

9 Podstawowe twierdzenia o różniczkowalności

9.1 Funkcje klasy C^1

Przypomnijmy, że:

Odwzorowanie $f : D \subset \mathbb{R}^d \rightarrow \mathbb{R}^k$ jest różniczkowalne w punkcie P_0 gdy istnieje odwzorowanie liniowe $L : \mathbb{R}^d \rightarrow \mathbb{R}^k$ oraz funkcja α określona w otoczeniu zera taka, że

$$\Delta f := f(P_0 + \Delta P) - f(P_0) = L(\Delta P) + \alpha(\Delta P) \|\Delta P\|, \quad \text{gdzie} \quad \lim_{\Delta P \rightarrow \vec{0}} \alpha(\Delta P) = 0.$$

Takie odwzorowanie liniowe L nazywamy różniczką (lub różniczką zupełną) odwzorowania f w punkcie P_0 i oznaczamy je symbolem $d_{P_0} f$

Wyrażenie postaci $\alpha(\Delta P) \cdot \|\Delta P\|$, w którym $\alpha(\Delta P)$ zmierza do zera nazywamy wyrażeniem typu o-małe od normy wektora "przyrostu argumentu": ΔP , zapisując je jako $o(\|\Delta P\|)$ przy $\Delta P \rightarrow 0$. Różniczkowalność oznacza więc istnienie takiego odwzorowania liniowego L , które przybliża przyrost całkowity f z dokładnością do o-małego od normy przyrostu argumentu. Przy badaniu różniczkowalności można to robić osobno dla każdej ze składowych odwzorowania f sprowadzając zagadnienie do przypadku f o wartościach skalarnych ($k=1$).

Mówimy, że odwzorowanie $f : D \rightarrow \mathbb{R}$ jest klasy C^1 w zbiorze otwartym D , co zapisujemy: $f \in C^1(D)$, gdy wszystkie pochodne cząstkowe $\frac{\partial f}{\partial x_j}(P)$ są funkcjami ciągłymi zmiennej $P \in D$.

Nieco słabszym warunkiem jest założenie przyjęte w następującym twierdzeniu:

Twierdzenie. Jeśli założymy istnienie pochodnych cząstkowych dla f w pewnym otoczeniu P i ich ciągłość w punkcie P , to f będzie różniczkowalna w punkcie P , zaś wartość różniczki $d_P f$ na wektorze $V = (v_1, \dots, v_d) \in \mathbb{R}^d$ będzie iloczynem skalarnym tego wektora V przez wektor gradientu f w tym punkcie. Czyli

$$d_P f(V) = \sum_{j=1}^d \frac{\partial f}{\partial x_j}(P) v_j. \quad (1)$$

Można ponadto wykazać, że suma, iloczyn funkcji różniczkowalnych - są różniczkowalne. W przypadku ilorazów- trzeba dodatkowo założyć, że mianownik jest funkcją ciągłą nieprzyjmującą wartości zero.

9.2 Różniczka złożenia

Złożenie odwzorowań różniczkowalnych jest różniczkowalne, a różniczką złożenia jest złożenie różniczek poszczególnych odwzorowań. Mówi o tym twierdzenie, często nazywane Regułą Łańcucha.:

Twierdzenie (Reguła łańcucha). Jeżeli odwzorowanie $f : D \subset \mathbb{R}^d \rightarrow \mathbb{R}^k$ jest różniczkowalne w punkcie wewnętrznym x zbioru D , $y = f(x)$ oraz $g : D_1 \subset \mathbb{R}^k \rightarrow \mathbb{R}^m$ określone w otoczeniu D_1 punktu y jest w tym punkcie różniczkowalne, to złożenie $g \circ f$ jest różniczkowalne w punkcie x , a jego różniczka jest złożeniem odpowiednich różniczek:

$$d_x(g \circ f) = d_{f(x)} g \circ d_x f. \quad (2)$$

Różniczkę -jako odwzorowanie liniowe- możemy utożsamiać z macierzą pochodnych cząstkowych, zwaną macierzą Jacobiego. Wiemy, jak wykonać mnożenie macierzy- to odpowiednik składania odwzorowań. Na przykład, dla $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ macierz Jacobiego, to wiersz złożony z pochodnych cząstkowych f , czyli gradient -a dokładnie -macierz transponowana do wektora gradientu.

