Rozdział 7

FUNKCJE WIELU ZMIENNYCH

Istniejące podręczniki

H. Margenau, G.M. Murphy: *Matematyka w fizyce i chemii*, rozdział 4, PWN, Warszawa 1962;
W. Kołodziej: *Analiza matematyczna*, rozdział 4, PWN, Warszawa 1978;
J.E. Marsden, A.J. Tromba *Vector Calculus*, Freeman and Company, New York 1988, rozdział 4.

7.1 Przestrzeń $\mathbb{R}^n$

Zbiór $\mathbb{R}^n$ jest *iloczynem kartezjańskim* n kopii zbioru $\mathbb{R}$, co się oznacza $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}$. Jego elementami są *ciągi n-elementowe* liczb rzeczywistych:

$$\mathbf{x} \in \mathbb{R}^n, \quad \mathbf{x} = \{x_1, \ldots, x_n\}, \quad x_k \in \mathbb{R}$$

zwane *wektorami* lub *punktami*. Umiemy je sobie wyobrazić na płaszczyźnie (wtedy $n = 2$) lub w przestrzeni trójwymiarowej (wtedy $n = 3$) jako odcinki zaczepione w jednym wspólnym punkcie, który nazywamy *początkiem układu współrzędnych*. Wtedy wektorem jest cały odcinek, a punktem jest drugi koniec odcinka (jeśli pierwszym końcem jest ten zaczepiony w początku układu).

W zbiorze $\mathbb{R}^n$ jest określone *dodawanie*: jeśli $\mathbf{x} = \{x_1, \ldots, x_n\}$, $\mathbf{y} = \{y_1, \ldots, y_n\}$, to

$$\mathbf{x} + \mathbf{y} = \{x_1 + y_1, \ldots, x_n + y_n\} \tag{7.1}$$

oraz *mnożenie przez skalary* $\alpha \in \mathbb{R}$:

$$\alpha\mathbf{x} = \{\alpha x_1, \ldots, \alpha x_n\}. \tag{7.2}$$

Jeśli te dwa działania spełniają pewne własności podane na wykładzie z algebry, to zbiór $\mathbb{R}^n$ nazywa się *przestrzenią wektorową*, *przestrzenią liniową* lub *linealem*. Własności owe można sprawdzić na podstawie podanych definicji obu działań, więc $\mathbb{R}^n$ jest linealem. W lineale $\mathbb{R}^n$ można wyróżnić n wektorów o wpółrzędnych składających się z zer i jedynek, przy czym w każdym z nich jest tylko jedna jedynka:

$$\mathbf{e}_1 = \{1, 0, 0, \ldots, 0, 0\}, \quad \mathbf{e}_2 = \{0, 1, 0, \ldots, 0, 0\}, \quad \ldots \quad, \mathbf{e}_n = \{0, 0, 0, \ldots, 0, 1\}. \tag{7.3}$$

Mają one tę własność, że dowolny wektor $\mathbf{x} \in \mathbb{R}^n$ da się zapisać jako kombinacja liniowa

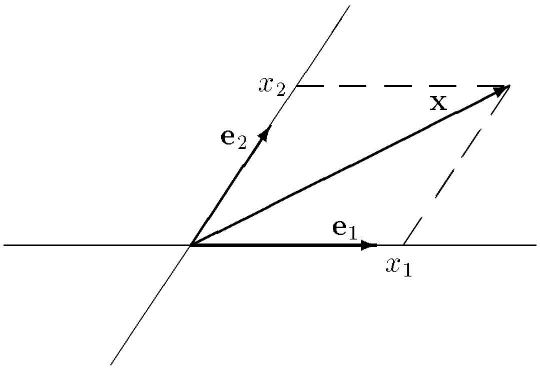
$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{e}_i. \tag{7.4}$$

Z tego powodu układ wektorów $\{\mathbf{e}_i\}$ jest *bazą*, a x_i są *współrzędnymi wektora* (lub *punktu*) $\mathbf{x}$ w tej bazie. Wzór (7.4) w trzech wymiarach przyjmuje postać

$$\{x_1, x_2, x_3\} = x_1 \{1,0,0\} + x_2 \{0,1,0\} + x_3 \{0,0,1\}.$$

Układ wektorów (7.3) nazywa się *bazą kanoniczną*.

Jeśli wprowadzić n prostych przechodzących przez początek układu współrzędnych, przy czym te proste nie leżą w jednej podprzestrzeni o wymiarze mniejszym niż n, a na każdej z tych prostych umieścimy po jednym wektorze $\mathbf{e}_i$, to nazywają się one *osiami układu współrzędnych*. Wtedy i-tą wpółrzędną wektora $\mathbf{x}$ znajduje się przez rzutowanie tego wektora na i-tą prostą równolegle do podprzestrzeni wyznaczonej przez pozostałe proste i znalezienie liczby x_i takiej, że ten rzut jest iloczynem $x_i \mathbf{e}_i$, patrz Rys. 7.1. Ujemna liczba x_i oznacza, że rzut wektora $\mathbf{x}$ na i-tą prostą ma zwrot przeciwny do $\mathbf{e}_i$.



Rysunek 7.1

W lineale $\mathbb{R}^n$ wprowadza się *iloczyn skalarny* wektorów

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i. \tag{7.5}$$

(Będę też nazywać go *skaloczynem*). Można sprawdzić, że tak określony iloczyn skalarny ma następujące własności:

1. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n: \quad \mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ – symetryczność,
2. $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n: \quad \mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$ – rozdzielność względem dodawania,
3. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,\ \alpha \in \mathbb{R}: \quad \mathbf{x} \cdot (\alpha \mathbf{y}) = \alpha(\mathbf{x} \cdot \mathbf{y})$ – jednorodność,
4. $\forall \mathbf{x} \in \mathbb{R}^n: \quad \mathbf{x} \cdot \mathbf{x} \geq 0$ – nieujemność,
5. $\mathbf{x} \cdot \mathbf{x} = 0 \quad \Rightarrow \quad \mathbf{x} = \mathbf{0}$ – niezwyrodniałość.

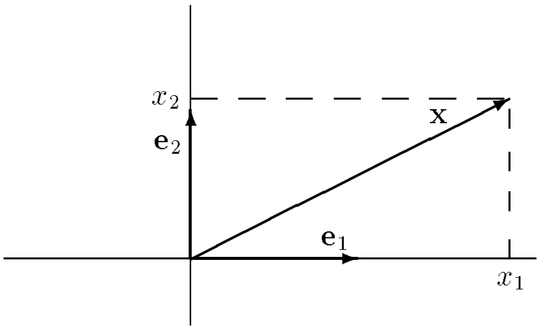
Za pomocą skaloczynu określa się *prostopadłość* dwóch wektorów:

$$\mathbf{x} \perp \mathbf{y} \quad \text{w.t.w. gdy} \quad \mathbf{x} \cdot \mathbf{y} = 0.$$

Można sprawdzić według definicji skaloczynu (7.5), że różne wektory bazowe (7.3) są prostopadłe $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ dla $i \neq j$, a każdy z nich ma kwadrat skalarny jeden: $\mathbf{e}_i \cdot \mathbf{e}_i = 1$. Obie te własności można zawrzeć w jednym wzorze: $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$, gdzie

$$\delta_{ij} = \begin{cases} 0 & \text{gdy} \quad i \neq j, \\ 1 & \text{gdy} \quad i = j. \end{cases}$$

nazywa się *deltą Kroneckera*. Stąd wynika, że po wprowadzeniu skaloczynu według wzoru (7.5) osie układu współrzędnych muszą być prostopadłe, a wybrane na nich wektory bazowe muszą być jednostkowe. W takim razie Rys. 7.1 powinien być teraz zastąpiony przez Rys. 7.2. [Leopold Kronecker (1823-1891) – niemiecki matematyk urodzony w Legnicy, autor prac z algebry i teorii liczb, zwolennik sprowadzenia matematyki do arytmetyki liczb naturalnych.]



Rysunek 7.2

Twierdzenie 1. Dla dowolnych $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ zachodzi nierówność

$$|\mathbf{x} \cdot \mathbf{y}| \leq (\mathbf{x} \cdot \mathbf{x})^{1/2} (\mathbf{y} \cdot \mathbf{y})^{1/2}$$

zwana *nierównością Schwarza*. [Hermann Amandus Schwarz (1843–1921) – matematyk niemiecki urodzony w Przewoźnikach koło Żar; główne jego prace są poświęcone analizie matematycznej i jej zastosowaniom w geometrii.]

Dowód. Z własności nieujemności skaloczynu dla dowolnych $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ i dowolnych $\alpha, \beta \in \mathbb{R}$ możemy napisać

$$(\alpha \mathbf{x} + \beta \mathbf{y}) \cdot (\alpha \mathbf{x} + \beta \mathbf{y}) \geq 0.$$

Korzystamy z jednorodności i rozdzielności względem dodawania:

$$\alpha^2 (\mathbf{x} \cdot \mathbf{x}) + 2\alpha\beta (\mathbf{x} \cdot \mathbf{y}) + \beta^2 (\mathbf{y} \cdot \mathbf{y}) \geq 0.$$

Teraz wybieramy $\alpha = \mathbf{y} \cdot \mathbf{y}$, $\beta = -\mathbf{x} \cdot \mathbf{y}$:

$$(\mathbf{y} \cdot \mathbf{y})^2 (\mathbf{x} \cdot \mathbf{x}) - 2(\mathbf{y} \cdot \mathbf{y})(\mathbf{x} \cdot \mathbf{y})^2 + (\mathbf{x} \cdot \mathbf{y})^2 (\mathbf{y} \cdot \mathbf{y}) \geq 0,$$

$$(\mathbf{y} \cdot \mathbf{y})^2 (\mathbf{x} \cdot \mathbf{x}) - (\mathbf{y} \cdot \mathbf{y})(\mathbf{x} \cdot \mathbf{y})^2 \geq 0,$$

$$(\mathbf{y} \cdot \mathbf{y})(\mathbf{x} \cdot \mathbf{x}) - (\mathbf{x} \cdot \mathbf{y})^2 \geq 0,$$

$$(\mathbf{y} \cdot \mathbf{y})(\mathbf{x} \cdot \mathbf{x}) \geq (\mathbf{x} \cdot \mathbf{y})^2,$$

$$(\mathbf{x}\cdot\mathbf{y})^2 \leq (\mathbf{x}\cdot\mathbf{x})(\mathbf{y}\cdot\mathbf{y}).$$

Skoro pierwiastek jest funkcją rosnącą, to po spierwiastkowaniu zachowany jest zwrot nierówności:

$$|\mathbf{x}\cdot\mathbf{y}| \leq (\mathbf{x}\cdot\mathbf{x})^{1/2}(\mathbf{y}\cdot\mathbf{y})^{1/2}. \blacksquare$$

Po wprowadzeniu iloczynu skalarnego w lineale $\mathbb{R}^n$ definiuje się *normę skaloczynową*, która zadaje *długość* wektora $\mathbf{x}$:

$$\|\mathbf{x}\| = (\mathbf{x}\cdot\mathbf{x})^{1/2}. \tag{7.6}$$

Za pomocą tej normy można nierówność Schwarza zapisać inaczej

$$|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\|\,\|\mathbf{y}\|. \tag{7.7}$$

Z wzorów (7.6) i (7.5) wynika, że norma wektora wyraża się przez jego współrzędne wzorem $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$, a stąd wynika nierówność

$$\forall j \in \{1, 2, \ldots n\}: \quad |x_j| \leq \|\mathbf{x}\|. \tag{7.8}$$

Norma ma następujące własności

1. $\forall\, \alpha \in \mathbb{R},\ \mathbf{x} \in \mathbb{R}^n: \quad \|\alpha\mathbf{x}\| = |\alpha|\ \|\mathbf{x}\|$ – dodatnia jednorodność,
2. $\forall\, \mathbf{x} \in \mathbb{R}^n: \ \|\mathbf{x}\| \geq 0$ – nieujemność,
3. $\|\mathbf{x}\| = 0 \quad\Rightarrow\quad \mathbf{x} = \mathbf{0}$ – niezwyrodniałość,
4. $\forall \mathbf{x},\mathbf{y} \in \mathbb{R}^n \quad \|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ – nierówność trójkąta.

Te wszystkie własności można wyprowadzić z podanych własności iloczynu skalarnego. My sprawdzimy jedynie własność 4. Korzytając z definicji normy (7.6) liczymy kwadrat lewej strony:

$$\|\mathbf{x}+\mathbf{y}\|^2 = (\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y}) = (\mathbf{x}\cdot\mathbf{x}) + (\mathbf{y}\cdot\mathbf{y}) + 2(\mathbf{x}\cdot\mathbf{y}) = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2(\mathbf{x}\cdot\mathbf{y}).$$

Korzystamy z nierówności Schwarza (7.7):

$$\|\mathbf{x}+\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\,\|\mathbf{x}\|\,\|\mathbf{y}\| = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2.$$

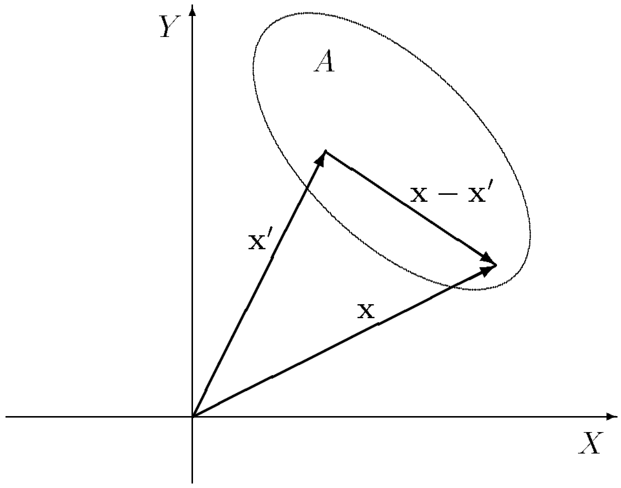
Skoro pierwiastek jest funkcją rosnącą, to po spierwiastkowaniu zachowany jest zwrot nierówności:

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|. \blacksquare$$

Mając do dyspozycji pojęcie normy możemy wprowadzić wiele następnych pojęć. *Odległością* między punktami $\mathbf{x}$ i $\mathbf{y}$ nazywamy

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|.$$

Niech $A \subset \mathbb{R}^n$; liczbę $\operatorname{diam} A = \sup\{\|\mathbf{x}-\mathbf{x}'\| : \mathbf{x},\mathbf{x}' \in A\}$ nazywamy *średnicą zbioru* A. Zbiór A nazywamy *ograniczonym*, jeśli jego średnica jest skończona. Taki zbiór jest ukazany na Rys. 7.3.



Rysunek 7.3

Zbiór

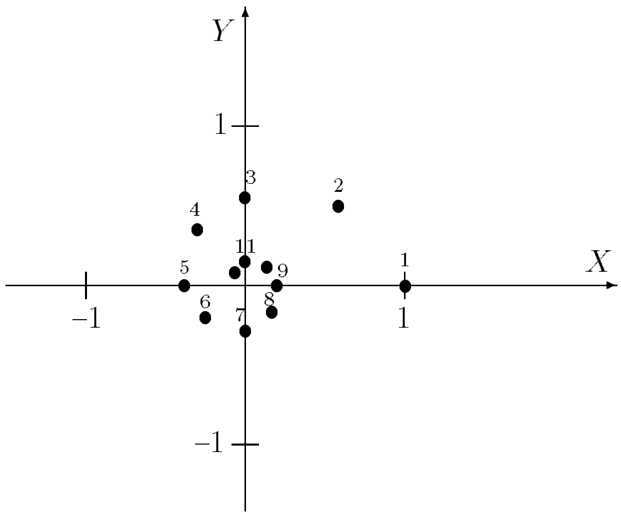
$$K(\mathbf{a}, r) = \{\mathbf{x} \in \mathbb{R}^n : ||\, \mathbf{x} - \mathbf{a}\, || < r\} \tag{7.9}$$

nazywamy *niby-kulą otwartą o środku* $\mathbf{a}$ *i promieniu* r. Dla $n = 1$, czyli dla prostej ta niby-kula jest odcinkiem otwartym o długości $2r$, dla $n = 2$ czyli w przypadku płaszczyzny niby-kula otwarta jest kołem bez brzegu, dla $n = 3$ niby-kula otwarta jest wnętrzem kuli. Zbiór

$$B(\mathbf{a}, r) = \{\mathbf{x} \in \mathbb{R}^n : ||\, \mathbf{x} - \mathbf{a}\, || \leq r\} \tag{7.10}$$

nazywamy *niby-kulą domkniętą o środku* $\mathbf{a}$ *i promieniu* r.

Jeśli każdej liczbie naturalnej jest przyporządkowany wektor, to mówimy, że został określony *ciąg wektorów* $\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_k, \ldots$ lub $\{\mathbf{x}_k\}$. Ilustrujemy to na Rys. 7.4, który odpowiada Rys. 2.2 z rozdziału o ciągach liczbowych. Teraz ciąg jest umieszczony na płaszczyźnie zamiast na linii prostej. Kropkami zaznaczono punkty odpowiadające poszczególnym elementom ciągu. Malutkie cyferki oznaczają numery tych elementów. Ma to być ciąg zbieżny do zera, czyli do początku układu współrzędnych.



Rysunek 7.4

Ciąg wektorów $\{\mathbf{x}_k\}$ nazywamy *zbieżnym* do $\mathbf{a} \in \mathbb{R}^n$, jeśli

$$\forall \epsilon > 0\ \exists N \in \mathbb{N}\ \forall k > N : ||\ \mathbf{x}_k - \mathbf{a}\ || < \epsilon. \tag{7.11}$$

Warunek ten można zapisać inaczej z użyciem niby-kuli otwartej:

$$\forall \epsilon > 0\ \exists N \in \mathbb{N}\ \forall k > N :\ \mathbf{x}_k \in K(\mathbf{a}, \epsilon). \tag{7.12}$$

Jeśli ten warunek będzie spełniony, to będziemy pisać

$$\lim_{k\to\infty} \mathbf{x}_k = \mathbf{a} \qquad \text{lub} \qquad \mathbf{x}_k \to \mathbf{a}. \tag{7.13}$$

Dla nas wektor jest n-ką liczb zwanych jego współrzędnymi $\mathbf{x} = \{x_1, \ldots, x_n\}$, więc dla ciągu wektorów współrzędne muszą mieć jeszcze jeden wskaźnik numerujący element ciągu. Jeżeli lineał jest dwuwymiarowy albo trójwymiarowy, to można tego uniknąć używając różnych liter dla różnych współrzędnych, a wtedy potrzebny jest tylko jeden wskaźnik numerujący element ciągu:

$$\begin{array}{ll}
\mathbf{r}_1 = \{x_1, y_1\} & \qquad \mathbf{r}_1 = \{x_1, y_1, z_1\}, \\
\mathbf{r}_2 = \{x_2, y_2\} & \qquad \mathbf{r}_2 = \{x_2, y_2, z_2\}, \\
\mathbf{r}_3 = \{x_3, y_3\} & \qquad \mathbf{r}_3 = \{x_3, y_3, z_3\}, \\
\quad \ldots\ldots & \qquad \quad \ldots\ldots\ldots \\
\mathbf{r}_k = \{x_k, y_k\} & \qquad \mathbf{r}_k = \{x_k, y_k, z_k\}, \\
\quad \ldots\ldots & \qquad \quad \ldots\ldots\ldots
\end{array}$$

Dla ogólnego n nie możemy już uniknąć dwóch wskaźników. Umówimy się, że pierwszy wskaźnik jest numerem współrzędnej, a drugi numeruje kolejny element ciągu:

$$\begin{array}{c}
\mathbf{x}_1 = \{x_{1,1},\ x_{2,1},\ x_{3,1},\ \ldots,\ x_{n,1}\} \\
\mathbf{x}_2 = \{x_{1,2},\ x_{2,2},\ x_{3,2},\ \ldots,\ x_{n,2}\} \\
\mathbf{x}_3 = \{x_{1,3},\ x_{2,3},\ x_{3,3},\ \ldots,\ x_{n,3}\} \\
\ldots\ldots\ldots\ldots \\
\mathbf{x}_k = \{x_{1,k},\ x_{2,k},\ x_{3,k},\ \ldots,\ x_{n,k}\} \\
\ldots\ldots\ldots\ldots
\end{array}$$

Twierdzenie 2. Ciąg $\{\mathbf{x}_k\}$ jest zbieżny do $\mathbf{a}$ w $\mathbb{R}^n$ w.t.w. gdy dla każdego $j \in \{1, 2, ..., n\}$ ciąg j-tych współrzędnych $\{x_{j,k}\}_{k=1}^{\infty}$ jest zbieżny do a_j w $\mathbb{R}$. Symbolicznie

$$\lim_{k\to\infty} \mathbf{x}_k = \mathbf{a} \quad \Leftrightarrow \quad \forall j \in \{1, 2, ..., n\} \quad \lim_{k\to\infty} x_{j,k} = a_j.$$

Dowód $\Rightarrow$. Załóżmy, że ciąg wektorów $\{\mathbf{x}_k\}$ jest zbieżny do wektora $\mathbf{a}$. Z definicji 7.11 to znaczy, że dla dowolnego $\epsilon > 0$ istnieje $N \in \mathbb{N}$ takie, że dla każdego $k > N$ zachodzi

$\| \mathbf{x}_k - \mathbf{a} \| < \epsilon$. Według wzoru (7.8) dla każdego $j \in \{1, 2, ..., n\}$ mamy $|x_{j,k} - a_j| \leq \| \mathbf{x}_k - \mathbf{a} \|$, więc dodając wszystkie wspomniane kwantyfikatory możemy napisać

$$\forall \epsilon > 0 \ \ \exists N \in \mathbb{N} \ \ \forall k > N \ \ \forall j \in \{1, 2, ...n\} : \ |x_{j,k} - a_j| < \epsilon,$$

a to oznacza, że ciąg j-tych współrzędnych $\{x_{j,k}\}_{k=1}^{\infty}$ jest zbieżny do a_j. □

$\Leftarrow$. Załóżmy, że dla każdego $j \in \{1, 2, ..., n\}$ ciąg $\{x_{j,k}\}_{k=1}^{\infty}$ jest zbieżny do a_j, tzn.

$$\forall j \in \{1, 2, ...n\} \ \ \forall \epsilon > 0 \ \ \exists N_j \in \mathbb{N} \ \ \forall k > N_j : \ |x_{j,k} - a_j| < \frac{\epsilon}{\sqrt{n}}.$$

Przyjęliśmy N_j zależne od j, gdyż w założeniu jest zbieżność dla każdej j-tej współrzędnej z osobna. Jeśli wybierzemy $N = \max\{N_1, N_2, ..., N_n\}$, to będziemy mogli napisać zdanie z kwantyfikatorem zawierającym j na końcu:

$$\forall \epsilon > 0 \ \ \exists N \in \mathbb{N} \ \ \forall k > N \ \ \forall j \in \{1, 2, ...n\} : \ |x_{j,k} - a_j| < \frac{\epsilon}{\sqrt{n}}.$$

Z umieszczonej na końcu nierówności wynika taka

$$\| \mathbf{x}_k - \mathbf{a} \| = \sqrt{|x_{1,k} - a_1|^2 + |x_{2,k} - a_2|^2 + \ \ldots + |x_{n,k} - a_n|^2} <$$

$$< \sqrt{\frac{\epsilon^2}{n} + \frac{\epsilon^2}{n} + \ \ldots + \frac{\epsilon^2}{n}} = \sqrt{\epsilon^2} = \epsilon.$$

W takim razie mamy prawdziwe zdanie

$$\forall \epsilon > 0 \ \exists N \in \mathbb{N} \ \forall k > N : \| \mathbf{x}_k - \mathbf{a} \| < \epsilon.$$

A to jest zdanie (7.11), więc ciąg wektorów $\mathbf{x}_k$ jest zbieżny do $\mathbf{a}$. ■

Zbiór $V \subset \mathbb{R}$ nazywa się *domknięty*, jeśli dla ciągu $\{\mathbf{x}_n\} \subset V$ z warunku $\mathbf{x}_k \to \mathbf{x}$ wynika $\mathbf{x} \in V$, czyli granice ciągów z V też należą do V.

