supp } f \subset \Omega$.

Zadanie 2.76. Niech F będzie dowolnym zbiorem domkniętym w $\mathbb{R}^n$. Istnieje wówczas funkcja $f \in C^\infty(\mathbb{R}^n)$ taka, że $f \geq 0$ i $F = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = 0\}$.