PRZYKŁAD 1. Sprawdzimy, czy niby-kula B określona wzorem (7.10) jest domknięta w sensie tej definicji, czyli zasługuje na podaną nazwę. Jeśli $\mathbf{x}_k \in B(\mathbf{a}, r)$ oraz $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}$, to na podstawie nierówności trójkąta możemy napisać

$$\| \mathbf{x} - \mathbf{a} \| = \| \mathbf{x} - \mathbf{x}_k + \mathbf{x}_k - \mathbf{a} \| \leq \| \mathbf{x} - \mathbf{x}_k \| + \| \mathbf{x}_k - \mathbf{a} \| .$$

Prawy składnik tej sumy jest nie większy od r, a lewy przy dowolnym $\epsilon > 0$ dla dostatecznie dużych k jest mniejszy od ϵ, więc mamy

$$\| \mathbf{x} - \mathbf{a} \| < r + \epsilon$$

dla dowolnego $\epsilon > 0$. Stąd wynika nierówność

$$\| \mathbf{x} - \mathbf{a} \| \leq r,$$

która mówi, że wektor $\mathbf{x}$ należy do niby-kuli B. ■

Zbiór $U \subset \mathbb{R}^n$ nazywa się *otwarty*, jeśli

$$\forall \mathbf{x} \in U \quad \exists \delta > 0 : K(\mathbf{x}, \delta) \subset U. \tag{7.14}$$

Innymi słowy, zbiór jest otwarty, jeśli wraz z dowolnym punktem zawiera też pewną niby-kulę otwartą.

PRZYKŁAD 2. Sprawdzimy, że niby-kula K określona wzorem (7.9) jest otwarta w sensie tej definicji. Weźmy dowolny punkt $\mathbf{x} \in K(\mathbf{a}, r)$. Na mocy (7.9) zachodzi nierówność $\| \mathbf{x} - \mathbf{a} \| < r$. Wybieramy

$$\delta = r - \| \mathbf{x} - \mathbf{a} \| . \tag{7.15}$$

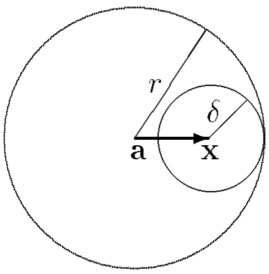
Rys. 7.5 ilustruje to w dwóch wymiarach. Twierdzimy, że zachodzi $K(\mathbf{x}, \delta) \subset K(\mathbf{a}, r)$. Weźmy punkt $\mathbf{x}' \in K(\mathbf{x}, \delta)$, co oznacza $\| \mathbf{x}' - \mathbf{x} \| < \delta$. Z nierówności trójkąta mamy

$$\| \mathbf{x}' - \mathbf{a} \| \leq \| \mathbf{x}' - \mathbf{x} \| + \| \mathbf{x} - \mathbf{a} \| < \delta + \| \mathbf{x} - \mathbf{a} \| .$$

Korzystamy z (7.15):

$$\| \mathbf{x}' - \mathbf{a} \| < r.$$

To oznacza, że $\mathbf{x}' \in K(\mathbf{a}, r)$. Pokazaliśmy wynikanie $\mathbf{x}' \in K(\mathbf{x}, \delta) \Rightarrow \mathbf{x}' \in K(\mathbf{a}, r)$, co dla dowolnego $\mathbf{x} \in K(\mathbf{a}, r)$ oznacza zawieranie zbiorów $K(\mathbf{x}, \delta) \subset K(\mathbf{a}, r)$. Zdanie (7.14) jest prawdziwe, więc niby-kula $K(\mathbf{a}, r)$ może być nazywana otwartą.



Rysunek 7.5

Twierdzenie 3. Jeśli ciągi $\{\mathbf{x}_k\}$ i $\{\mathbf{y}_k\}$ są zbieżne, to suma i różnica tych ciągów jest zbieżna oraz zachodzą wzory:

$$\lim_{k\to\infty} (\mathbf{x}_k + \mathbf{y}_k) = \lim_{k\to\infty} \mathbf{x}_k + \lim_{k\to\infty} \mathbf{y}_k,$$

$$\lim_{k\to\infty} (\mathbf{x}_k - \mathbf{y}_k) = \lim_{k\to\infty} \mathbf{x}_k - \lim_{k\to\infty} \mathbf{y}_k,$$

tzn. granica sumy jest sumą granic, a granica różnicy jest różnicą granic.

Dowód. Udowodnimy równość pierwszą. Wybierzmy liczbę $\epsilon > 0$. Niech $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{g}$ i $\lim_{k\to\infty} \mathbf{y}_k = \mathbf{h}$. To oznacza, że istnieje takie $N \in \mathbb{N}$, że dla $k > N$ zachodzą nierówności

$$||\mathbf{x}_k - \mathbf{g}|| < \frac{1}{2}\epsilon \quad \text{oraz} \quad ||\mathbf{y}_k - \mathbf{h}|| < \frac{1}{2}\epsilon.$$

Z nierówności trójkąta

$$||(\mathbf{x}_k + \mathbf{y}_k) - (\mathbf{g} + \mathbf{h})|| = ||(\mathbf{x}_k - \mathbf{g}) + (\mathbf{y}_k - \mathbf{h})|| \leq ||\mathbf{x}_k - \mathbf{g}|| + ||\mathbf{y}_k - \mathbf{h}|| < \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon.$$

Tym samym pokazaliśmy

$$||(\mathbf{x}_k + \mathbf{y}_k) - (\mathbf{g} + \mathbf{h})|| < \epsilon$$

dla $n > N$. To znaczy, że suma ciągów $\{\mathbf{x}_k\}$ i $\{\mathbf{y}_k\}$ jest zbieżna, a jej granicą jest suma granic $\mathbf{g} + \mathbf{h} = \lim_{k\to\infty} \mathbf{x}_k + \lim_{k\to\infty} \mathbf{y}_k$. Pierwsza równość jest udowodniona. ■

Twierdzenie 4. Jeśli ciąg $\{\mathbf{x}_k\}$ jest zbieżny, to dla dowolnego $c \in \mathbb{R}$ ciąg $\{c\,\mathbf{x}_k\}$ jest zbieżny oraz zachodzi wzór:

$$\lim_{k\to\infty} (c\,\mathbf{x}_k) = c \lim_{k\to\infty} \mathbf{x}_k$$

tzn. liczbę stałą można wyciągnąć przed znak granicy.

Dowód. Wybierzmy liczbę $\epsilon > 0$. Niech $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{g}$. To oznacza, że istnieje takie $N \in \mathbb{N}$, że dla $k > N$ zachodzi nierówność

$$||\mathbf{x}_k - \mathbf{g}|| < \frac{\epsilon}{|c|}.$$

Z własności normy i z tej nierówności

$$||c\mathbf{x}_k - c\mathbf{g}|| = ||c\,(\mathbf{x}_k - \mathbf{g})|| = |c|\,||\mathbf{x}_k - \mathbf{g}|| < |c|\,\frac{\epsilon}{|c|} = \epsilon.$$

Tym samym pokazaliśmy

$$||c\mathbf{x}_k - c\mathbf{g}|| < \epsilon$$

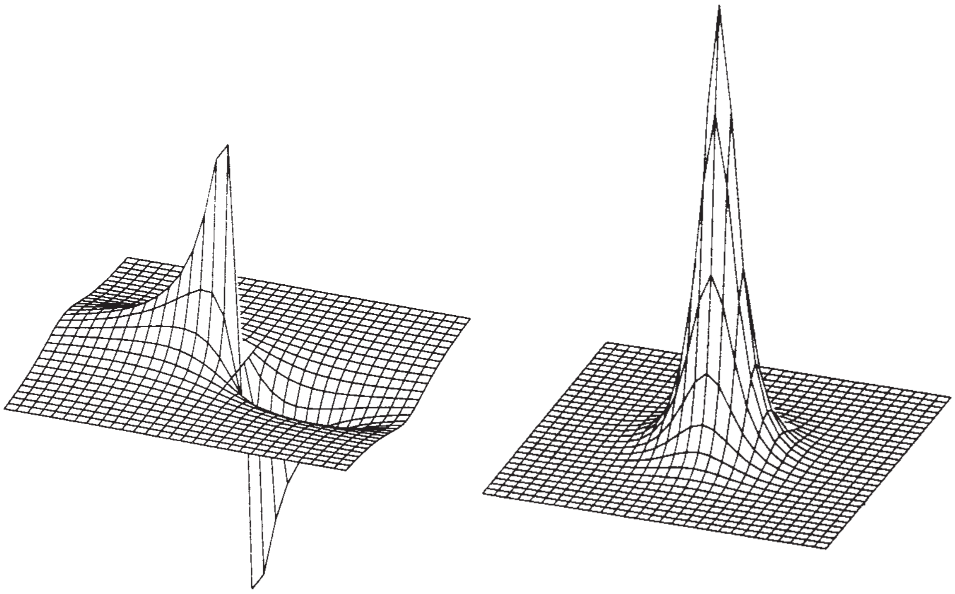
dla $k > N$. To znaczy, że ciąg $\{c\,\mathbf{x}_k\}$ jest zbieżny, a jego granicą jest $c\,\mathbf{g} = c \lim_{k\to\infty} \mathbf{x}_k$ ■

7.2 Funkcje o wartościach rzeczywistych

Przez *funkcję wielu zmiennych o wartościach rzeczywistych* i o dziedzinie A będziemy rozumieć odwzorowanie $f : A \to \mathbb{R}$, jeśli $A \subset \mathbb{R}^n$. Zbiór par uporządkowanych $(\mathbf{x},\, f(\mathbf{x}))$ nazywa się *wykresem funkcji f*; jego oznaczenie

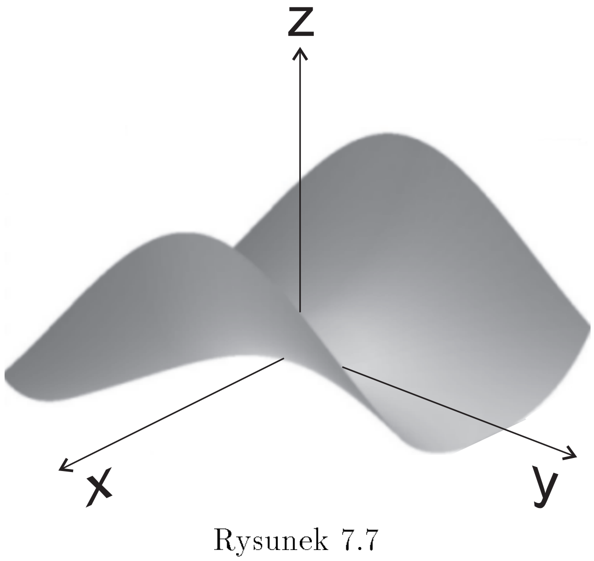
$$\mathrm{graf}\, f = \{(\mathbf{x},\, f(\mathbf{x})) : \mathbf{x} \in A\}.$$

Wykres jest podzbiorem lineału $\mathbb{R}^{n+1}$. Tak więc dla sporządzenia wykresu funkcji jednej zmiennej potrzebna jest płaszczyzna, a dla funkcji dwóch zmiennych – przestrzeń trójwymiarowa. Na ogół dziedzinę funkcji dwóch zmiennych umieszcza się w płaszczyźnie (X, Y), a wartości funkcji – na osi Z. Wtedy wykresem funkcji jest jakaś powierzchnia (zob. Rys. 7.6, na którym ukazany jest wykres wzięty z publikacji fizycznej).



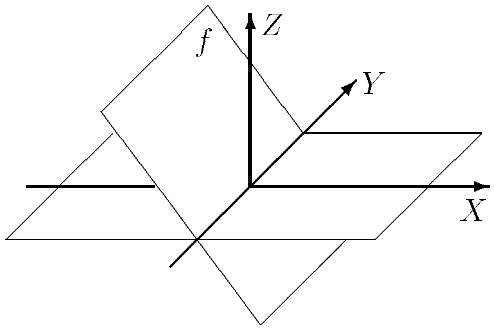
Rysunek 7.6

PRZYKŁAD 1. Wykresem funkcji $f(x, y) = x^2 - y^2$ jest powierzchnia siodłowa ukazana na Rys. 7.7. Przecięcie wykresu płasczyzną (X, Z) oznacza przyrównanie y do zera, co daje funkcję jednej zmiennej $f(x, 0) = x^2$; jej wykresem jest parabola o ramionach zwrócnych ku górze. Przecięcie wykresu płaszczyzną (Y, Z) oznacza $x = 0$; wtedy wykresem funkcji jednej zmiennej $f(0, y) = -y^2$ jest parabola o ramionach zwróconych ku dołowi. Przy pewnej spostrzegawczości obie te parabole można zobaczyć na Rys. 7.7.

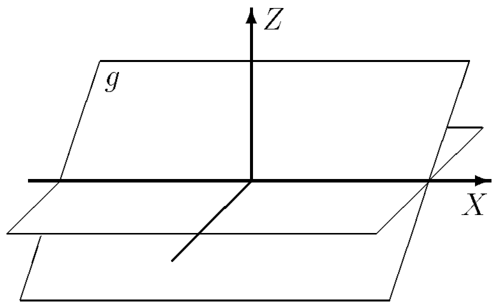


Rysunek 7.7

PRZYKŁAD 2. Wykresem funkcji $f(x, y) = -x$ jest płaszczyzna przechodząca przez oś Y i nachylona do płaszczyzny (X, Y) pod kątem 45^0 (zob. Rys. 7.8), a wykresem funkcji $g(x, y) = y$ jest płaszczyzna przechodząca przez oś X i też nachylona do płaszczyzny (X, Y) pod kątem 45^0 (zob. Rys. 7.9).



Rysunek 7.8



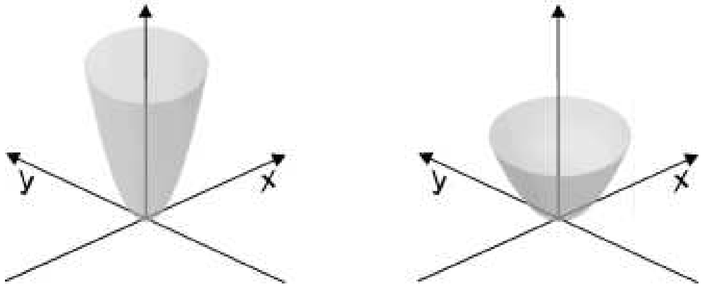
Rysunek 7.9

Jeśli $c \in \mathbb{R}$ jest ustaloną liczbą, to podzbiór dziedziny $L_c \subset D_f$ określony jako

$$L_c = \{\mathbf{x} \in D_f : f(\mathbf{x}) = c\}$$

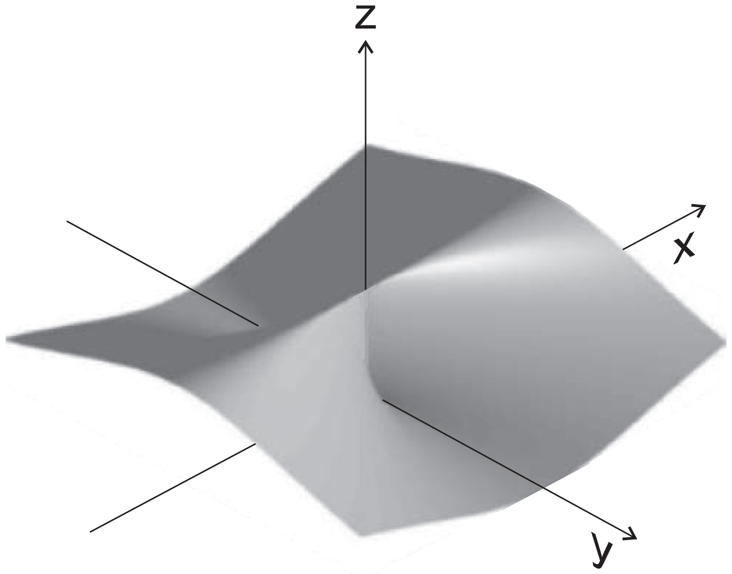
nazywamy *poziomicą funkcji f o wartości c*. Zauważmy, że do wykreślania poziomic funkcji nie jest potrzebny lineał o wymiar wyższy, bo wykreślamy je w samej dziedzinie. Dla funkcji jednej zmiennej poziomicami są punkty, dla funkcji dwóch zmiennych – linie krzywe. Na przykład dla ukazanej na Rys. 7.8 funkcji f poziomicami są linie proste równoległe do osi Y, natomiast dla funkcji g z Rys. 7.9 poziomicami są proste równoległe do osi X. Gdyby wybierać poziomice w jednakowych odstępach wartości funkcji, np. $c_k = k,\ k \in \mathbb{Z}$, to odległości między poziomicami są miarą tego, jak stromy jest wykres. Dla podanych dwóch przykładów odległości między takimi poziomicami są stałe, co dowodzi tego, że wykresami są płaszczyzny.

PRZYKŁAD 3. Dla funkcji $h(x, y) = x^2 + y^2$ wykresem jest paraboloida obrotowa, bo zawężenie tej funkcji do jednej zmiennej $h(x, 0) = x^2$ daje na płaszczyźnie (X, Z) parabolę, a wzór określający całą funkcję h ma symetrię obrotową wokół osi Z. Na Rys. 7.10 ukazano tę paraboloidę dwa razy. Po lewej stronie większą jej część, a po prawej stronie jej część bizszą wierzchołka w pewnym powiększeniu. Poziomicami tej funkcji są okręgi wyznaczone przez równania $x^2 + y^2 = c$ o promieniach $r = \sqrt{x^2 + y^2} = \sqrt{c}$. Tym razem odstępy między poziomicami nie są stałe – maleją wraz ze wzrostem r, bo parabola jest coraz bardziej stroma w miarę oddalania się od jej wierzchołka.



Rysunek 7.10

PRZYKŁAD 4. Funkcja dana wzorem $f(x, y) = \frac{x^2}{x^2+y^2}$ o dziedzinie $\mathbb{R}^2 \setminus (0, 0)$ ma wykres ukazany na Rys. 7.11.



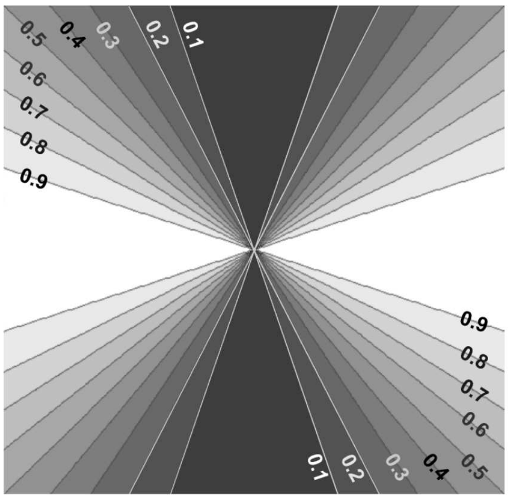
Rysunek 7.11

Aby znaleźć jej poziomice, przyrównajmy wartość funkcji do stałej $c^2 \leq 1$ i przekształcajmy otrzymane równanie:

$$\frac{x^2}{x^2+y^2} = c^2, \qquad x^2 = c^2\left(x^2+y^2\right),$$

$$y^2 = \frac{1-c^2}{c^2}x^2, \qquad y = \pm\frac{\sqrt{1-c^2}}{|c|}x.$$

Widzimy, że poziomicami są proste przechodzące przez początek układu, patrz Rys. 7.12.



Rysunek 7.12

Dla funkcji trzech zmiennych nie jesteśmy w stanie narysować jej wykresu, bo do tego potrzebny jest lineał czterowymiarowy, natomiast możemy wskazywać jej poziomice. Na przykład dla funkcji $f(x,y,z) = x^2+y^2+z^2$ poziomicami wyznaczonymi przez równania $x^2+y^2+z^2 = c$ są sfery o promieniach $r = \sqrt{c}$. Te sfery są coraz gęstsze w miarę oddalania się od początku układu. W elektrostatyce poziomicami funkcji potencjału elektrycznego są *powierzchnie ekwipotencjalne.*

Mając pojęcie odległości w $\mathbb{R}^n$ możemy w analogii do funkcyj jednej zmiennej zdefiniować granicę funkcji w punkcie i ciągłość funkcji. *Definicja Cauchy'ego granicy funkcji*: funkcja $f : A \to \mathbb{R}$ dla $A \subset \mathbb{R}^n$ *ma granicę* b *w punkcie* $\mathbf{a}$, jeśli

$$\forall \epsilon > 0 \ \ \exists \delta > 0 : \ 0 < ||\,\mathbf{x}-\mathbf{a}\,|| < \delta \Rightarrow |f(\mathbf{x}) - b\,| < \epsilon. \tag{7.16}$$

Piszemy wtedy

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = b.$$

Widzimy, że jest to dostosowanie definicji Cauchy'ego granicy funkcji do przypadku wielu wymiarów.

Twierdzenie 1. Funkcja f ma granicę b w punkcie $\mathbf{a}$ w.t.w. gdy

$$\text{dla każdego ciągu } \ \mathbf{x}_k \in A,\ \mathbf{x}_k \to \mathbf{a},\ \mathbf{x}_k \neq \mathbf{a} \ \text{ zachodzi } \ f(\mathbf{x}_k) \to b. \tag{7.17}$$

Zdanie (7.17) nazywa się *definicją Heinego granicy funkcji.* Dowód jest przeniesieniem dowodu o równoważności definicyj Heinego i Cauchy'ego dla funkcji jednej zmiennej.

Dowód $\Longleftarrow$ nie wprost. Przypuśćmy, że warunek (7.16) nie jest spełniony, tzn.

$$\exists \epsilon > 0 \ \ \forall \delta > 0 : \ 0 < ||\,\mathbf{x}-\mathbf{a}\,|| < \delta \ \wedge \ |f(\mathbf{x}) - b\,| \geq \epsilon.$$

W szczególności po podstawieniu $\delta = \frac{1}{k}$ zauważymy, że istnieje ciąg $\{\mathbf{x}_k\}$ taki, że

$$0 < ||\,\mathbf{x}_k - \mathbf{a}\,|| < \frac{1}{k} \tag{7.18}$$

oraz

$$|f(\mathbf{x}_k) - b\,| \geq \epsilon. \tag{7.19}$$

Nierówności (7.18) mówią, że $\mathbf{x}_k \to \mathbf{a}$ oraz $\mathbf{x}_k \neq \mathbf{a}$. Nierówność (7.19) natomiast mówi, że nie zachodzi $f(\mathbf{x}_k) \to b$. W takim razie nie może zachodzić warunek (7.17). □

$\Longrightarrow$. Weźmy dowolne $\epsilon > 0$ i dowolny ciąg $\{\mathbf{x}_k\}$ o własności $\lim\limits_{k\to\infty} \mathbf{x}_k = \mathbf{a}$ oraz $\mathbf{x}_k \neq \mathbf{a}$. Na mocy warunku (7.16) istnieje $\delta > 0$ taka, że $0 < ||\,\mathbf{x}-\mathbf{a}\,|| < \delta \Rightarrow |f(\mathbf{x}) - b\,| < \epsilon$. Ze względu na założenie $\lim\limits_{k\to\infty} \mathbf{x}_k = \mathbf{a}$ nierówność $||\,\mathbf{x}_k - \mathbf{a}\,|| < \delta$ jest spełniona dla wszystkich k większych od pewnego $N \in \mathbb{N}$. Dla tych samych $k > N$ mamy więc nierówność $|f(\mathbf{x}_k) - b\,| < \epsilon$, co oznacza $\lim\limits_{k\to\infty} f(\mathbf{x}_k) = b$. W ten sposób pokazaliśmy prawdziwość zdania (7.17). ■

PRZYKŁAD 4a. Wiemy, że funkcja dana wzorem $f(x,y) = \frac{x^2}{x^2+y^2}$ ma dziedzinę $\mathbb{R}^2 \setminus (0,0)$. Jej wykres jest ukazany na Rys. 7.11. Badamy, czy ma ona granicę w zerze, czyli w punkcie $\mathbf{0}$=(0,0). Jeśli wybierzemy ciąg wektorów $\mathbf{r}_k = (0, y_k)$, czyli podążymy do zera po osi Y, to ciąg wartości funkcji $\{f(0,y_k)\} = \{\frac{0}{y^2}\}$ przy $y_k \to 0$ ma granicę zero. Jeśli podążymy do zera po prostej $y = \alpha x$, czyli wybierzemy ciąg punktów $\mathbf{r}_k = (x_k, \alpha x_k)$, to otrzymamy ciąg wartości funkcji $\{f(x_k, \alpha x_k)\} = \{{x_k}^2/({x_k}^2 + \alpha^2 {x_k}^2)\} = \{\frac{1}{1+\alpha^2}\}$. Granicą tego ciągu $\{f(\mathbf{r}_k)\}$ jest $\frac{1}{1+\alpha^2}$. W ten sposób dla różnych $\alpha \in \mathbb{R}$ otrzymujemy różne liczby z przedziału $(0, 1\rangle$. Widzimy, ze dla różnych ciągów dążących do zera otrzymujemy różne granice wartości funkcji, warunek (7.17) nie jest spełniony, więc funkcja nie ma granicy w zerze. Rys. 7.11 to potwierdza, bo widać na nim, że punktowi (0,0) nie można przypisać jednej wartości, bo jest nad nim odcinek (0,1) osi Z.

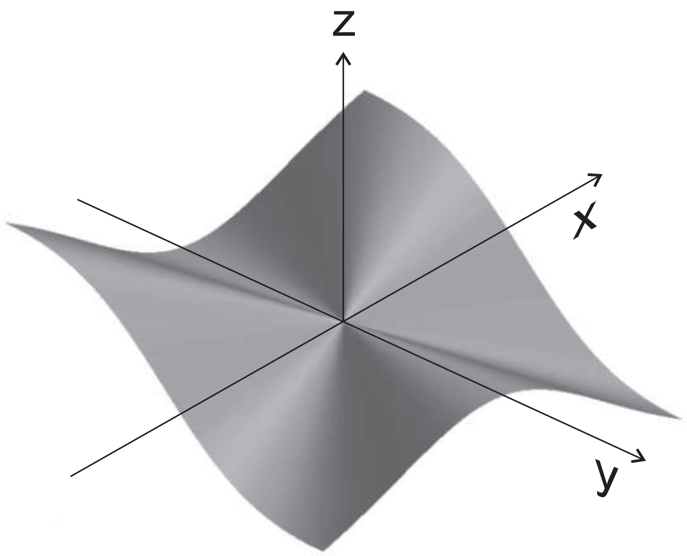
Niech $A \subset \mathbb{R}^n$. Mówimy, że funkcja $f : A \to \mathbb{R}$ jest *ciągła w punkcie* $\mathbf{a} \in A$, jeśli

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a}),$$

czyli, gdy granica funkcji w punkcie jest równa wartości funkcji w tym punkcie. Funkcja z przykładu 1 nie jest ciągła, bo nie ma granicy w zerze, więc granica w zerze nie może być porównywana z wartością w zerze.

PRZYKŁAD 5. Badamy, czy funkcja

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



Rysunek 7.13

ma granicę w zerze. Jej wykres jest ukazany na Rys. 7.13. Mamy nierówność $\frac{x^2}{x^2+y^2} \le 1$, mnożymy ją stronami przez $|y|$:

$$\frac{x^2}{x^2 + y^2}|y| \le |y|,$$

skąd

$$\left|\frac{x^2 y}{x^2 + y^2}\right| \le |y|. \tag{7.20}$$

Jeśli $\| \mathbf{r} - \mathbf{0} \| = \sqrt{x^2 + y^2} < \delta$, to $|y| \leq \sqrt{x^2 + y^2} < \delta$. Dla dowolnego $\epsilon > 0$ wystarczy dobrać $\delta = \epsilon$, a wtedy dzięki (7.20) mamy

$$|f(\mathbf{r}) - 0| = \left| \frac{x^2 y}{x^2 + y^2} \right| \leq |y| < \epsilon.$$

Wykazaliśmy spełnienie zdania (7.16), więc $\lim_{\mathbf{r}\to\mathbf{0}} f(\mathbf{r}) = 0$. Skoro granica funkcji pokrywa się z wartością funkcji w zerze, to funkcja jest ciągła w zerze.

Niech $U \subset \mathbb{R}^n$ będzie zbiorem otwartym i niech $f : U \to \mathbb{R}$ będzie funkcją n zmiennych. Oznaczamy

$$\frac{\partial f(\mathbf{x})}{\partial x_j} = \lim_{h\to 0} \frac{f(\mathbf{x} + h\mathbf{e}_j) - f(\mathbf{x})}{h} \tag{7.21}$$

i jeśli ta granica istnieje, to nazywamy ją *j-tą pochodną cząstkową funkcji f* albo *pochodną cząstkową funkcji f względem j-tej zmiennej.* Wyrażenie w liczniku po prawej stronie można zapisać inaczej

$$f(\mathbf{x} + h\mathbf{e}_j) - f(\mathbf{x}) = f(x_1, ..., x_{j-1}, x_j + h, x_{j+1}, ..., x_n) - f(x_1, ..., x_{j-1}, x_j, x_{j+1}, ..., x_n),$$

więc widzimy, że tę pochodną definiuje się jak pochodną funkcji jednej zmiennej x_j przy ustalonych pozostałych zmiennych.

PRZYKŁAD 6. Dla funkcji dwóch zmiennych $f(x, y) = \cos xy + x \cos y$ można rozważać dwie pochodne cząstkowe:

$$\frac{\partial f}{\partial x} = -y \sin xy + \cos y, \qquad \frac{\partial f}{\partial y} = -x \sin xy - x \sin y.$$

Pochodne te wyliczyliśmy korzystając z reguł różniczkowania funkcyj jednej zmiennej.

PRZYKŁAD 7. Przy funkcji dwóch zmiennych $f(x, y) = x^{1/3} y^{1/3}$ dwie pochodne cząstkowe w zerze dla ostrożności liczymy wprost z definicji:

$$\frac{\partial f(0,0)}{\partial x} = \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} = \lim_{h\to 0} \frac{h^{1/3} \cdot 0 - 0}{h} = 0,$$
$$\frac{\partial f(0,0)}{\partial y} = \lim_{h\to 0} \frac{f(0,h) - f(0,0)}{h} = \lim_{h\to 0} \frac{0 \cdot h^{1/3} - 0}{h} = 0.$$

Na osi X dwie pochodne cząstkowe liczymy też z definicji:

$$\frac{\partial f(x_0,0)}{\partial x} = \lim_{h\to 0} \frac{f(x_0 + h, 0) - f(x_0, 0)}{h} = \lim_{h\to 0} \frac{(x_0 + h)^{1/3} \cdot 0 + 0}{h} = 0,$$
$$\frac{\partial f(x_0,0)}{\partial y} = \lim_{h\to 0} \frac{f(x_0, h) - f(x_0, 0)}{h} = \lim_{h\to 0} \frac{x_0^{1/3} \cdot h^{1/3}}{h} = \lim_{h\to 0} x_0^{1/3} \cdot h^{-2/3},$$

a ta granica nie istnieje. Widzimy, że na osi X istnieje tylko pochodna cząstkowa względem x. Podobnie przekonamy się, że na osi Y istnieje tylko pochodna cząstkowa względem y. Poza osiami współrzędnych te pochodne istnieją i nawet są funkcjami ciągłymi, o czym wiemy z reguł różniczkowania funkcyj jednej zmiennej:

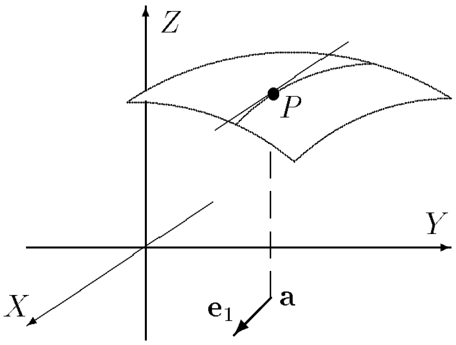
$$\frac{\partial f(x,y)}{\partial x} = \frac{1}{3} x^{-2/3} y^{1/3}, \qquad \frac{\partial f(x,y)}{\partial y} = \frac{1}{3} x^{1/3} y^{-2/3}.$$

Inne oznaczenia na j-tą pochodną cząstkową funkcji $f : \mathbb{R}^n \to \mathbb{R}$:

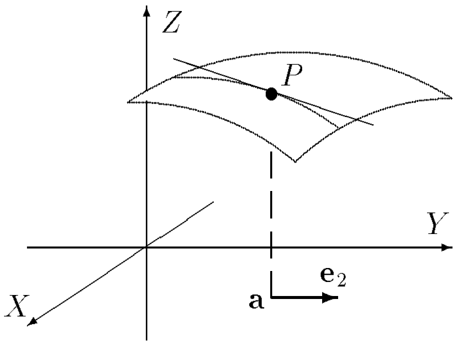
$$\partial_j f(\mathbf{x}), \quad \partial_{x_j} f(\mathbf{x}), \quad f'_{x_j}(\mathbf{x}).$$

Litera ∂ czytana jest „de kręcone". To oznaczenie wymyślono specjalnie po to, aby odróżnić ten rodzaj pochodnej od pochodnej funkcji jednej zmiennej, którą nazywa się *pochodną zwyczajną*.

W przypadku funkcji dwóch zmiennych dla interpretacji pochodnej cząstkowej względem x trzeba rozważać płaszczyznę równoległą do płaszczyzny (X, Z), przechodzącą przez punkt $P = \{\mathbf{a}, f(\mathbf{a})\}$. Przekrój tej płaszczyzny z wykresem funkcji wyznacza pewną krzywą, zob. Rys. 7.14. Pochodna cząstkowa $\partial_x f(\mathbf{a})$ jest tangensem kąta nachylenia stycznej do tej krzywej w punkcie P. Dla interpretacji pochodnej względem y trzeba rozważać płaszczyznę równoległą do płaszczyzny (Y, Z), przechodzącą przez punkt $P = \{\mathbf{a}, f(\mathbf{a})\}$. Przekrój tej płaszczyzny z wykresem funkcji wyznacza inną krzywą, zob. Rys. 7.15. Pochodna cząstkowa $\partial_x y(\mathbf{a})$ jest tangensem kąta nachylenia stycznej do tej krzywej w punkcie P.



Rysunek 7.14



Rysunek 7.15

Istnienie pochodnych cząstkowych w danym punkcie nie wystarcza do nazwania funkcji różniczkowalną w tym punkcie. Definicja tego pojęcia jest bardziej wymagająca. Niech $U \subset \mathbb{R}^n$ oraz $f : U \to \mathbb{R}$. Mówimy, że funkcja f jest*różniczkowalna w punkcie* $\mathbf{a} \in U$, jeśli dla wszystkich $j \in \{1, 2, \dots, n\}$ istnieją pochodne cząstkowe $\partial_j f(\mathbf{a})$ oraz jeśli

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{f(\mathbf{x}) - f(\mathbf{a}) - \sum_{j=1}^{n} \partial_j f(\mathbf{a}) \cdot (x_j - a_j)}{|| \, \mathbf{x} - \mathbf{a} \, ||} = 0. \tag{7.22}$$

Ta definicja jest wzięta z faktu, iż definicję pochodnej funkcji jednej zmiennej

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

można zapisać w sposób równoważny

$$\lim_{x \to a} \frac{f(x) - f(a) - f'(a)(x - a)}{x - a} = 0. \tag{7.23}$$

Jeśli funkcja jednej zmiennej f jest różniczkowalna w punkcie a, to różnicę $f(x) - f(a)$ dobrze przybliża iloczyn $f'(a)(x - a)$. Przybliża tak dobrze, że różnica między nimi dąży do zera szybciej, niż różnica argumentów $x - a$.

PRZYKŁAD 7a. Sprawdźmy, czy funkcja z Przykładu 4 jest różniczkowalna w zerze. Tym razem $\mathbf{x} = (x, y)$. Według definicji (7.22) mamy liczyć granicę

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{f(x,y) - f(0,0) - \left[\frac{\partial f}{\partial x}(0,0)\cdot(x-0) + \frac{\partial f}{\partial y}(0,0)\cdot(y-0)\right]}{\sqrt{(x-0)^2+(y-0)^2}}$$

$$= \lim_{(x,y)\to(0,0)} \frac{x^{1/3}y^{1/3} - 0 - 0\cdot x - 0\cdot y}{\sqrt{x^2+y^2}}.$$

Gdy dążymy do zera po osi X, otrzymujemy granicę

$$\lim_{x\to 0}\frac{x^{1/3}\cdot 0 - 0 - 0\cdot x}{|x|} = 0.$$

Jeśli zaś podążamy po prostej $y = x$, to mamy granicę

$$\lim_{x\to 0}\frac{x^{1/3}x^{1/3} - 0 - 0\cdot x - 0\cdot x}{\sqrt{x^2+x^2}} = \lim_{x\to 0}\frac{x^{2/3}}{\sqrt{2}|x|} = 2^{-1/2}\lim_{x\to 0}|x|^{-1/3},$$

a ta granica nie istnieje. W takim razie rozważana funkcja nie jest różniczkowalna w zerze. Istnienie dwóch pochodnych cząstkowych nie jest więc wystarczające dla różniczkowalności funkcji w danym punkcie.

Twierdzenie 2. Niech $U \subset \mathbb{R}^n$ oraz $f : U \to \mathbb{R}$. Jeśli pochodne cząstkowe $\partial_j f$ istnieją dla wszystkich $j \in \{1, \dots, n\}$ i są ciągłe w punkcie $\mathbf{a}$, to funkcja f jest różniczkowalna w $\mathbf{a}$.

Dowodu nie będzie.

Funkcja z Przykładu 4 nie ma ciągłych pochodnych cząstkowych w zerze, więc nic dziwnego, że nie jest tam różniczkowalna.

PRZYKŁAD 8. Niech $f(x,y) = xy^2e^{xy}$. Ta funkcja ma za dziedzinę całą płaszczyznę (X, Y). Pochodne cząstkowe liczymy według reguł różniczkowania funkcyj jednej zmiennej:

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(xy^2)e^{xy} + xy^2\frac{\partial}{\partial x}(e^{xy}) = y^2e^{xy} + xy^2(ye^{xy}) = y^2(1+xy)e^{xy},$$

$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}(xy^2)e^{xy} + xy^2\frac{\partial}{\partial y}(e^{xy}) = 2xye^{xy} + xy^2(xe^{xy}) = xy(2+xy)e^{xy}.$$

Pochodne te istnieją na całej dziedzinie i są ponadto funkcjami ciągłymi.

Teraz czas na definicję następnego pojęcia. Jeśli funkcja f jest różniczkowalna w punkcie $\mathbf{x}_0$, to układ pochodnych cząstkowych

$$\{\partial_1 f(\mathbf{x}_0), \dots, \partial_n f(\mathbf{x}_0)\} = \nabla f(\mathbf{x}_0)$$

stanowi wektor w $\mathbb{R}^n$, który nazywamy *gradientem funkcji* f i oznaczamy $\nabla f = \operatorname{grad} f$. Trójkąt odwrócony wierzchołkiem na dół czytamy jako *nabla*. Ta nazwa wzięła się z nazwy strunowego instrumentu hinduskiego. Gradient zapisuje się również za pomocą wektorów bazowych:

$$\nabla f = \text{grad } f = \sum_{j=1}^{n} \mathbf{e}_j \frac{\partial f}{\partial x_j}.$$

W przestrzeni trójwymiarowej ten wzór znany jest w postaci

$$\text{grad } f = \mathbf{i}\frac{\partial f}{\partial x} + \mathbf{j}\frac{\partial f}{\partial y} + \mathbf{k}\frac{\partial f}{\partial z}$$

z trzema prostopadłymi wektorami jednostkowymi **i**, **j**, **k**.

We wzorze (7.22) rozpoznajemy współrzędne gradientu, więc możemy ten wzór zapisać za pomocą iloczynu gradientu z wektorem $\mathbf{r} - \mathbf{a}$:

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{x}) - f(\mathbf{a}) - [\nabla f(\mathbf{a})] \cdot (\mathbf{x} - \mathbf{a})}{|| \, \mathbf{x} - \mathbf{a} \, ||} = 0. \tag{7.24}$$

Teraz lepiej widać podobieństwo do wzoru (7.23).

Niech $\langle a, b\rangle$ będzie przedziałem domkniętym w $\mathbb{R}$ i niech dana będzie funkcja $\boldsymbol{\sigma}$: $\langle a, b\rangle \to \mathbb{R}^n$. Zwykle będziemy zakładać, że funkcja $\boldsymbol{\sigma}(t) = \{\sigma_1(t), \sigma_2(t), \ldots \sigma_n(t)\}$ ma wszystkie współrzędne różniczkowalne i ich pochodne są ciągłe. Zbiór

$$K = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \boldsymbol{\sigma}(t), \, t \in \langle a, b\rangle\}$$

nazywamy *krzywą* w $\mathbb{R}^n$, a funkcję $\boldsymbol{\sigma}$ nazywamy *parametryzacją* krzywej K. Można powiedzieć, że krzywa K to obraz funkcji $\boldsymbol{\sigma}$. W fizyce często posługujemy się takimi funkcjami $\mathbf{r} = \boldsymbol{\sigma}(t)$ opisującymi ruch jakiejś cząstki w naszej przestrzeni $\mathbb{R}^3$. W takim razie w terminologii fizycznej parametryzacja $\boldsymbol{\sigma}(t)$ to ruch, a krzywa K to tor tego ruchu. Jeśli $U \subset \mathbb{R}^n$ i mamy funkcję $f : U \to \mathbb{R}$, a funkcja $\boldsymbol{\sigma}$ ma obraz zawarty w U, to można rozważać funkcję złożoną $f \circ \boldsymbol{\sigma} : \langle a, b\rangle \to \mathbb{R}$:

$$(f \circ \boldsymbol{\sigma})(t) = f(\boldsymbol{\sigma}(t)) = f(\sigma_1(t), \sigma_2(t), \ldots, \sigma_n(t)).$$

Funcja złożona jest funkcją jednej zmiennej. Przy liczeniu pochodnej zwyczajnej funkcji złożonej trzeba uwzględnić obecność zmiennej t w każdym z n argumentów funkcji f, więc wystąpi n iloczynów pochodnych funkcji zewnętrznej i wewnętrznej:

$$\frac{d(f \circ \boldsymbol{\sigma})}{dt} = \frac{\partial f}{\partial x_1}\frac{d\sigma_1}{dt} + \frac{\partial f}{\partial x_2}\frac{d\sigma_2}{dt} + \ldots + \frac{\partial f}{\partial x_n}\frac{d\sigma_n}{dt} = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}\frac{d\sigma_j}{dt}. \tag{7.25}$$

Ten wzór można zapisać krócej za pomocą gradientu i skaloczynu:

$$\frac{d(f \circ \boldsymbol{\sigma})}{dt} = \nabla f \cdot \frac{d\boldsymbol{\sigma}}{dt}. \tag{7.26}$$

Dwa ostatnie wzory są bardzo ważne w rachunkach przybliżonych, które matematycy uważają za nieścisłe. Jeśli równość (7.25) pomnożymy stronami przez dt, to otrzymamy różniczkę funkcji złożonej

$$d(f \circ \boldsymbol{\sigma}) = \frac{\partial f}{\partial x_1}d\sigma_1 + \frac{\partial f}{\partial x_2}d\sigma_2 + \ldots + \frac{\partial f}{\partial x_n}d\sigma_n.$$

Różniczka $d\sigma_j$ oznacza dowolnie mały przyrost współrzędnej x_j wektora wodzącego, więc można to zapisać inaczej:

$$df = \frac{\partial f}{\partial x_1}\, dx_1 + \frac{\partial f}{\partial x_2}\, dx_2 + \ldots + \frac{\partial f}{\partial x_n}\, dx_n.$$

Fizycy nazywają to wyrażenie *różniczką zupełną* funkcji wielu zmiennych i najczęściej wykorzystują w rachunku błędów w postaci

$$\Delta f = \frac{\partial f}{\partial x_1}\, \Delta x_1 + \frac{\partial f}{\partial x_2}\, \Delta x_2 + \ldots + \frac{\partial f}{\partial x_n}\, \Delta x_n.$$

Niech $U \subset \mathbb{R}^n$, $f : U \to \mathbb{R}$. *Pochodną dążnościową* funkcji f w punkcie $\mathbf{a}$ *z dążnością* $\mathbf{v}$ nazywamy

$$\lim_{h\to 0} \frac{f(\mathbf{a} + h\mathbf{v}) - f(\mathbf{a})}{h}, \tag{7.27}$$

jeśli ta granica istnieje. Pochodną dążnościową w punkcie $\mathbf{a}$ oznaczamy $D_{\mathbf{v}} f(\mathbf{a})$. We wzorze (7.27) można zobaczyć złożenie funkcji f z parametryzacją pewnej szczególnej krzywej $\boldsymbol{\sigma}(t) = \mathbf{a} + t\mathbf{v}$. Ta krzywa jest linią prostą, a jeśli uznamy, że parametr t jest czasem, to ta parametryzacja odpowiada ruchowi jednostajnemu z prędkością $\mathbf{v}$. Wyrażenie $f \circ \boldsymbol{\sigma}$ to zamiana funkcji n zmiennych na funkcję jednej zmiennej. Pochodna dążnościowa to pochodna tej nowej funkcji w zerze:

$$D_{\mathbf{v}} f(\mathbf{a}) = \frac{d}{dt} f(\mathbf{a} + t\mathbf{v})|_{t=0}. \tag{7.28}$$

Można powiedzieć, że jest to szybkość zmiany wielkości f, jeśli cząstka przechodzi przez punkt $\mathbf{a}$ z prędkością $\mathbf{v}$. W tym zapisie widać zależność od wektora $\mathbf{v}$. Im większa będzie prędkość $\mathbf{v}$, tym szybciej będzie cząstka przechodzić przez punkt $\mathbf{a}$, więc tym większa będzie szybkość zmiany wielkości f, czyli większa będzie pochodna dążnościowa.

Lemat 1. Jeśli funkcja $f : U \to \mathbb{R}$ jest różniczkowalna w punkcie $\mathbf{a}$, to

$$D_{\mathbf{v}} f(\mathbf{a}) = \operatorname{grad} f(\mathbf{a}) \cdot \mathbf{v}.$$

Dowód. Pochodną dążnościową wyrażamy wzorem (7.28). Widzimy, że pod argumentem funkcji f stoi parametryzacja prostej $\boldsymbol{\sigma}(t) = \mathbf{a} + t\mathbf{v}$. Jej pochodna wynosi $d\boldsymbol{\sigma}/dt = \mathbf{v}$. Wystarczy to podstawić do wzoru (7.26)

$$D_{\mathbf{v}} f(\mathbf{a}) = \left.\frac{d(f \circ \boldsymbol{\sigma})}{dt}\right|_{t=0} = \operatorname{grad} f(\mathbf{a}) \cdot \left.\frac{d\boldsymbol{\sigma}}{dt}\right|_{t=0} = \operatorname{grad} f(\mathbf{a}) \cdot \mathbf{v}. \;\blacksquare$$

Ten wynik można zapisać za pomocą nabli: $D_{\mathbf{v}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{v}$. Iloczyn skalarny jest przemienny, więc można to zapisać jeszcze inaczej:

$$D_{\mathbf{v}} f(\mathbf{a}) = \mathbf{v} \cdot \nabla f(\mathbf{a}).$$

Z tego wzoru widać, że pochodna dążnościowa jest liniowa względem $\mathbf{v}$, tzn. *jednorodna*:

$$D_{\alpha\mathbf{v}} f = (\alpha\mathbf{v}) \cdot \nabla f = \alpha\, (\mathbf{v} \cdot \nabla f) = \alpha\, D_{\mathbf{v}} f$$

(tu potwierdza się to, co wspomnieliśmy wcześniej: im większa prędkość $\mathbf{v}$, tym większa pochodna dążnościowa) oraz *addytywna*:

$$D_{\mathbf{v}_+\mathbf{v}_2} f = (\mathbf{v}_1 + \mathbf{v}_2) \cdot \nabla f = \mathbf{v}_2 \cdot \nabla f + \mathbf{v}_1 \cdot \nabla f = D_{\mathbf{v}_1} f(\mathbf{a}) + D_{\mathbf{v}_2} f(\mathbf{a}).$$

Z wzoru na wartość iloczynu skalarnego mamy

$$\mathbf{v} \cdot \nabla f = ||\, \mathbf{v} \,||\; ||\, \nabla f \,|| \cos \alpha,$$

Przy ustalonych długościach czynników ten skaloczyn jest największy dla $\alpha = 0$. Jeśli wybierzemy wektor $\mathbf{v}$ o długości takiej samej jak gradient funkcji, to otrzymamy

$$\mathbf{v} \cdot \nabla f = ||\, \nabla f \,||\; ||\, \nabla f \,|| \cos \alpha,$$

Podstawmy za wektor $\mathbf{v}$ dokładnie gradient funkcji, wtedy $\alpha = 0$ oraz $\cos \alpha = 1$, skąd

$$D_{\nabla f} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \nabla f(\mathbf{a}) = ||\, \nabla f(\mathbf{a}) \,||^2 .$$

To oznacza, że dla różnych wektorów $\mathbf{v}$ o długości takiej samej jak $\operatorname{grad} f$ pochodna dążnościowa jest największa dla $\mathbf{v} = \nabla f$. Stąd wyciągamy wniosek, że gradient jest wektorem, który pokazuje kierunek najszybszego wzrostu funkcji.

PRZYKŁAD 9. Niech $f(x, y) = x^2 - 4xy$. Znajdziemy gradient tej funkcji w punkcie $\mathbf{r}_1 = (3, -2)$:

$$\nabla f(x, y) = (2x - 4y, -4x).$$

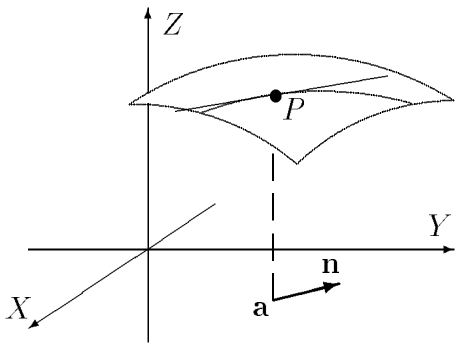
W takim razie we wskazanym punkcie $\nabla f(3, -2) = (6 + 8, -12) = (14, -12)$. Teraz wyliczymy pochodną dążnościową z wektorem łączącym punkty $\mathbf{r}_1$ i $\mathbf{r}_2 = (4, 1)$. Otóż $\mathbf{v} = \mathbf{r}_2 - \mathbf{r}_1 = (4 - 3, 1 - (-2)) = (1, 3)$, więc

$$\mathbf{v} \cdot \nabla f(3, -2) = (1, 3) \cdot (14, -12) = 14 - 36 = -22.$$

Jeśli wybierzemy $\mathbf{v} = \mathbf{n}$, gdzie $||\, \mathbf{n} \,|| = 1$, to pochodna dążnościowa $D_{\mathbf{n}} f$ nazywa się *pochodną kierunkową*. Pochodna cząstkowa jest jej szczególnym przypadkiem:

$$D_{\mathbf{e}_j} f = \mathbf{e}_j \cdot \nabla f = \frac{\partial f}{\partial x_j}.$$

Pochodna kierunkowa opisuje szybkość zmiany funkcji f w kierunku $\mathbf{n}$. Niech f będzie funkcją dwóch zmiennych . Weźmy płaszczyznę równoległą do wektorów $\mathbf{e}_3$ i $\mathbf{n}$ oraz przechodzącą przez punkt $P = \{\mathbf{a}, f(\mathbf{a})\}$. Przekrój tej płaszczyzny z wykresem funkcji wyznacza pewną krzywą, zob. Rys. 7.16. Pochodna kierunkowa $\mathbf{n} \cdot \nabla f(\mathbf{a})$ jest tangensem kąta nachylenia stycznej do tej krzywej w punkcie P.



Rysunek 7.16

PRZYKŁAD 9a. Dla tej samej funkcji, co w Przykładzie 6 znajdziemy pochodną kierunkową w kierunku wektora $\mathbf{v}$. Wprowadzamy wektor jednostkowy o potrzebnym kierunku:

$$\mathbf{n} = \frac{\mathbf{v}}{||\,\mathbf{v}\,||} = \frac{1}{\sqrt{1^2+3^2}}\,(1,3) = \frac{1}{\sqrt{10}}\,(1,3)$$

i obliczamy pochodną kierunkową

$$\mathbf{n}\cdot\nabla f(3,-2) = \frac{1}{\sqrt{10}}\mathbf{v}\cdot\nabla f(3,-2) = \frac{1}{\sqrt{10}}(-22) = -\frac{22}{\sqrt{10}} \approx -6,96.$$

Twierdzenie 3. Niech $U \subset \mathbb{R}^n$, $f : U \to \mathbb{R}$, niech S będzie poziomicą funkcji f i niech $\mathbf{x}_0 \in S$. Jeśli $\boldsymbol{\sigma}(t)$ jest krzywą leżącą na S taką, że $\boldsymbol{\sigma}(0) = \mathbf{x}_0$, to

$$\mathbf{v}\cdot\nabla f(\mathbf{x}_0) = 0 \tag{7.29}$$

dla $\mathbf{v} = \boldsymbol{\sigma}'(0)$.

Dowód. Jeśli krzywa $\boldsymbol{\sigma}$ leży całkowicie na poziomicy S, to funkcja jednej zmiennej $(f\circ\boldsymbol{\sigma})(t)$ jest stała, więc jej pochodna jest zerem. Korzystamy ze wzoru (7.26):

$$0 = \frac{d}{dt}f(\boldsymbol{\sigma}(t))|_{t=0} = \nabla f\cdot\boldsymbol{\sigma}'(0) = \nabla f\cdot\mathbf{v}. \ \blacksquare$$

To twierdzenie jest bardzo ważne, bo mówi, że gradient jako wektor jest prostopadły do poziomic. Skoro krzywa $\boldsymbol{\sigma}$ leży na poziomicy, to wektor $\mathbf{v}$ jako styczny do krzywej jest też styczny do poziomicy. Skoro poza tym krzywa jest dowolna, to możemy uważać, że wektor $\mathbf{v}$ może być dowolnym wektorem stycznym do poziomicy. Równość (7.29) oznacza prostopadłość wektorów ∇f i $\mathbf{v}$. Gradient jest więc prostopadły do każdego wektora stycznego do poziomicy, a tak właśnie rozumie się prostopadłość do poziomicy.

PRZYKŁAD 10. Funkcja $f(x,y,z) = (x-a)^2+(y-b)^2+(z-c)^2$ jako swoje poziomice ma sfery o wspólnym środku w punkcie $\mathbf{t} = (a,b,c)$. Wyliczmy gradient tej funkcji

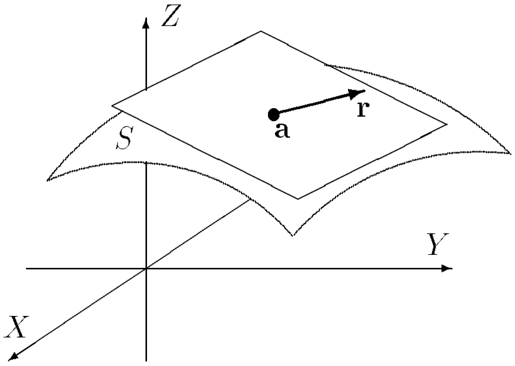
$$\nabla f(\mathbf{r}) = (2(x-a),\,2(y-b),\,2(z-c)) = 2\,[(x,y,z)-(a,b,c)] = 2(\mathbf{r}-\mathbf{t}).$$

Stwierdzamy, że gradient jest proporcjonalny do wektora $\mathbf{r}-\mathbf{t}$. A ten wektor (zaczepiony w punkcie $\mathbf{r}$) jest prostopadły do sfery o środku w punkcie $\mathbf{t}$ przechodzącej przez punkt $\mathbf{r}$.

Niech $U \subset \mathbb{R}^3$ $f : U \to \mathbb{R}$. Wiemy, że dla takiej funkcji poziomicami są na ogół powierzchnie. Zbiór końców wszystkich wektorów stycznych do powierzchni w ustalonym punkcie stanowi płaszczyznę. Jeśli S jest poziomicą funkcji f oraz $\mathbf{a}\in S$, to na mocy tego, co już zauważyliśmy, wektor $\mathbf{r}-\mathbf{a}$ jest styczny do S w punkcie $\mathbf{a}$ w.t.w. gdy jest prostopadły do ∇f:

$$\nabla f(\mathbf{a})\cdot(\mathbf{r}-\mathbf{a}) = 0. \tag{7.30}$$

Wektor $\mathbf{r}$ ma swój koniec na płaszczyźnie stycznej do S, zob. Rys. 7.17. W takim razie mamy prawo nazywać równość (7.30) *równaniem płaszczyzny stycznej do poziomicy* funkcji f w punkcie $\mathbf{a}$.



Rysunek 7.17

Niech $U \subset \mathbb{R}^n$ i niech funkcja $f : U \to \mathbb{R}$ będzie różniczkowalna w każdym punkcie $\mathbf{x} \in U$. Z definicji oznacza to, że istnieją pochodne cząstkowe $\partial_j f$ i one również są funkcjami z U w $\mathbb{R}$. Jeśli wszystkie pochodne cząstkowe są funkcjami ciągłymi, to mówimy, że funkcja f *jest klasy* C^1 i piszemy $f \in C^1$. Możemy rozważać z kolei ich pochodne:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \partial_i \partial_j f = f''_{x_i x_j},$$

które są *drugimi pochodnymi cząstkowymi funkcji f* albo *pochodnymi cząstkowymi drugiego rzędu*. Uwaga: najpierw liczona jest pochodna stojąca na prawo, czyli w tym wzorze względem x_j. Tych drugich pochodnych jest n^2. Jeśli wszystkie drugie pochodne są ciągłe, to mówimy, ze funkcja f *jest klasy* C^2.

Twierdzenie 4. Jeśli $U \subset \mathbb{R}^n$, a $f : U \to \mathbb{R}$ jest klasy C^2, to drugie pochodne mieszane są przemienne, tzn.

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Dowód. Wprowadźmy oznaczenie

$$S(\Delta x_i, \Delta x_j) =$$

$$= f(x_1, \ldots, x_{i0} + \Delta x_i, \ldots, x_{j0} + \Delta x_j, \ldots, x_n) - f(x_1, \ldots, x_{i0}, \ldots, x_{j0} + \Delta x_j, \ldots, x_n)$$
$$- f(x_1, \ldots, x_{i0} + \Delta x_i, \ldots, x_{j0}, \ldots, x_n) + f(x_1, \ldots, x_{i0}, \ldots, x_{j0}, \ldots, x_n).$$

Jeśli x_{j0} i Δx_j są ustalone, to definiujemy fukcję jednej zmiennej

$$g(x_i) = f(x_1, \ldots, x_i, \ldots, x_{j0} + \Delta x_j, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_{j0}, \ldots, x_n). \quad (7.31)$$

Wtedy możemy napisać

$$S(\Delta x_i, \Delta x_j) = g(x_{i0} + \Delta x_i) - g(x_{i0}).$$

Z twierdzenia o wartości pośredniej (Tw. 3 w § 4.2) mamy

$$g(x_{i0} + \Delta x_i) - g(x_{i0}) = g'(\tilde{x}_i)\Delta x_i, \qquad \text{dla} \quad \tilde{x}_i \in (x_{i0}, x_{i0} + \Delta x_i).$$

Wykorzystujemy wzór (7.31):

$$S(\Delta x_i, \Delta x_j) = \left[\frac{\partial f}{\partial x_i}(\ldots, \tilde{x}_i, \ldots, x_{j0} + \Delta x_j \ldots) - \frac{\partial f}{\partial x_i}(\ldots \tilde{x}_i \ldots, x_{j0} \ldots)\right] \Delta x_i.$$

Stosujemy drugi raz twierdzenie o wartości pośredniej:

$$S(\Delta x_i, \Delta x_j) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\ldots, \tilde{x}_i, \ldots, \tilde{x}_j \ldots)\Delta x_i \Delta x_j.$$

Stąd

$$\frac{S(\Delta x_i, \Delta x_j)}{\Delta x_i \Delta x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}(\ldots, \tilde{x}_i, \ldots, \tilde{x}_j \ldots).$$

Z założenia drugie pochodne cząstkowe są ciągłe, więc

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\ldots, x_{i0}, \ldots, x_{j0} \ldots) = \lim_{(\Delta x_i, \Delta x_j) \to (0,0)} \frac{S(\Delta x_i, \Delta x_j)}{\Delta x_i \Delta x_j}.$$

Jeśli inaczej określimy funkcję g we wzorze (7.31) z jednakowym argumentem x_j, to podobnie pokażemy, że

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\ldots, x_{i0}, \ldots, x_{j0} \ldots) = \lim_{(\Delta x_i, \Delta x_j) \to (0,0)} \frac{S(\Delta x_i, \Delta x_j)}{\Delta x_j \Delta x_i}.$$

Skoro mnożenie jest przemienne i prawe strony są równe, to lewe tak samo i otrzymujemy

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\ldots, x_{i0}, \ldots, x_{j0} \ldots) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\ldots, x_{i0}, \ldots, x_{j0} \ldots). \quad \blacksquare$$

7.3 Funkcje o wartościach wektorowych

Niech $m, n \in \mathbb{N}$ i $U \subset \mathbb{R}^n$. Przez *funkcję wielu zmiennych o wartościach wektorowych* rozumiemy funkcję $\mathbf{f} : U \to \mathbb{R}^m$.Będziemy też zapisywać

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})),$$

gdzie *współrzędne* f_k są funkcjami n zmiennych o wartościach rzeczywistych (skalarnych).

Mając pojęcie odległości w $\mathbb{R}^n$ i w $\mathbb{R}^m$ możemy w analogii do funkcyj o wartościach rzeczywistych zdefiniować granicę funkcji w punkcie i ciągłość funkcji. Funkcja $\mathbf{f} : U \to \mathbb{R}^m$ dla $U \subset \mathbb{R}^n$ *ma granicę* $\mathbf{b}$ *w punkcie* $\mathbf{a}$, jeśli

$$\forall \epsilon > 0 \;\; \exists\, \delta > 0 : \;\; 0 < ||\, \mathbf{x} - \mathbf{a} \,|| < \delta \Rightarrow ||\, \mathbf{f}(\mathbf{x}) - \mathbf{b} \;\, || < \epsilon. \tag{7.32}$$

Piszemy wtedy

$$\lim_{\mathbf{x} \to \mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{b}.$$

Widzimy, że jest to dostosowanie definicji Cauchy'ego granicy funkcji do przypadku wielu wymiarów.

Twierdzenie 1. Funkcja $\mathbf{f}$ ma granicę $\mathbf{b}$ w punkcie $\mathbf{a}$ w.t.w. gdy

$$\text{dla każdego ciągu } \; \mathbf{x}_k \in U, \; \mathbf{x}_k \to \mathbf{a}, \; \mathbf{x}_k \neq \mathbf{a} \; \text{ zachodzi } \; \mathbf{f}(\mathbf{x}_k) \to \mathbf{b}. \tag{7.33}$$

Widzimy, że druga część tego twierdzenia jest odpowiednikiem definicji Heinego granicy funkcji. Dowód jest powtórzeniem dowodu Tw. 1 z poprzedniego paragrafu z zamianą modułu na normę.

Lemat 1 (*granica funkcji pomnożonej przez stałą*). Niech $U \subset \mathbb{R}^n$, niech funkcja $\mathbf{f} : U \to \mathbb{R}^m$ ma granicę w punkcie $\mathbf{a}$ i niech $c \in \mathbb{R}$. Wtedy funkcja $\mathbf{h}(\mathbf{x}) = c\mathbf{f}(\mathbf{x})$ ma granicę w $\mathbf{a}$ oraz

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{h}(\mathbf{x}) = c \lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}).$$

Dowód. Dla $c = 0$ teza lematu jest w sposób oczywisty prawdziwa, bo funkcja $\mathbf{h}$ jest wtedy stała i równa zero. Niech więc $c \neq 0$. Mamy pokazać odpowiednik zdania (7.32) dla funkcji $\mathbf{h}$ przy założeniu prawdziwości samego zdania (7.32). Potrzebna nierówność $||\, \mathbf{h} - c\mathbf{b} \,|| < \epsilon$ wynika z (7.32), gdy zamiast ϵ napisać tam $\epsilon/|c|$:

$$||\, \mathbf{f}(\mathbf{x}) - \mathbf{b} \,|| < \frac{\epsilon}{|c|} \;\Rightarrow\; |c|\, ||\, \mathbf{f}(\mathbf{x}) - \mathbf{b} \,|| < \epsilon \;\Rightarrow ||\, c\mathbf{f}(\mathbf{x}) - c\mathbf{b} \,|| < \epsilon \;\Rightarrow ||\, \mathbf{h}(\mathbf{x}) - c\mathbf{b} \,|| < \epsilon.$$

Razem z kwantyfikatorami możemy więc napisać

$$\forall \epsilon > 0 \;\; \exists\, \delta > 0 : \;\; 0 < ||\, \mathbf{x} - \mathbf{a} \,|| < \delta \Rightarrow ||\, \mathbf{h}(\mathbf{x}) - c\mathbf{b} \,|| < \epsilon,$$

co oznacza, że funkcja $\mathbf{h}$ ma granicę w $\mathbf{a}$ oraz

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{h}(\mathbf{x}) = c\mathbf{b} = c \lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}). \;\; \blacksquare$$

Lemat 2 (*granica sumy funkcyj*). Niech $U \subset \mathbb{R}^n$, niech funkcje $\mathbf{f}, \mathbf{g} : U \to \mathbb{R}^m$ mają granice w punkcie $\mathbf{a}$. Wtedy funkcja $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})$ ma granicę w $\mathbf{a}$ oraz

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{h}(\mathbf{x}) = \lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) + \lim_{\mathbf{x}\to\mathbf{a}} \mathbf{g}(\mathbf{x}).$$

Dowód. Piszemy odpowiedniki zdania (7.32) dla funkcyj $\mathbf{f}$ i $\mathbf{g}$:

$$\forall \epsilon > 0 \;\; \exists\, \delta_1 > 0 : \;\; 0 < ||\, \mathbf{x} - \mathbf{a} \,|| < \delta_1 \;\Rightarrow ||\, \mathbf{f}(\mathbf{x}) - \mathbf{b} \,|| < \frac{\epsilon}{2}, \tag{7.34}$$

$$\forall \epsilon > 0 \;\; \exists\, \delta_2 > 0 : \;\; 0 < ||\, \mathbf{x} - \mathbf{a} \,|| < \delta_2 \;\Rightarrow ||\, \mathbf{g}(\mathbf{x}) - \mathbf{d} \,|| < \frac{\epsilon}{2}. \tag{7.35}$$

Dzięki nierówności trójkąta możemy napisać

$$||\, \mathbf{h}(\mathbf{x}) - (\mathbf{b} + \mathbf{d}) \,|| = ||\, \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x}) - \mathbf{b} - \mathbf{d} \,|| = ||\, \mathbf{f}(\mathbf{x}) - \mathbf{b} + \mathbf{g}(\mathbf{x}) - \mathbf{d} \,||$$

$$\leq ||\, \mathbf{f}(\mathbf{x}) - \mathbf{b} \,|| + ||\, \mathbf{g}(\mathbf{x}) - \mathbf{d} \,|| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

jeśli prawdziwe są oba zdania (7.34) i (7.35). A one są prawdziwe dla $\delta = \min\{\delta_1, \delta_2\}$. W takim razie razem z kwantyfikatorami można napisać

$$\forall \epsilon > 0 \;\; \exists\, \delta > 0 : \;\; 0 < ||\, \mathbf{x} - \mathbf{a} \,|| < \delta \;\Rightarrow\; ||\, \mathbf{h}(\mathbf{x}) - \mathbf{b} - \mathbf{d} \,|| < \epsilon,$$

co oznacza, że funkcja $\mathbf{h}$ ma granicę w $\mathbf{a}$ oraz

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{h}(\mathbf{x}) = \mathbf{b} + \mathbf{d} = \lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) + \lim_{\mathbf{x}\to\mathbf{a}} \mathbf{g}(\mathbf{x}). \;\; \blacksquare$$

Niech $U \subset \mathbb{R}^n$. Mówimy, że funkcja $\mathbf{f} : U \to \mathbb{R}^m$ jest *ciągła w punkcie* $\mathbf{a} \in U$, jeśli

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}),$$

czyli, gdy granica funkcji w punkcie jest równa wartości funkcji w tym punkcie.

Twierdzenie 2. Funkcja $\mathbf{f} : U \to \mathbb{R}^m$ jest ciągła w punkcie $\mathbf{a} \in U$ w.t.w. gdy każda współrzędna $f_k : U \to \mathbb{R}$ jest ciągła w punkcie $\mathbf{a}$.

Dowód. Według definicji Heinego (7.33) należy pokazać zbieżność $\mathbf{f}(\mathbf{x}_k) \to \mathbf{f}(\mathbf{a})$ jako wektorów w $\mathbb{R}^m$. Twierdzenie 2 z § 7.1 mówi, że taka zbieżność jest równoważna zbieżności ciągów j-tych współrzędnych $f_j(\mathbf{x}_k) \to f_j(\mathbf{a})$ dla każdego $j \in \{1, ..., m\}$. A to na mocy Tw. 1 z poprzedniego paragrafu oznacza ciągłość każdej funkcji f_j w punkcie $\mathbf{a}$. ■

W przypadku różniczkowania rozważanych w tym paragrafie funkcyj z $\mathbb{R}^n$ w $\mathbb{R}^m$ trzeba wprowadzić pochodne cząstkowe poszczególnych funkcyj-współrzędnych względem poszczególnych zmiennych

$$\frac{\partial f_k}{\partial x_j}, \quad k \in \{1, ..., m\},\ j \in \{1, ..., n\},$$

które najwygodniej jest ułożyć w następującą macierz $m \times n$

$$Df(\mathbf{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_2}{\partial x_n}(\mathbf{a}) \\ \dots\dots & \dots\dots & \dots & \dots\dots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) & \frac{\partial f_m}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}, \tag{7.36}$$

którą będziemy nazywać *pochodną macierzową funkcji* $\mathbf{f}$. Wtedy zapis $D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})$ oznacza mnożenie macierzy $D\mathbf{f}(\mathbf{a})$ przez wektor-kolumnę $(\mathbf{x} - \mathbf{a})^T$ o n elementach:

$$D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_2}{\partial x_n}(\mathbf{a}) \\ \dots\dots & \dots\dots & \dots & \dots\dots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) & \frac{\partial f_m}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{a}) \end{pmatrix} \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \\ \dots\dots \\ x_n - a_n \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a})(x_1 - a_1) + \frac{\partial f_1}{\partial x_2}(\mathbf{a})(x_2 - a_2) + \dots + \frac{\partial f_1}{\partial x_n}(\mathbf{a})(x_n - a_n) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a})(x_1 - a_1) + \frac{\partial f_2}{\partial x_2}(\mathbf{a})(x_2 - a_2) + \dots + \frac{\partial f_2}{\partial x_n}(\mathbf{a})(x_n - a_n) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a})(x_1 - a_1) + \frac{\partial f_m}{\partial x_2}(\mathbf{a})(x_2 - a_2) + \dots + \frac{\partial f_m}{\partial x_n}(\mathbf{a})(x_n - a_n) \end{pmatrix},$$

co daje wektor-kolumnę o m elementach.

PRZYKŁAD 1. Niech funkcja $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^2$ będzie dana wzorem $\mathbf{f}(x, y, z) = (ze^x, -ye^z)$. Jest to funkcja ciągła, bo jej obie funkcje-współrzędne są funkcjami ciągłymi. Macierz jej pochodnych cząstkowych ma postać

$$D\mathbf{f}(x, y, z) = \begin{pmatrix} ze^x & 0 & e^x \\ 0 & -e^z & -ye^z \end{pmatrix}.$$

Te pochodne są są funkcjami ciągłymi, więc $\mathbf{f} \in C^1$.

Dla $m = 1$ nie piszemy tłustego f, bo to jest funkcja o wartościach rzeczywistych; w tym przypadku macierz Df jest jednowierszowa:

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

i dlatego pochodna macierzowa takiej funkcji pokrywa się z gradientem.

Wspomniane wcześniej krzywe $\boldsymbol{\sigma} : \mathbb{R} \to \mathbb{R}^m$ są funkcjami rozważanego obecnie typu, chociaż jako funkcje jednej zmiennej nie mają pochodnych cząstkowych. Dla nich macierz (7.36) należałoby zapisać jako kolumnę

$$D\boldsymbol{\sigma}(a) = \begin{pmatrix} \frac{d\sigma_1}{dx}(a) \\ \frac{d\sigma_2}{dx}(a) \\ \dots\dots \\ \frac{d\sigma_m}{dx}(a) \end{pmatrix} = \frac{d\boldsymbol{\sigma}}{dx}(\mathbf{a}).$$

Ten wektor w $\mathbb{R}^m$ można interpretować jako prędkość posuwania się po krzywej.

PRZYKŁAD 2. Niech funkcja $\boldsymbol{\sigma} : \mathbb{R} \to \mathbb{R}^3$ będzie dana wzorem $\boldsymbol{\sigma}(t) = (r\cos\omega t,\ r\sin\omega t,\ vt)$. Jest to funkcja ciągła. Opisuje ona *linię śrubową* (*helisę*) nawijającą się wokół trzeciej osi. Macierz jej pochodnych jest następującą kolumną

$$D\boldsymbol{\sigma}(t) = \begin{pmatrix} -r\omega\sin\omega t \\ r\omega\cos\omega t \\ v \end{pmatrix}.$$

Wyliczone pochodne są funkcjami ciągłymi, więc jest to funkcja klasy C^1.

Niech $U \subset \mathbb{R}^n$ oraz $\mathbf{f} : U \to \mathbb{R}^m$. Mówimy, że funkcja $\mathbf{f}$ jest *różniczkowalna w punkcie* $\mathbf{a} \in U$, jeśli dla wszystkich $j \in \{1, 2, \dots, n\}, k \in \{1, \dots, m\}$ istnieją pochodne cząstkowe $\partial_j f_k(\mathbf{a})$ oraz jeśli

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})}{||\ \mathbf{x} - \mathbf{a}\ ||} = 0. \tag{7.37}$$

Jest to z niewielką zmianą przepisanie wzoru (7.24) odpowiadającego definicji (7.22). Podobnie da się sformułować odpowiednik Tw. 2 z poprzedniego paragrafu.

Twierdzenie 3. Niech $U \subset \mathbb{R}^n$ oraz $\mathbf{f} : U \to \mathbb{R}^m$. Jeśli pochodne cząstkowe $\partial_j f_k$ istnieją dla wszystkich $j \in \{1, \dots, n\},\ k \in \{1, \dots, m\}$ i są ciągłe w punkcie $\mathbf{a}$, to funkcja $\mathbf{f}$ jest różniczkowalna w $\mathbf{a}$.

Dowód można znaleźć w książce: J.E. Marsden, A.J. Tromba *Vector Calculus*, Freeman and Company, New York 1988, w § 2.7.

Lemat 3 (*pochodna funkcji pomnożonej przez stałą*). Niech $U \subset \mathbb{R}^n$, niech funkcja $\mathbf{f} : U \to \mathbb{R}^m$ będzie różniczkowalna w punkcie $\mathbf{a}$ i niech $c \in \mathbb{R}$. Wtedy funkcja $\mathbf{h}(\mathbf{x}) = c\mathbf{f}(\mathbf{x})$ jest różniczkowalna w $\mathbf{a}$ oraz

$$D\mathbf{h}(\mathbf{a}) = cD\mathbf{f}(\mathbf{a}).$$

Dowód. Mamy wykazać, że

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{a}) - cD\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})}{||\, \mathbf{x}-\mathbf{a} \,||} = 0,$$

czyli że

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{c\mathbf{f}(\mathbf{x}) - c\mathbf{f}(\mathbf{a}) - cD\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})}{||\, \mathbf{x}-\mathbf{a} \,||} = 0.$$

Stałą c można wyciągnąć przed ułamek dzięki własnościom dodawania wektorów, a potem dzięki Lematowi 1 przed znak granicy. W takim razie dostajemy równość

$$c \lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})}{||\, \mathbf{x}-\mathbf{a} \,||} = 0,$$

która jest spełniona na mocy założonej różniczkowalności funkcji $\mathbf{f}$. ■

Lemat 4 (*pochodna sumy*). Niech $U \subset \mathbb{R}^n$, niech funkcje $\mathbf{f}, \mathbf{g} : U \to \mathbb{R}^m$ będą różniczkowalne w punkcie $\mathbf{a}$. Wtedy funkcja $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})$ jest różniczkowalna w $\mathbf{a}$ oraz

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

Dowód. Przekształcamy wyrażenie na granicę, która powinna być zerem:

$$\begin{aligned}
&\lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{a}) - [D\mathbf{f}(\mathbf{a}) + D\mathbf{g}(\mathbf{a})](\mathbf{x}-\mathbf{a})}{||\, \mathbf{x}-\mathbf{a} \,||} \\
&\quad = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - \mathbf{g}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}) - D\mathbf{g}(\mathbf{a})(\mathbf{x}-\mathbf{a})}{||\, \mathbf{x}-\mathbf{a} \,||} \\
&\quad = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}) + \mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{a}) - D\mathbf{g}(\mathbf{a})](\mathbf{x}-\mathbf{a})}{||\, \mathbf{x}-\mathbf{a} \,||}
\end{aligned}$$

Korzystamy z Lematu 2:

$$\begin{aligned}
&\lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{a}) - [D\mathbf{f}(\mathbf{a}) + D\mathbf{g}(\mathbf{a})](\mathbf{x}-\mathbf{a})}{||\, \mathbf{x}-\mathbf{a} \,||} \\
&\quad = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})}{||\, \mathbf{x}-\mathbf{a} \,||} + \lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{a}) - D\mathbf{g}(\mathbf{a})](\mathbf{x}-\mathbf{a})}{||\, \mathbf{x}-\mathbf{a} \,||}.
\end{aligned}$$

Obie granice z założenia są zerami, więc liczona granica jest zerem. Skoro tak, to funkcja $\mathbf{h}$ jest różniczkowalna i zachodzi potrzebna równość. ■

Twierdzenie 4 (*pochodna funkcji złożonej*). Niech $U \subset \mathbb{R}^n$ oraz $V \subset \mathbb{R}^m$ będą zbiorami otwartymi. Niech będą dane funkcje $\mathbf{g} : U \to V$, $\mathbf{f} : V \to \mathbb{R}^p$, dzięki czemu określone jest złożenie $\mathbf{f} \circ \mathbf{g}$. Niech $\mathbf{g}$ będzie różniczkowalna w punkcie $\mathbf{a}$ zaś $\mathbf{f}$ w punkcie $\mathbf{b} = \mathbf{g}(\mathbf{a})$. Wtedy funkcja $\mathbf{f} \circ \mathbf{g}$ jest różniczkowalna w punkcie $\mathbf{a}$ oraz zachodzi równość

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) = D\mathbf{f}(\mathbf{b}) \cdot D\mathbf{g}(\mathbf{a}), \tag{7.38}$$

gdzie kropka $\cdot$ oznacza mnożenie macierzy.

Dowód można znaleźć w książce: J.E. Marsden, A.J. Tromba *Vector Calculus*, Freeman and Company, New York 1988, w § 2.7.

Jeśli wprowadzimy oznaczenie $\mathbf{y} = \mathbf{g}(\mathbf{x})$ i uznamy, że $\mathbf{f}$ jest funkcją zmiennej $\mathbf{y} \in V$, to według wzoru (7.38) pochodne cząstkowe funkcji złożonej wyrażają się jako sumy

$$\frac{\partial(\mathbf{f} \circ \mathbf{g})_l}{\partial x_j} = \sum_{k=1}^{m} \frac{\partial f_l}{\partial y_k} \frac{\partial g_k}{\partial x_j}. \tag{7.39}$$

Wzór (7.25) jest szczególnym przypadkiem tego wzoru, trzeba tylko zauważyć, że tam $p = 1$ oraz $\mathbf{g} = \boldsymbol{\sigma}$, a zmienne oznaczone są innymi literami.

PRZYKŁAD 3. Niech $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ zaś $f : \mathbb{R}^m \to \mathbb{R}$ niech będzie dana wzorem $f(\mathbf{y}) = \sin(\mathbf{y} \cdot \mathbf{y}) = \sin(y_1^2 + y_2^2 + \ldots + y_m^2)$. Wtedy $h : \mathbb{R}^n \to \mathbb{R}$ jako złożenie $h = f \circ \mathbf{g}$ jest dana wzorem $h(\mathbf{x}) = \sin[\mathbf{g}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x})]$. Pochodna macierzowa, czyli gradient funkcji f wynosi

$$\begin{aligned}\nabla f(\mathbf{y}) &= (2y_1 \cos(\mathbf{y} \cdot \mathbf{y}), 2y_2 \cos(\mathbf{y} \cdot \mathbf{y}), \ldots, 2y_m \cos(\mathbf{y} \cdot \mathbf{y})) \\ &= 2\cos(\mathbf{y} \cdot \mathbf{y})\ (y_1, y_2, \ldots, y_m) = 2\cos(\mathbf{y} \cdot \mathbf{y})\, \mathbf{y}.\end{aligned}$$

W takim razie na podstawie (7.39) pochodną macierzową funkcji h możemy zapisać

$$Dh(\mathbf{x}) = \nabla f(\mathbf{y}) \cdot D\mathbf{g}(\mathbf{x}) = 2\cos[\mathbf{g}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x})]\ \mathbf{g}(\mathbf{x}) \cdot D\mathbf{g}(\mathbf{x}).$$

Jako iloczyn wiersza $\mathbf{y} = \mathbf{g}(\mathbf{x})$ i macierzy prostokątnej $D\mathbf{g}(\mathbf{x})$ jest to macierz-wiersz. I to dobrze, bo wartości funkcji h są rzeczywiste, więc jej pochodna macierzowa jest gradientem ∇h.

Twierdzenie 5 (*wzór Taylora*). Niech $U \subset \mathbb{R}^n$ i funkcja $f : U \to \mathbb{R}$ ma ciągłe pochodne cząstkowe trzeciego rzędu. Wtedy zachodzi

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)\, h_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0)\, h_i h_j + R_2(\mathbf{h}, \mathbf{x}_0),$$

gdzie $R_2(\mathbf{h}, \mathbf{x}_0)/||\mathbf{h}||^2 \to 0$ przy $\mathbf{h} \to 0$.

Dowód. W wyrażeniu $f(\mathbf{x}_0 + t\mathbf{h})$ widzimy funkcję złożoną $(f \circ \boldsymbol{\sigma})$, gdzie $\boldsymbol{\sigma}(t) = \mathbf{x}_0 + t\mathbf{h}$. W tym wypadku $D\boldsymbol{\sigma}(t) = \frac{d\boldsymbol{\sigma}}{dt} = \mathbf{h}$. Stosujemy wzór (7.39) na pochodną funkcji złożonej:

$$\frac{d}{dt} f(\mathbf{x}_0 + t\mathbf{h}) = [Df(\mathbf{x}_0 + t\mathbf{h})] \cdot \mathbf{h} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_0 + t\mathbf{h})\, h_i.$$

Całkujemy tę równość stronami względem t:

$$f(\mathbf{x}_0 + t\mathbf{h}) + C = \int \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_0 + t\mathbf{h})\, h_i\, dt = \sum_{i=1}^{n} \int \frac{\partial f}{\partial x_i}(\mathbf{x}_0 + t\mathbf{h})\, h_i\, dt.$$

Całkujemy przez części podstawiając $v' = 1,\ u = \frac{\partial f}{\partial x_i} h_i$; stąd $u' = \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0 + t\mathbf{h})\, h_i h_j$ i wybieramy $v = t - 1$:

$$f(\mathbf{x}_0 + t\mathbf{h}) + C = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_0 + t\mathbf{h})\, h_i(t-1) - \sum_{i,j=1}^{n} \int (t-1) \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0 + t\mathbf{h})\, h_i h_j\, dt.$$

Teraz liczymy całkę oznaczoną w granicach od $t = 0$ do $t = 1$:

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = -\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)\, h_i(-1) + \sum_{i,j=1}^{n} \int_0^1 (1-t) \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0 + t\mathbf{h})\, h_i h_j\, dt.$$

W ten sposób udowodniliśmy tożsamość

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)\, h_i + R_1(\mathbf{h}, \mathbf{x}_0), \tag{7.40}$$

gdzie

$$R_1(\mathbf{h}, \mathbf{x}_0) = \sum_{i,j=1}^{n} \int_0^1 (1-t) \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0 + t\mathbf{h})\, h_i h_j\, dt. \tag{7.41}$$

Całkujemy wyrażenie (7.41) przez części podstawiając

$$u = \frac{\partial^2 f}{\partial x_i \partial x_j}\, h_i h_j \quad \text{oraz} \quad v' = 1 - t.$$

Wówczas

$$u' = \sum_{k=1}^{n} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}\, h_i h_j h_k \quad \text{oraz} \quad v = -\frac{(t-1)^2}{2}$$

i otrzymujemy całkę oznaczoną

$$R_1(\mathbf{h}, \mathbf{x}_0) =$$

$$= -\sum_{i,j=1}^{n} \frac{(t-1)^2}{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0 + t\mathbf{h})\, h_i h_j \Big|_0^1 + \sum_{i,j,k=1}^{n} \int_0^1 \frac{(t-1)^2}{2} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\mathbf{x}_0 + t\mathbf{h})\, h_i h_j h_k\, dt.$$

$$= \sum_{i,j=1}^{n} \frac{1}{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0)\, h_i h_j + \sum_{i,j,k=1}^{n} \int_0^1 \frac{(t-1)^2}{2} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\mathbf{x}_0 + t\mathbf{h})\, h_i h_j h_k\, dt.$$

W takim razie wykazaliśmy, że

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)\, h_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0)\, h_i h_j + R_2(\mathbf{h}, \mathbf{x}_0), \tag{7.42}$$

gdzie

$$R_2(\mathbf{h}, \mathbf{x}_0) = \sum_{i,j,k=1}^{n} \left[\int_0^1 \frac{(t-1)^2}{2} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\mathbf{x}_0 + t\mathbf{h})\, dt \right] h_i h_j h_k\, . \tag{7.43}$$

Szacujemy moduł tej reszty korzystając z nierówności trójkąta

$$|R_2(\mathbf{h},\mathbf{x}_0)| = \left|\sum_{i,j,k=1}^{n}\left[\int_0^1 \frac{(t-1)^2}{2}\frac{\partial^3 f}{\partial x_i\partial x_j\partial x_k}(\mathbf{x}_0+t\mathbf{h})\,dt\right] h_i h_j h_k\right|$$
$$\leq \sum_{i,j,k=1}^{n}\left|\int_0^1 \frac{(t-1)^2}{2}\frac{\partial^3 f}{\partial x_i\partial x_j\partial x_k}(\mathbf{x}_0+t\mathbf{h})\,dt\right| |h_i|\,|h_j|\,|h_k|.$$

Każda współrzędna h_j wektora $\mathbf{h}$ spełnia nierówność $|h_j| \leq ||\mathbf{h}||$, więc

$$|R_2(\mathbf{h},\mathbf{x}_0)| \leq \sum_{i,j,k=1}^{n}\left|\int_0^1 \frac{(t-1)^2}{2}\frac{\partial^3 f}{\partial x_i\partial x_j\partial x_k}(\mathbf{x}_0+t\mathbf{h})\,dt\right| ||\mathbf{h}||^3.$$

wobec tego istnieje stała

$$M = \sum_{i,j,k=1}^{n}\left|\int_0^1 \frac{(t-1)^2}{2}\frac{\partial^3 f}{\partial x_i\partial x_j\partial x_k}(\mathbf{x}_0+t\mathbf{h})\,dt\right| \geq 0$$

taka, że

$$|R_2(\mathbf{h},\mathbf{x}_0)| \leq M\,||\mathbf{h}||^3.$$

Stąd wynika

$$\frac{|R_2(\mathbf{h},\mathbf{x}_0)|}{||\mathbf{h}||^2} \leq M\,||\mathbf{h}|| \to 0 \quad \text{gdy} \quad \mathbf{h}\to 0.$$

Wzór (7.42) jest tym wzorem, jaki mieliśmy udowodnić, a ostatnia zbieżność jest tą, jaką należało wykazać. ■

Równość (7.40) nazywa się *wzorem Taylora pierwszego rzędu*, a (7.42) *wzorem Taylora drugiego rzędu*. Przez analogię można napisać *wzór Taylora trzeciego rzędu*:

$$f(\mathbf{x}_0+\mathbf{h}) = f(\mathbf{x}_0) + \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(\mathbf{x}_0)\,h_i + \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_i\partial x_j}(\mathbf{x}_0)\,h_i h_j$$
$$+\frac{1}{3!}\sum_{i,j,k=1}^{n}\frac{\partial^3 f}{\partial x_i\partial x_j\partial x_k}(\mathbf{x}_0)\,h_i h_j h_k + R_3(\mathbf{h},\mathbf{x}_0), \tag{7.44}$$

i dalszych rzędów. W ten sposób znaleźliśmy odpowiednik poznanego wcześniej wzoru Taylora dla funkcji wielu zmiennych.

PRZYKŁAD 4. Znajdziemy wzór Taylora dugiego rzędu dla $f(x,y) = \sin(x+2y)$ wokół punktu $\mathbf{x}_0 = (0,0)$. Mamy $f(0,0) = 0$ tudzież

$$\partial_x f(x,y) = \cos(x+2y), \qquad \partial_y f(x,y) = 2\cos(x+2y),$$

$$\partial_x f(0,0) = \cos(0+2\cdot 0) = 1, \quad \partial_y f(0,0) = 2\cos(0+2\cdot 0) = 2,$$
$$\partial_x\partial_x f(0,0) = -\sin(0+2\cdot 0) = 0, \quad \partial_x\partial_y f(0,0) = -2\sin(0+2\cdot 0) = 0,$$
$$\partial_y\partial_y f(0,0) = -4\sin(0+2\cdot 0) = 0.$$

Zatem

$$f(\mathbf{0}+\mathbf{h}) = h_1 + 2h_2 + R_2(\mathbf{h},\mathbf{0}).$$

Niech $U \subset \mathbb{R}^n$ i $f : U \to \mathbb{R}$ jest daną funkcją. Punkt $\mathbf{a} \in U$ nazywa się *minimum lokalnym*, jeśli istnieje niby-kula otwarta $K(\mathbf{a}, r)$ taka, że dla wszystkich $\mathbf{x} \in K(\mathbf{a}, r)$ zachodzi $f(\mathbf{x}) \geq f(\mathbf{a})$. Jeśli ponadto zachodzi $f(\mathbf{x}) > f(\mathbf{a})$ dla $\mathbf{x} \neq \mathbf{a}$, to punkt $\mathbf{a}$ nazywa się *ścisłym minimum lokalnym.* Podobnie $\mathbf{b} \in U$ nazywa się *maksimum lokalnym*, jeśli istnieje niby-kula otwarta $K(\mathbf{b}, r)$ taka, że dla wszystkich $\mathbf{x} \in K(\mathbf{b}, r)$ zachodzi $f(\mathbf{x}) \leq f(\mathbf{b})$. Jeśli ponadto $f(\mathbf{x}) < f(\mathbf{b})$ dla $\mathbf{x} \neq \mathbf{b}$, to punkt $\mathbf{b}$ nazywa się *ścisłym maksimum lokalnym.* Mówimy, że punkt $\mathbf{a}$ jest *ekstremum lokalnym*, jeśli jest minimum lokalnym albo maksimum lokalnym. Jeśli w $\mathbf{x}_0$ zeruje się pochodna $Df(\mathbf{x}_0) = 0$, to punkt $\mathbf{x}_0$ nazywa się *punktem krytycznym.* Punkt krytyczny, który nie jest lokalnym ekstremum, nazywa się *punktem siodłowym.*

Twierdzenie 6. Jeśli $U \subset \mathbb{R}^n$ jest otwarty, funkcja $f : U \to \mathbb{R}$ jest różniczkowalna oraz $\mathbf{a} \in U$ jest lokalnym ekstremum, to $Df(\mathbf{a}) = 0$. Krócej: każde lokalne ekstremum jest punktem krytycznym.

Dowód. Załóżmy, że funkcja f osiąga w $\mathbf{a}$ ekstremum lokalne. Wtedy dla każdego $\mathbf{h} \in \mathbb{R}^n$ funkcja $g(t) = f(\mathbf{a} + t\mathbf{h})$ ma ekstremum lokalne w $t = 0$. Na mocy Tw. 1 z § 4.2 zachodzi $g'(0) = 0$. Z drugiej strony g jest funkcją złożoną $g = f \circ \boldsymbol{\sigma}$ dla $\boldsymbol{\sigma}(t) = \mathbf{a} + \mathbf{h}t$ oraz $D\boldsymbol{\sigma}(0) = d\boldsymbol{\sigma}/dt = \mathbf{h}$. Na mocy Tw. 4 mamy

$$g'(0) = [\nabla f(\mathbf{a})] \cdot \mathbf{h}.$$

W ten sposób $[\nabla f(\mathbf{a})] \cdot \mathbf{h} = 0$ dla dowolnych $\mathbf{h}$. Jedynym wektorem prostopadłym do wszystkich wektorów jest wektor zerowy, więc $\nabla f(\mathbf{a}) = 0$. ■

PRZYKŁAD 5. Znajdziemy punkty krytyczne funkcji $f(x, y) = x^2y + xy^2$. Różniczkowanie daje

$$\frac{\partial f}{\partial x} = 2xy + y^2, \qquad \frac{\partial f}{\partial y} = x^2 + 2xy.$$

Przyrównujemy te pochodne do zera

$$2xy + y^2 = 0, \qquad 2xy + x^2 = 0, \tag{7.45}$$

$$y^2 = -2xy, \qquad x^2 = -2xy.$$

Stąd $x^2 = y^2$, $x = \pm y$. Podstawienie $x = y$ do pierwszego z równań (7.45) daje

$$2y^2 + y^2 = 3y^2 = 0,$$

więc $y = 0$, a stąd $x = 0$. Kiedy $x = -y$, to samo równanie daje

$$-2y^2 + y^2 = -y^2 = 0,$$

więc $y = 0$, a stąd $x = 0$. W ten sposób jedynym punktem krytycznym jest początek układu $(0, 0)$. Czy to jest ekstremum? Dla prostej $x = y$ dostajemy $f(x, x) = 2x^3$, a ta funkcja jednej zmiennej jest zarówno dodatnia jak i ujemna w pobliżu zera. Dlatego stwierdzamy, że punkt krytyczny $(0, 0)$ nie jest ekstremum lokalnym – trzeba go nazwać punktem siodłowym. Ten przykład pokazuje, że w Twierdzeniu 6 wynikanie w drugą stronę nie jest prawdziwe.

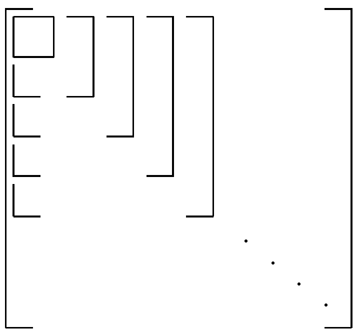
Niech $U \subset \mathbb{R}^n$, a funkcja $f : U \to \mathbb{R}$ ma wszystkie pochodne cząstkowe drugiego rzędu $f''_{x_i.x_j}(\mathbf{a})$ w punkcie $\mathbf{a}$. *Macierzą Hessego* funkcji f w punkcie $\mathbf{a}$ jest macierz kwadratowa określona wzorem

$$Hf(\mathbf{a}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \ldots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\mathbf{a}) & \ldots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{a}) \\ \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{a}) & \ldots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{pmatrix}, \tag{7.46}$$

[Ludwig Otto Hesse (1811-1874) – matematyk niemiecki, wprowadził swoją macierz drugich pochodnych w 1844 roku.] Macierz Hessego jest przeważnie stosowana w punktach krytycznych $\mathbf{x}_0 \in U$, dlatego pierwsze pochodne cząstkowe są tam zerami i wzór Taylora można zapisać

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \frac{1}{2} \sum_{i,j=1}^{n} Hf(\mathbf{x}_0)_{ij} h_i h_j + R_2(\mathbf{h}, \mathbf{x}_0).$$

W ten sposób w punkcie krytycznym macierz Hessego występuje w pierwszym nie stałym wyrazie w rozwinięciu Taylora danej funkcji. Wyznacznik macierzy Hessego nazywa się *wyznacznikiem Hessego* albo *hesjanem* funkcji f. Do następnego twierdzenia potrzebny jest hesjan i jego podwyznaczniki coraz mniejszego rzędu powstające przez odrzucenie ostatniego wiersza i ostatniej kolumny, zob. Rys. 7.18.



Rysunek 7.18

Twierdzenie 7. Niech $U \subset \mathbb{R}^n$, niech funkcja $f : U \to \mathbb{R}$ będzie klasy C^3 i niech $\mathbf{x}_0 \in U$ będzie punktem krytycznym tej funkcji. Jeśli element $(Hf)_{11}(\mathbf{x}_0)$ macierzy Hesego jest dodatni i wszystkie podwyznaczniki coraz wyższego rzędu leżące na głównej przekątnej aż do hesjanu są dodatnie, to $\mathbf{x}_0$ jest ścisłym minimum lokalnym funkcji f. Jeśli element $(Hf)_{11}$ macierzy Hessego jest ujemny, a podwyznaczniki coraz wyższego rzędu leżące na głównej przekątnej są na przemian dodatnie i ujemne, to $\mathbf{x}_0$ jest ścisłym maksimum lokalnym funkcji f.

Dowód (dla szczególnego przypadku $n = 2$) można znaleźć w książce: J.E. Marsden, A.J. Tromba *Vector Calculus*, Freeman and Company, New York 1988, w § 4.2.

PRZYKŁAD 6. Rozważmy funkcję $f : \mathbb{R}^2 \to \mathbb{R}$ daną wzorem $f(x,y) = x^2-2xy+2y^2$. Jej pochodna macierzowa wynosi $\nabla f(x,y) = (2x-2y, -2x+4y)$. Szukamy punktów krytycznych przyrównując pochodną do zera:

$$2x - 2y = 0, \quad -2x + 4y = 0.$$

Z pierwszego równania mamy $y = x$; wstawiamy to do drugiego: $-2x+4x = 0$, co daje $x = 0$. W takim razie też $y = 0$ i dostajemy tylko jeden punkt krytyczny $\mathbf{r}_0 = (0,0)$.

Macierz Hessego ma postać

$$Hf(\mathbf{r}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x \partial x}(\mathbf{r}) & \frac{\partial^2 f}{\partial x \partial y}(\mathbf{r}) \\ \frac{\partial^2 f}{\partial y \partial x}(\mathbf{r}) & \frac{\partial^2 f}{\partial y \partial y}(\mathbf{r}) \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}.$$

Jej wyznacznik, czyli hesjan wynosi $\det H = 8 - 4 = 4$. Jako funkcja jest on stały, więc tyle samo wynosi w punkcie krytycznym. Jego podwyznacznik z głównej przekątnej to pierwszy element tej przekątnej wynoszący 2. Skoro oba te wyznaczniki są dodatnie, to na mocy Tw. 7 funkcja f ma lokalne minimum w początku układu współrzędnych. Widać to także z jej zapisania w innej postaci $f(x,y) = (x-y)^2 + y^2$, skąd $f(x,y) > 0$ dla $(x,y) \neq (0,0)$

Weźmy funkcję $g : \mathbb{R}^2 \to \mathbb{R}$ przeciwną do poprzedniej: $g(x,y) = -x^2 + 2xy - 2y^2$. Możemy z góry powiedzieć, że jej jedyny punkt krytyczny jest taki sam, jak funkcji f, a z powodu jej przeciwnego znaku – w tym punkcie jest lokalne maksimum. Macierz Hessego tej funkcji

$$Hg(\mathbf{r}) = \begin{pmatrix} \frac{\partial^2 g}{\partial x \partial x}(\mathbf{r}) & \frac{\partial^2 g}{\partial x \partial y}(\mathbf{r}) \\ \frac{\partial^2 g}{\partial y \partial x}(\mathbf{r}) & \frac{\partial^2 g}{\partial y \partial y}(\mathbf{r}) \end{pmatrix} = \begin{pmatrix} -2 & 2 \\ 2 & -4 \end{pmatrix}$$

ma wyznacznik tego samego znaku: $\det Hg = 4$, ale pierwszy element głównej przekątnej jest ujemny. Według Tw. 7 funkcja g ma lokalne maksimum – zgodnie z pierwszym spostrzeżeniem.

7.3.1 Funkcja kwadratowa

Zajmiemy się teraz uogólnieniem trójmianu kwadratowego na funkcje wielu zmiennych. Niech funkcja $f : \mathbb{R}^n \to \mathbb{R}$ będzie dana wzorem

$$f(\mathbf{x}) = a\,\mathbf{x}^2 + \mathbf{b} \cdot \mathbf{x} + c, \tag{7.47}$$

gdzie $a, c \in \mathbb{R}$, $a \neq 0$, $\mathbf{b} \in \mathbb{R}^n$. Jest to funkcja bardzo podobna do trójmianu kwadratowego jednej zmiennej. Tym razem $\mathbf{x}^2$ jest kwadratem skalarnym wektora $\mathbf{x}$, a stała $\mathbf{b}$ jest wektorem stałym pommnożonym skalarnie przez $\mathbf{x}$. Pochodna macierzowa tej funkcji wynosi

$$\nabla f = 2a\,\mathbf{x} + \mathbf{b}.$$

Szukamy punktów krytycznych tej funkcji przyrównując pochodną do zera:

$$2a\,\mathbf{x}_0 + \mathbf{b} = 0, \qquad \mathbf{x}_0 = -\frac{\mathbf{b}}{2a}.$$

Okazuje się, że jest jeden punkt krytyczny $\mathbf{x}_0$.

Macierz Hessego ma postać

$$Hf(\mathbf{x}) = \begin{pmatrix} \partial_1\partial_1 f & \partial_1\partial_1 f & \dots & \partial_1\partial_n f \\ \partial_2\partial_1 f & \partial_2\partial_2 f & \dots & \partial_2\partial_n f \\ \dots & \dots & \dots & \dots \\ \partial_n\partial_1 f & \partial_n\partial_2 f & \dots & \partial_n\partial_n f \end{pmatrix} = \begin{pmatrix} 2a & 0 & \dots & 0 \\ 0 & 2a & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 2a \end{pmatrix}.$$

Jej wyznacznik wynosi $\det H = (2a)^n$, jego coraz mniejsze podwyznaczniki wynoszą kolejno $(2a)^{n-1}, \dots, (2a)^2$, $2a$. Jeśli $a < 0$, to te liczby są na przemian ujemne i dodatnie, więc według Tw. 7 w punkcie krytycznym $\mathbf{x}_0$ jest maksimum. Jeśli zaś $a > 0$, to liczby $(2a)^{n-1}, \dots, (2a)^3$, $(2a)^2$, $2a$ są dodatnie i według Tw. 7 w punkcie $\mathbf{x}_0$ jest minimum. Widzimy w tym podobieństwo do trójmianu kwadratowego $ax^2 + bx + c$, który dla $a < 0$ ma jedno maksimum, zaś dla $a > 0$ ma jedno minimum.

Pytamy teraz, czy są miejsca zerowe funkcji (7.47). Podstawiamy wartość krytyczną $\mathbf{x}_0$ do funkcji:

$$f(\mathbf{x}_0) = a\left(-\frac{\mathbf{b}}{2a}\right)^2 + \mathbf{b}\left(-\frac{\mathbf{b}}{2a}\right) + c = \frac{\mathbf{b}^2}{4a} - \frac{\mathbf{b}^2}{2a} + c = -\frac{\mathbf{b}^2}{4a} + c.$$

Niech $a > 0$, wiemy już, że w $\mathbf{x}_0$ jest minimum. Jeśli owa wartość minimalna jest większa od zera, to funkcja (7.47) nie może osiągać zera. Zatem dla przyjmowania wartości zero przez tę funkcję musi być spełniony warunek

$$-\frac{\mathbf{b}^2}{4a} + c \le 0,$$

skąd

$$-\mathbf{b}^2 + 4ac \le 0, \qquad \mathbf{b}^2 - 4ac \ge 0. \tag{7.48}$$

Jeśli $a < 0$, to w $\mathbf{x}_0$ jest maksimum. Aby mogły być miejsca zerowe funkcji (7.47), to owo maksimum musi być nieujemne:

$$-\frac{\mathbf{b}^2}{4a} + c \ge 0,$$

skąd

$$-\mathbf{b}^2 + 4ac \le 0, \qquad \mathbf{b}^2 - 4ac \ge 0. \tag{7.49}$$

Oba warunki (7.48) i (7.49) są jednakowe, więc dla każdego $a \ne 0$ warunkiem istnienia miejsc zerowych jest nierówność

$$\Delta = \mathbf{b}^2 - 4ac \ge 0. \tag{7.50}$$

Wyrażenie po lewej stronie można nazwać *wyróżnikiem*. Jest on bardzo podobny do wyróżnika funkcji kwadratowej jednej zmiennej. Jedyna różnica jest w symbolu $\mathbf{b}$. Tam była to liczba rzeczywista, tu jest wektor z $\mathbb{R}^n$.

Na koniec znajdziemy miejsca zerowe funkcji (7.47). Przkształćmy ją tak, żeby znaleźć kwadrat pewnego jednomianu:

$$f(\mathbf{x}) = a\left(\mathbf{x}^2 + \frac{\mathbf{b}}{a}\cdot\mathbf{x}\right) + c = a\left(\mathbf{x}^2 + \frac{\mathbf{b}}{a}\cdot\mathbf{x} + \frac{\mathbf{b}^2}{4a^2}\right) - \frac{\mathbf{b}^2}{4a} + c = a\left(\mathbf{x} + \frac{\mathbf{b}}{2a}\right)^2 - \frac{\mathbf{b}^2}{4a} + c.$$

Przyrównujemy to do zera i przekształcamy dalej

$$a\left(\mathbf{x}+\frac{\mathbf{b}}{2a}\right)^2-\frac{\mathbf{b}^2}{4a}+c=0.$$

$$a\left(\mathbf{x}+\frac{\mathbf{b}}{2a}\right)^2=\frac{\mathbf{b}^2}{4a}-c.$$

$$\left(\mathbf{x}+\frac{\mathbf{b}}{2a}\right)^2=\frac{\mathbf{b}^2-4ac}{4a^2}=\frac{\Delta}{4a^2}.$$

Po lewej stronie mamy kwadrat wektora, po prawej pewien skalar – nieujemny na mocy warunku (7.50). Jeśli po prawej stronie dopiszemy kwadrat wektora jednostkowego $\mathbf{n}^2=1$, to otrzymamy

$$\left(\mathbf{x}+\frac{\mathbf{b}}{2a}\right)^2=\frac{\Delta}{4a^2}\mathbf{n}^2=\left(\frac{\sqrt{\Delta}}{2a}\mathbf{n}\right)^2.$$

Skoro kwadraty wektorów są równe, to i same wektory mogą być równe:

$$\mathbf{x}+\frac{\mathbf{b}}{2a}=\frac{\sqrt{\Delta}}{2a}\mathbf{n}.$$

Stąd

$$\mathbf{x}=-\frac{\mathbf{b}}{2a}+\frac{\sqrt{\Delta}}{2a}\mathbf{n}=\frac{-\mathbf{b}+\mathbf{n}\sqrt{\Delta}}{2a}. \tag{7.51}$$

Otrzymaliśmy to, czego szukaliśmy, czyli zbiór wektorów, dla których funkcja (7.47) przyjmuje wartość zero. Skoro wektor jednostkowy $\mathbf{n}$ jest dowolny, to mamy nieskończenie wiele takich wektorów. Ich końce układają się w niby-sferę o środku w punkcie $-\mathbf{b}/2a$ i o promieniu $\sqrt{\Delta}/2|a|$. Jest to jedna z poziomic rozważanej funkcji.

7.3.2 Ekstrema warunkowe

Czasami pojawia się zagadnienie szukania największej wartości funkcji poddanej pewnym *więzom* lub *warunkom dodatkowym.* Typowym zadaniem jest na przykład pytanie o wartość maksymalną jednej funkcji $f(x,y)$ przy dodatkowym warunku $g(x,y)=c$ zadanym przez inną funkcję. Taka wartość maksymalna nazywa się *maksimum warunkowym.* Celem tego fragmentu wykładu jest rozwinięcie metod pozwalających rozwiązać tego typu zadania. Potrzebne jest nowe oznaczenie. Niech $S\subset U\subset\mathbb{R}^n$, $f:U\to\mathbb{R}$; symbol $f|S$ oznacza *zawężenie funkcji f do zbioru S*, czyli tę samą funkcję f, ale o mniejszej dziedzinie S zamiast U.

Twierdzenie 8. Niech $U\subset\mathbb{R}^n$ i niech $f,g:U\to\mathbb{R}$ dwie funkcje klasy C^1. Niech $\mathbf{x}_0\in U$, $g(\mathbf{x}_0)=c$, $\nabla g(\mathbf{x}_0)\neq 0$ i niech S będzie poziomicą dla funkcji g odpowiadającą wartości c. Jeśli funkcja $f|S$ ma ekstremum lokalne w $\mathbf{x}_0$, to istnieje $\lambda\in\mathbb{R}$ takie, że

$$\nabla f(\mathbf{x}_0)=\lambda\nabla g(\mathbf{x}_0). \tag{7.52}$$

D o w ó d dla przypadku $n=3$. Płaszczyzna styczna do S w punkcie $\mathbf{x}_0$ to płaszczyzna przechodząca przez ten punkt i prostopadła do $\nabla g(\mathbf{x}_0)$. Weźmy krzywą K leżącą na poziomicy S, tzn. $\boldsymbol{\sigma}(t) \in S$ dla każdego $t \in \langle a, b\rangle$ taką, że $\boldsymbol{\sigma}(0) = \mathbf{x}_0$. Jeśli funkcja $f|S$ ma ekstremum w $\mathbf{x}_0$, to funkcja $f(\boldsymbol{\sigma}(t))$ ma ekstremum w punkcie $t = 0$. Na mocy wzoru (7.26) mamy

$$0 = \frac{d}{dt} f(\boldsymbol{\sigma}(t))\bigg|_{t=0} = \nabla f(\mathbf{x}_0) \cdot \boldsymbol{\sigma}'(0).$$

W takim razie wektor $\nabla f(\mathbf{x}_0)$ jest prostopadły do wektora stycznego do każdej krzywej w S, więc również prostopadły do płaszczyzny stycznej do S w $\mathbf{x}_0$. Wiemy, że $\nabla g(\mathbf{x}_0)$ jest innym wektorem prostopadłym do tej samej płaszczyzny. Wobec tego oba wektory są równoległe do siebie. Skoro $\nabla g(\mathbf{x}_0) \neq 0$, to możemy napisać równość

$$\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0).$$

Gdyby było $\nabla g(\mathbf{x}_0) = 0$, zaś $\nabla f(\mathbf{x}_0) \neq 0$ to napisana równość nie mogłaby być spełniona dla żadnego $\lambda \in \mathbb{R}$. Dla $\nabla f(\mathbf{x}_0) = 0$ ta równość może być spełniona dla $\lambda = 0$. W takim przypadku ekstremum funkcji $f|S$ wypada w punkcie krytycznym funkcji f. ■

Liczba λ występująca we wzorze (7.52) nazywa się *mnożnikiem Lagrange'a.* Wiemy już, że wektor $\nabla f(\mathbf{x}_0)$ jest prostopadły do poziomicy funkcji f w punkcie $\mathbf{x}_0$, a z tego twierdzenia wiemy, że przy podanych założeniach jest również prostopadły do poziomicy funkcji g w tym samym punkcie. To oznacza również, że obie poiomice są do siebie styczne.

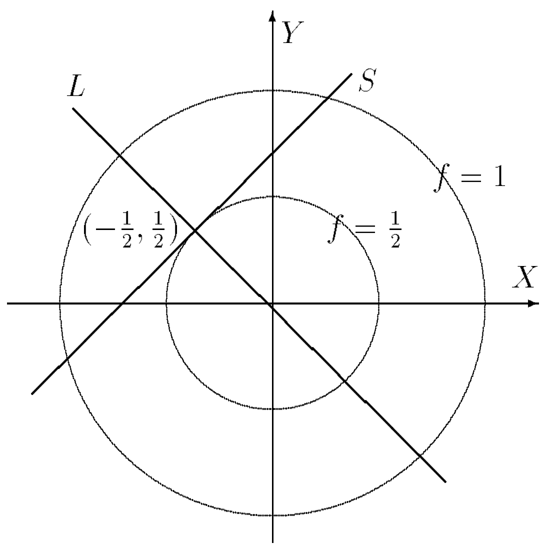
Otrzymany wynik mówi nam, że aby znaleźć ekstrema warunkowe funkcji f, trzeba go szukać wśród tych punktów $\mathbf{x}_0$, gdzie spełniony jest warunek (7.52). Ponadto trzeba znaleźć mnożnik λ. W takim razie trzeba rozwiązać układ $n + 1$ równań

$$\begin{aligned}
\frac{\partial f}{\partial x_1}(x_1, \ldots, x_n) &= \lambda \frac{\partial g}{\partial x_1}(x_1, \ldots, x_n),\\
\frac{\partial f}{\partial x_2}(x_1, \ldots, x_n) &= \lambda \frac{\partial g}{\partial x_2}(x_1, \ldots, x_n),\\
&\cdots\\
&\cdots\\
\frac{\partial f}{\partial x_n}(x_1, \ldots, x_n) &= \lambda \frac{\partial g}{\partial x_n}(x_1, \ldots, x_n),\\
g(x_1, \ldots, x_n) &= c
\end{aligned}$$

względem niewiadomych $x_1, \ldots, x_n$ i λ.

PRZYKŁAD 7. Niech $S \subset \mathbb{R}^2$ będzie prostą przechodzącą przez punkt (–1,0) nachyloną pod kątem 45^o do osi X i niech $f : \mathbb{R}^2 \to \mathbb{R}$ będzie dana wzorem $f(x, y) = x^2 + y^2$. Chcemy znaleźć ekstrema funkcji $f|S$.

Zbiór S można zapisać $S = \{(x, y) \in \mathbb{R}^2 : \ y = x + 1\}$, więc wybieramy $g(x, y) = y - x - 1$ oraz $c = 0$. Mamy $\nabla g(x, y) = (-1, 1) \neq 0$. Ekstrema warunkowe funkcji $f|S$ muszą się znaleźć wśród punktów, w ktorych $\nabla f(x, y) = (2x, 2y)$ jest prostopadłe do S, tzn. nachylone pod kątem $–45^o$. Jest to możliwe tylko wtedy, gdy $x = -y$, czyli gdy punkt (x, y) leży na prostej L przechodzącej przez początek układu i nachylonej pod kątem $–45^o$.To może się zdarzyć tylko w jednym punkcie przecięcia prostych S i L, patrz Rys. 7.19. Powołanie się na poziomice funkcji f wskazuje, że ten punkt $\left(-\frac{1}{2}, \frac{1}{2}\right)$ jest względnym minimum funkcji $f|S$. Nie jest to wcale minimum funkcji f, bo ona ma jedno minimum w początku układu.



Rysunek 7.19

PRZYKŁAD 8. Niech $S \subset \mathbb{R}^2$ będzie okręgiem o promieniu 1 i środku w początku układu i niech $f : \mathbb{R}^2 \to \mathbb{R}$ będzie dana wzorem $f(x, y) = x^2 - y^2$. Chcemy znaleźć ekstrema funkcji $f|S$.

Zbiór S jest poziomicą funkcji g o wartości 1, gdzie $g : \mathbb{R}^2 \to \mathbb{R}^2$, $g(x, y) = x^2 + y^2$. Funkcja f ma jako wykres powierzchnię siodłową z Rys. 7.5, a jako poziomice hiperbole, dla których osie X, Y są osiami symetrii zwierciadlanej, patrz Rys. 7.20. Poziomicami funkcji g są okręgi. Warunek $\nabla f(\mathbf{x}_0) = \lambda g(\mathbf{x}_0)$ oznacza, że poziomice obu funkcyj są do siebie styczne s punkcie $\mathbf{x}_0$. Zatem ekstremami funkcji $f|S$ mogą być cztery punkty $(0, \pm 1)$, $(\pm 1, 0)$. Licząc wartości $f(0, 1) = f(0, -1) = -1$, $f(1, 0) = f(-1, 0) = 1$ znajdujemy, że dwa punkty leżące na osi Y są minimami, a dwa na osi X maksimami.

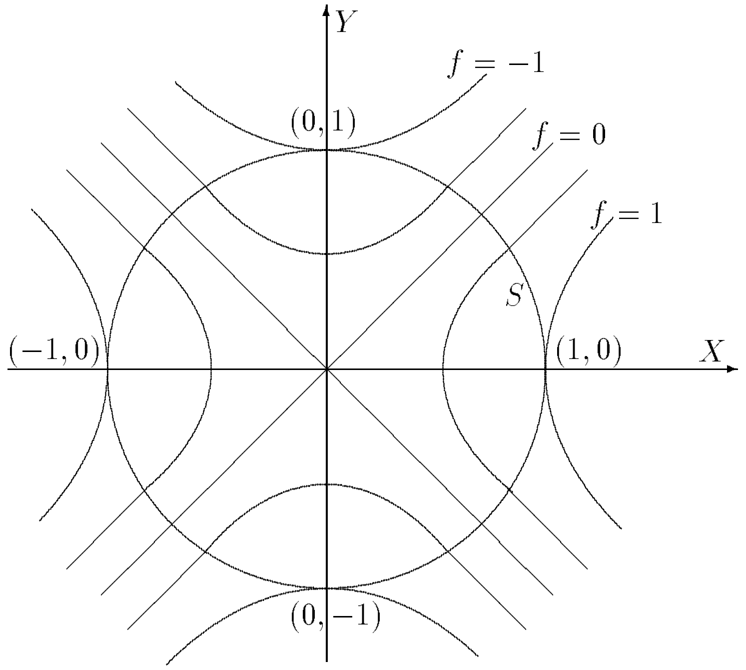
To było rozwiązanie geometryczne, bo wykorzystywalismy informacje z wykresów. Teraz rozwiążemy ten przykład metodą mnożników Lagrange'a. Wyliczmy gradienty

$$\nabla f(x, y) = (2x, -2y)\,, \qquad \nabla g(x, y) = (2x, 2y).$$

Na okręgu S, czyli dla $x^2 + y^2 = 1$ zachodzi $\nabla g(x, y) \neq 0$. Mamy znaleźć λ takie, że

$$(2x, -2y) = \lambda(2x, 2y) \quad \text{oraz} \quad x^2 + y^2 = 1.$$

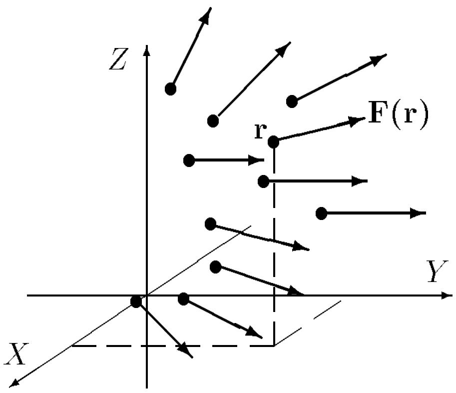
Są to trzy równania na trzy niewiadome x, y, λ. Z równości $2x = \lambda\, 2x$ mamy $x = 0$ lub $\lambda = 1$. Jeśli $x = 0$, to $y = \pm 1$, a $-2y = \lambda\, 2y$ daje $\lambda = -1$. Jeśli $\lambda = 1$, to $y = 0$, a $x = \pm 1$. W ten sposob otrzymujemy jak przedtem cztery punkty $(0, \pm 1)$, $(\pm 1, 0)$. Ta metoda tylko umiejscawia ekstrema. O tym, czy są to maksima, czy minima, musimy rozstrzygnąć innymi metodami.



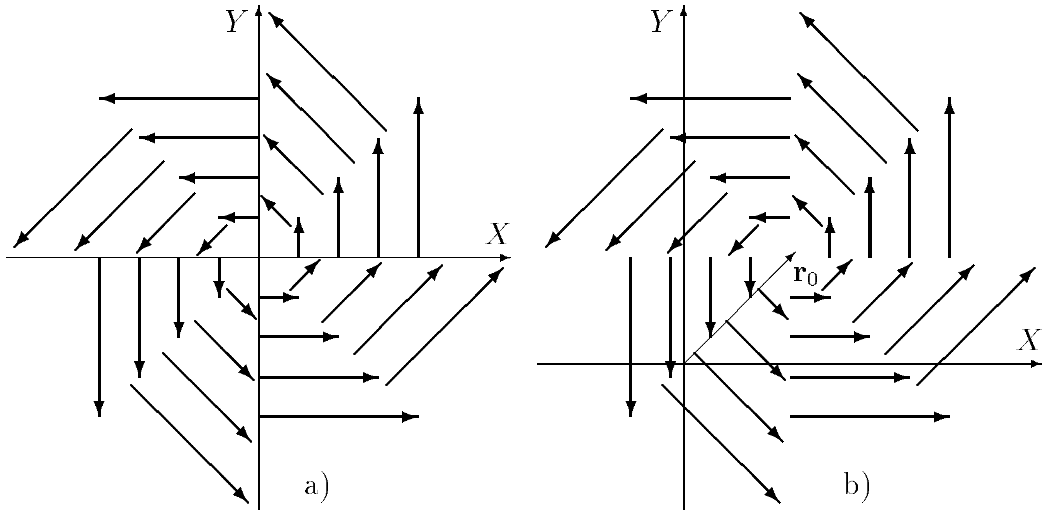
Rysunek 7.20

7.4 Pola wektorowe

Polem wektorowym nazywamy funkcję $\mathbf{F}$, której dziedzina i przeciwdziedzina mają ten sam wymiar. A dokładniej dla $U \in \mathbb{R}^n$ mamy $\mathbf{F} : U \to \mathbb{R}^n$. Mówimy, że pole wektorowe każdemu punktowi $\mathbf{x} \in U$ przypisuje wektor $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^n$. Taką funkcję możemy zobrazować przypisując możliwie wielu punktom strzałki, zob. Rys. 7.21. Tym razem strzałka przedstawiająca wektor $\mathbf{F}(\mathbf{x})$ zaczepiana jest w punkcie $\mathbf{x}$ zamiast w początku układu, co jest normalnym zwyczajem dla wektorów. W tym kontekście funkcja $f : U \to \mathbb{R}$, która każdemu punktowi przypisuje liczbę rzeczywistą, nazywa się *polem skalarnym*. Fizycznym przykładem pola wektorowego jest natężenie pola elektrycznego, a pola skalarnego potencjał elektryczny.



Rysunek 7.21



Rysunek 7.22

PRZYKŁAD 1. Niech $\mathbf{V}(\mathbf{r})$ będzie polem prędkości punktów obracającej się bryły sztywnej, według wzoru

$$\mathbf{V}(x, y, z) = a(-y\mathbf{e}_1 + x\mathbf{e}_2) \tag{7.53}$$

z pewną stałą a o fizycznym wymiarze odwrotności czasu. Możemy sprawdzić, że w każdym punkcie $\mathbf{r} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$ jest ono prostopadłe do wektora wodzącego: $\mathbf{V} \cdot \mathbf{r} = 0$. Jego wartość rośnie liniowo wraz z wartością wektora $\mathbf{r}$. Jego ilustracją jest Rys. 7.22a wykonany dla płaszczyzny (X, Y). Pole $\mathbf{V}(\mathbf{r})$ nie zależy od zmiennej z, więc ma symetrię względem przeusunięć wzdłuż osi Z. To oznacza, że w każdej płaszczyźnie równoległej do płaszczyzny (X, Y) wygląda tak samo, jak na Rys. 7.22a. Jest ono polem zakręconym o osi obracania pokrywającej się z osią Z. Dziedziną U, w której obowiązuje wzór (7.53), jest obszar zajęty przez bryłę sztywną.

Pole $\mathbf{V}'(\mathbf{r}) = \mathbf{V}(\mathbf{r} - \mathbf{r}_0)$, czyli dane wzorem

$$\mathbf{V}'(x, y, z) = a\left[-(y - y_0)\mathbf{e}_1 + (x - x_0)\mathbf{e}_2\right]$$

jest podobnym polem wirowym, ale „zakręconym” wokół innej osi, mianowicie równoległej do osi Z i przechodzącej przez koniec wektora $\mathbf{r}_0$. Jego wygląd na płaszczyźnie (X, Y) ukazany jest na Rys. 7.22b. Pole $\mathbf{V}'$ można zapisać inaczej

$$\mathbf{V}'(x, y, z) = a(-y\mathbf{e}_1 + x\mathbf{e}_2) - a(-y_0\mathbf{e}_1 + x_0\mathbf{e}_2) = \mathbf{V}(x, y, z) - \mathbf{V}_0, \tag{7.54}$$

skąd widzimy, że różni się od pola $\mathbf{V}$ tylko o pewien wektor stały.

Rozważmy teraz pole wektorowe $\mathbf{F}$ klasy C^1 na $\mathbb{R}^3$:

$$\mathbf{F}(x, y, z) = F_1(x, y, z)\,\mathbf{e}_1 + F_2(x, y, z)\,\mathbf{e}_2 + F_3(x, y, z)\,\mathbf{e}_3.$$

Nowe pole wektorowe

$$\operatorname{rot}\mathbf{F} = \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}\right)\mathbf{e}_1 + \left(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}\right)\mathbf{e}_2 + \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right)\mathbf{e}_3 \tag{7.55}$$

nazywamy *rotacją pola* $\mathbf{F}$. Ten wzór jest łatwiejszy do zapamiętania, jeśli go przepisać z oznaczeniem operatorowym. Wprowadźmy *operator wektorowy nabla*:

$$\boldsymbol{\nabla} \;=\; \mathbf{e}_1\frac{\partial}{\partial x}+\mathbf{e}_2\frac{\partial}{\partial y}+\mathbf{e}_3\frac{\partial}{\partial z} \;=\; \mathbf{e}_1\frac{\partial}{\partial x_1}+\mathbf{e}_2\frac{\partial}{\partial x_2}+\mathbf{e}_3\frac{\partial}{\partial x_3}.$$

Słowo 'operator' oznacza, że on działa na funkcje o wartościach skalarnych lub wektorowych, aby je zróżniczkować. Jeśli spojrzymy na niego jak na wektor o współrzędnych $(\partial_x,\, \partial_y,\, \partial_z) = (\partial_1,\, \partial_2,\, \partial_3)$, to możemy rotację zapisać jako iloczyn wektorowy nabli i wektora $\mathbf{F}$ z użyciem symbolu wyznacznika

$$\boldsymbol{\nabla}\times\mathbf{F} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ \partial_1 & \partial_2 & \partial_3 \\ F_1 & F_2 & F_3 \end{vmatrix} = (\partial_2F_3-\partial_3F_2)\mathbf{e}_1+(\partial_3F_1-\partial_1F_3)\mathbf{e}_2+(\partial_1F_2-\partial_2F_1)\mathbf{e}_3. \quad (7.56)$$

W ten sposób

$$\mathrm{rot}\,\mathbf{F} = \boldsymbol{\nabla}\times\mathbf{F}.$$

Zauważmy, że pole rot$\mathbf{F}$ jest klasy C^{k-1}, jeśli pole $\mathbf{F}$ jest klasy C^k.

Właściwie nabla występująca w zapisie gradientu też jest operatorem wektorowym, bo działając na pole skalarne f daje pole wektorowe $\boldsymbol{\nabla} f$ – odpowiada to mnożeniu wektora przez skalar, co daje wektor.

Twierdzenie 1. Niech $U \subset \mathbb{R}^3$. Dla dowolnej funkcji $f : U \to \mathbb{R}$ klasy C^2 zachodzi

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla} f) = \mathbf{0},$$

tzn. rotacja gradientu jest wektorem zerowym.

Dowód. Do wzoru (7.55) należy podstawić $F_i = \partial f/\partial x_i$:

$$\mathrm{rot(grad)}f = \left(\frac{\partial^2 f}{\partial x_2\partial x_3}-\frac{\partial^2 f}{\partial x_3\partial x_2}\right)\mathbf{e}_1+\left(\frac{\partial^2 f}{\partial x_3\partial x_1}-\frac{\partial^2 f}{\partial x_1\partial x_3}\right)\mathbf{e}_2+\left(\frac{\partial^2 f}{\partial x_1\partial x_2}-\frac{\partial^2 f}{\partial x_2\partial x_1}\right)\mathbf{e}_3.$$

Skoro drugie pochodne są ciągłe, to na mocy Tw. 1 z § 8.2 pochodne mieszane są przemienne, więc wyrażenia w nawiasach są zerami. Otrzymujemy wektor o zerowych współrzędnych, więc wektor zerowy. ■

PRZYKŁAD 1a. Wyliczmy rotację pola $\mathbf{V}$ z Przykładu 1.

$$\boldsymbol{\nabla}\times\mathbf{V} = (-\partial_3V_2)\mathbf{e}_1+(\partial_3V_1)\mathbf{e}_2+(\partial_1V_2-\partial_2V_1)\mathbf{e}_3 = (\partial_1V_2-\partial_2V_1)\mathbf{e}_3$$

$$= a(1+1)\mathbf{e}_3 = 2a\,\mathbf{e}_3.$$

Zauważamy, że rotacja dała pole jednorodne, tzn. niezależne od położenia, Jego kierunek jest równoległy do osi Z, czyli do osi obrotu płyty. To jest cecha charakterystyczna wprowadzonego odwzorowania – jest ono niezerowe wtedy, gdy pole wektorowe ma coś wspólnego z obracaniem, a jego kierunek jest równoległy do osi obrotu. Zresztą stąd jest nazwa tego operatora: rotacja.

Można zapytać, dlaczego rotacja tego pola nie wyróżnia jakoś osi Z, skoro to jest oś obrotu. Odpowiedzią na to jest rotacja pola $\mathbf{V}'$, która jest taka sama, jak rotacja pola $\mathbf{V}$, gdyż według wzoru (7.54) oba te pola różnią się o wektor stały, więc wszelkie ich pochodne muszą być jednakowe. Stąd

$$\nabla \times \mathbf{V}' = 2a\,\mathbf{e}_3.$$

A osią obrotu dla pola $\mathbf{V}'$ jest prosta przechodząca przez koniec wektora $\mathbf{r}_0$. stały

PRZYKŁAD 2. Rozważmy pole wektorowe $\mathbf{v}$ w walcu o osi pokrywającej się z osią Z, ale z osią Z usuniętą z dziedziny:

$$\mathbf{v}(x,y,z) = A\,\frac{-y\mathbf{e}_1 + x\mathbf{e}_2}{\rho^2} = A\left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\right), \tag{7.57}$$

gdzie $\rho^2 = x^2 + y^2$. Zauważmy taki związek $\mathbf{v} = \frac{A}{a\rho^2}\mathbf{V}$ z polem z Przykładu 1. Jako równoległe do pola $\mathbf{V}$ to pole jest znowu styczne do okręgów o środkach na osi Z, ale jego wartość maleje wraz z oddalaniem się od osi, więc rysunek 7.22a już nie jest dla niego odpowiedni.

Rozważmy ciecz w jakimś przepływie ustalonym. Jeśli każdemu punktowi przypiszemy prędkość cieczy w tym punkcie, to otrzymamy pole prędkości $\mathbf{v}(\mathbf{r})$ cieczy. W takiej sytuacji warunek $\operatorname{rot}\mathbf{v} = 0$ oznacza fizycznie, że ciecz jest wolna od wirów, albo jest *bezwirowa*. Uzasadnienie tego poglądu i użycie słowa „bezwirowa” zależy od twierdzenia Stokesa, które poznamy w następnym rozdziale. Pole (7.57) przybliża pole prędkości wody w ruchu okrężnym, jaki występuje w wiadrze wody z dziurą wywierconą w środku jego dna, czyli wtedy, gdy pojawia się wir. Wówczas prędkość jest coraz większa w miarę zbliżania się do osi wiru. Dlaczego jest napisane „przybliża”? Otóż dlatego, że w takim wiadrze prędkość nie jest idealnie styczna do okręgów o środkach na osi wiadra, lecz musi mieć niewielką składową w stronę osi.

Wyliczmy rotację tego pola:

$$\nabla \times \mathbf{v} = (\partial_1 v_2 - \partial_2 v_1)\,\mathbf{e}_3 = \left(\frac{x^2+y^2-x\cdot 2x}{(x^2+y^2)^2} - \frac{-(x^2+y^2)+y\cdot 2y}{(x^2+y^2)^2}\right)\mathbf{e}_3$$

$$= \frac{-x^2+y^2+x^2-y^2}{(x^2+y^2)^2}\,\mathbf{e}_3 = \mathbf{0}.$$

Widzimy, że rotacja jest zerem w całej dziedzinie, czyli poza osią Z. Ten wynik oznacza, że nie ma innych wirów poza osią jednego wiru.

Innym operatorem działającym na pola wektorowe w $\mathbb{R}^3$ jest *dywergencja* określona jako

$$\operatorname{div}\mathbf{F} = \nabla\cdot\mathbf{F} = \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{\partial F_3}{\partial x_3}. \tag{7.58}$$

Jest to iloczyn skalarny nabli i wektora $\mathbf{F}$. Zauważmy, że dywergencja jest polem skalarnym, więc w pewnym sensie jest czymś odwrotnym do gradientu. Następny przykład pomoże podać znaczenie fizyczne tego operatora.

PRZYKŁAD 3. Wyobraźmy sobie, że $\mathbf{F}$ jest polem prędkości gazu, danym przez

$$\mathbf{F}(\mathbf{r}) = a(x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3) = a\,\mathbf{r}.$$

Ten wzór oznacza, że gaz rozbiega się tym szybciej, im dalej jest od początku układu. Tak zachowują się galaktyki w naszym wszechświecie, a współczynnik a nazywa się *stałą Hubble'a*. Dla takiego pola $\mathbf{F}$:

$$\operatorname{div}\mathbf{F} = a(1+1+1) = 3a.$$

To oznacza, że gaz powiększa trzykrotnie swoją objętość na jednostkę czasu. Słowo dywergencja wzięło się z łacińskiego *divergare* = rozbiegać się. Zresztą jeszcze przed ostatnią wojną polscy matematycy zamiast „dywergencja" mówili „rozbieżność".

Pole

$$\mathbf{G}(\mathbf{r}) = \mathbf{F}(\mathbf{r}-\mathbf{r}_0) = a\,(\mathbf{r}-\mathbf{r}_0)$$

ma środek rozbiegania się w punkcie $\mathbf{r}_0$ zamiast w początku układu. Jednakże można je zapisać inaczej

$$\mathbf{G}(\mathbf{r}) = a\mathbf{r} - a\mathbf{r}_0 = \mathbf{F}(\mathbf{r}) - \mathbf{F}_0,$$

czyli różni się od pola $\mathbf{F}$ tylko o wektor stały, więc ma te same pochodne, czyli tę samą dywergencję, co pole $\mathbf{F}$:

$$\operatorname{div}\mathbf{G} = 3a.$$

Nic dziwnego więc, że pole $\mathbf{F}$ ma dywergencję jednakową w każdym punkcie.

Twierdzenie 2. Niech $U \subset \mathbb{R}^3$. Dla dowolnej funkcji $\mathbf{G} : U \to \mathbb{R}^3$ klasy C^2 zachodzi

$$\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\mathbf{G}) = \mathbf{0},$$

tzn. dywergencja rotacji jest zerem.

Dowód. Do wzoru (7.58) należy podstawić $F_1 = \partial_2 G_3 - \partial_3 G_2$ i podobnie dla F_2, F_3 z cykliczną zamianą wskaźników:

$$\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\mathbf{G}) = \frac{\partial}{\partial x_1}\left(\frac{\partial G_3}{\partial x_2} - \frac{\partial G_2}{\partial x_3}\right) + \frac{\partial}{\partial x_2}\left(\frac{\partial G_1}{\partial x_3} - \frac{\partial G_3}{\partial x_1}\right) + \frac{\partial}{\partial x_3}\left(\frac{\partial G_2}{\partial x_1} - \frac{\partial G_1}{\partial x_2}\right)$$

$$= \frac{\partial^2 G_3}{\partial x_1 \partial x_2} - \frac{\partial^2 G_2}{\partial x_1 \partial x_3} + \frac{\partial^2 G_1}{\partial x_2 \partial x_3} - \frac{\partial^2 G_3}{\partial x_2 \partial x_1} + \frac{\partial^2 G_2}{\partial x_3 \partial x_1} - \frac{\partial^2 G_1}{\partial x_3 \partial x_2} = 0,$$

gdyż pochodne mieszane są przemienne i upraszczają się parami. ■

PRZYKŁAD 3a. Na podstawie Twierdzenia 2 wnioskujemy, że pole $\mathbf{F}$ z Przykładu 3 nie może być rotacją innego pola wektorowego, bo wtedy miałoby zerową dywergencję.

Wprowadza się jeszcze *operator Laplace'a* zwany też *laplasjanem*:

$$\Delta = \frac{\partial^2}{\partial {x_1}^2} + \frac{\partial^2}{\partial {x_2}^2} + \frac{\partial^2}{\partial {x_3}^2},$$

który może działać zarówno na pola skalarne, jak i wektorowe, bo sam jest operatorem skalarnym. Natomiast w działaniu na pola skalarne można go zapisać jeszcze inaczej

$$\Delta f = \boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} f) = \frac{\partial^2 f}{\partial {x_1}^2} + \frac{\partial^2 f}{\partial {x_2}^2} + \frac{\partial^2 f}{\partial {x_3}^2}.$$

[Pierre Simon Laplace (1749-1827) – francuski matematyk, fizyk i astronom. W wieku 26 lat miał stanowisko profesora w szkole artyleryjskiej w Paryżu. W okresie Rewolucji Francuskiej był ministrem spraw wewnętrznych. W astronomii zasłużył się dla mechaniki nieba, w matematyce dla teorii prawdopodobieństwa i równań różniczkowych cząstkowych, w fizyce dla teorii powszechnego ciążenia.]

Niech f będzie polem skalarnym, a $\mathbf{F}, \mathbf{G}$ – polami wektorowymi na $U \subset \mathbb{R}^3$. Poznane przez nas operatory różniczkowe zbudowane z nabli i laplasjanu spełniają następujące tożsamości

$$\operatorname{div}(f\mathbf{F}) = f \operatorname{div}\mathbf{F} + \mathbf{F} \cdot \operatorname{grad} f,$$

$$\operatorname{div}(\mathbf{F} \times \mathbf{G}) = \mathbf{G} \cdot \operatorname{rot} \mathbf{F} - \mathbf{F} \cdot \operatorname{rot} \mathbf{G}, \tag{7.59}$$

$$\operatorname{rot}(f\mathbf{F}) = f \operatorname{rot} \mathbf{F} + (\operatorname{grad} f) \times \mathbf{F},$$

$$\operatorname{rot}(\mathbf{F} \times \mathbf{G}) = \mathbf{F} \operatorname{div} \mathbf{G} - \mathbf{G} \operatorname{div} \mathbf{F} + (\mathbf{G} \cdot \boldsymbol{\nabla})\mathbf{F} - (\mathbf{F} \cdot \boldsymbol{\nabla})\mathbf{G}. \tag{7.60}$$

$$\operatorname{rot}(\operatorname{rot} \mathbf{F}) = \operatorname{grad}(\operatorname{div} \mathbf{F}) - \Delta\mathbf{F},$$

Udowodnienie większości z nich przenosimy na ćwiczenia, a teraz udowodnimy tożsamość (7.59). Lewa strona wynosi

$$L = \partial_1(\mathbf{F} \times \mathbf{G})_1 + \partial_2(\mathbf{F} \times \mathbf{G})_2 + \partial_2(\mathbf{F} \times \mathbf{G})_2$$
$$= \partial_1(F_2G_3 - F_3G_2) + \partial_2(F_3G_1 - F_1G_3) + \partial_3(F_1G_2 - F_2G_1).$$

Prawa strona wynosi

$$P = G_1(\boldsymbol{\nabla} \times \mathbf{F})_1 + G_2(\boldsymbol{\nabla} \times \mathbf{F})_2 + G_3(\boldsymbol{\nabla} \times \mathbf{F})_3 - F_1(\boldsymbol{\nabla} \times \mathbf{G})_1 - F_2(\boldsymbol{\nabla} \times \mathbf{G})_2 - F_3(\boldsymbol{\nabla} \times \mathbf{G})_3$$
$$= G_1(\partial_2F_3 - \partial_3F_2) + G_2(\partial_3F_1 - \partial_1F_3) + G_3(\partial_1F_2 - \partial_2F_1)$$
$$-F_1(\partial_2G_3 - \partial_3G_2) - F_2(\partial_3G_1 - \partial_1G_3) - F_3(\partial_1G_2 - \partial_2G_1)$$
$$= G_1\partial_2F_3 - G_1\partial_3F_2 + G_2\partial_3F_1 - G_2\partial_1F_3 + G_3\partial_1F_2 - G_3\partial_2F_1$$
$$-F_1\partial_2G_3 + F_1\partial_3G_2 - F_2\partial_3G_1 + F_2\partial_1G_3) - F_3\partial_1G_2 + F_3\partial_2G_1.$$

Łączymy wyrazy z pochodnymi względem tych samych zmiennych:

$$P = (\partial_1F_2)G_3 + F_2(\partial_1G_3) - (\partial_1F_3)G_2 - F_3(\partial_1G_2)$$
$$+(\partial_2F_3)G_1 + F_3(\partial_2G_1) - (\partial_2F_1)G_3 - F_1(\partial_2G_3)$$
$$+(\partial_3F_1)G_2 + F_1(\partial_3G_2) - (\partial_3F_2)G_1 - F_2(\partial_3G_1).$$

Korzystamy z wzoru na pochodną iloczynu funkcyj

$$P = \partial_1(F_2G_3) - \partial_1(F_3G_2) + \partial_2(F_3G_1) - \partial_2(F_1G_3) + \partial_3(F_1G_2) - \partial_3(F_2G_1)$$

$$= \partial_1(F_2G_3 - F_3G_2) + \partial_2(F_3G_1 - F_1G_3) + \partial_3(F_1G_2 - F_2G_1).$$

Zauważamy, że $L = P$.

Udowodnimy jeszcze tożsamość (7.60). Obie strony są wektorami. Wyliczmy piewrwszą współrzędną wektora po lewej stronie:

$$L_1 = \partial_2(\mathbf{F}\times\mathbf{G})_3 - \partial_3(\mathbf{F}\times\mathbf{G})_2 = \partial_2(F_1G_2 - F_2G_1) - \partial(F_3G_1 - F_1G_3).$$

Wykorzystujemy wzór Lebniza na pochodną iloczynu funkcyj

$$L_1 = G_2\partial_2F_1 + F_1\partial_2G_2 - G_1\partial_2F_2 - F_2\partial_1G_1 - G_1\partial_3F_3 - F_3\partial_3G_1 + G_3\partial_3F_1 + F_1\partial_3G_3$$

$$= F_1\,(\partial_2G_2 + \partial_3G_3) - G_1\,(\partial_2F_2 + \partial_3F_3) + (G_2\partial_2 + G_3\partial_3)F_1 - (F_2\partial_2 + F_3\partial_3)G_1.$$

Liczymy pierwszą współrzędną prawej strony:

$$P_1 = F_1\,\mathrm{div}\mathbf{G} - G_1\,\mathrm{div}\mathbf{F} + (\mathbf{G}\cdot\boldsymbol{\nabla})F_1 - (\mathbf{F}\cdot\boldsymbol{\nabla})G_1 = F_1(\partial_1G_1 + \partial_2G_2 + \partial_3G_3)$$

$$-G_1(\partial_1F_1 + \partial_2F_2 + \partial_3F_3) + (G_1\partial_1 + G_2\partial_2 + G_3\partial_3)F_1 - (F_1\partial_1 + F_2\partial_2 + F_3\partial_3)G_1.$$

Wyrazy $F_1\partial_1G_1$ oraz $G_1\partial_1F_1$ występują po dwa razy z przeciwnymi znakami, więc się upraszczają i zostaje

$$P_1 = F_1\,(\partial_2G_2 + \partial_3G_3) - G_1\,(\partial_2F_2 + \partial_3F_3) + (G_2\partial_2 + G_3\partial_3)F_1 - (F_2\partial_2 + F_3\partial_3)G_1.$$

Zauważamy, że $L_1 = P_1$. Podobnie sprawdza się tożsamość dla pozostałych współrzędnych.