Bowiem wektory traktujemy jako macierze kolumnowe. Iloczyn macierzy wierszowej i kolumnowej - to iloczyn skalarny związanych z nimi wektorów. Reguła Łańcucha pozwala więc na wyliczenie pochodnej ze złożenia

$$\phi(t) = f(x(t), y(t), z(t)).$$

Różniczka f w punkcie $P_0 = \gamma(t_0)$, to odwzorowanie liniowe:

$$\mathbb{R}^3 \ni (x, y, z) \rightarrow ax + by + cz \in \mathbb{R},$$

gdzie np. $a = \frac{\partial f}{\partial x}(P_0), \dots, c = \frac{\partial f}{\partial z}(P_0)$. Odwzorowaniem wewnętrznym tego złożenia, określającego funkcję ϕ , jest zestawienie trzech funkcji: $\mathbb{R} \ni t \rightarrow \gamma(t) := (x(t), y(t), z(t)) \in \mathbb{R}^3$. Różniczka tego odwzorowania liczona w punkcie t_0 działa jako mnożenie liczby rzeczywistej t przez wektor pochodnych: $\gamma'(t_0) = (x'(t_0), y'(t_0), z'(t_0))$ - czyli przez wektor styczny do krzywej γ . Złożenie tych dwu odwzorowań liniowych przekształca zmienną $t \in \mathbb{R}$ na liczbę t pomnożoną przez $ax'(t_0) + by'(t_0) + cz'(t_0)$ - czyli przez iloczyn skalarny gradientu f (liczonego w punkcie $\gamma(t_0)$). Pochodna takiego odwzorowania - to wynik podstawienia wartości $t = 1$ do różniczki tego odwzorowania, czyli jest to ten czynnik, przez który mnożymy t licząc różniczkę $d_{t_0}\phi(t)$. Mamy więc wyprowadzony dzięki Regule Łańcucha wzór na pochodną:

$$\frac{d}{dt}f(x(t), y(t), z(t))|_{t=t_0} = \frac{\partial f}{\partial x}(\gamma(t_0))x'(t_0) + \frac{\partial f}{\partial y}(\gamma(t_0))y'(t_0) + \frac{\partial f}{\partial z}(\gamma(t_0))z'(t_0) \quad (3)$$

Analogiczny, nieco krótszy wzór otrzymamy w przypadku dwu zmiennych (pomijamy wtedy ostatni składnik po prawej stronie równości (3)). Przykład (bardzo prosty) uzyskamy biorąc $f(x, y) = xy$. W punkcie $P_0 = (x_0, y_0)$ mamy

$$\frac{\partial f}{\partial x}(P_0) = y_0, \quad \frac{\partial f}{\partial y}(P_0) = x_0.$$

Dla zestawienia $\gamma(t) = (x(t), y(t))$ dwu funkcji różniczkowalnych $x(\cdot), y(\cdot)$ zmiennej $t \in \mathbb{R}$ mamy złożenie $f \circ \gamma$ równe po prostu iloczynowi tych dwu funkcji: $\phi(t) = f(x(t), y(t)) = x(t)y(t)$. Liczenie $\phi'(t)$ z reguły łańcucha daje dokładnie tę samą wartość, co wzór Leibniza na pochodną iloczynu: $(xy)'(t) = x'(t)y(t) + x(t)y'(t)$. Podobne zastosowanie funkcji $g(x, y) = \frac{x}{y}$ da nam nowy dowód wzoru na pochodną ilorazu dwu funkcji:

$$\left(\frac{x}{y}\right)'(t) = \frac{x'(t)y(t) - x(t)y'(t)}{y^2(t)}.$$

Proszę policzyć pochodne cząstkowe z tej funkcji g w punkcie $P_0 = (x_0, y_0)$ i sprawdzić, jak działa w tym przypadku nasze zastosowanie reguły Łańcucha.

Najprostszą postacią ("niezakrzywioną") krzywej jest równanie odcinka łączącego punkty P_0, P_1 przestrzeni $\mathbb{R}^d$. Oznaczmy przez $[P_0 : P_1]$ odcinek w przestrzeni $\mathbb{R}^d$ o końcach w punktach P_0, P_1 . Naturalną parametryzacją dla tego odcinka jest odwzorowanie afiniczne $[0, 1] \ni t \rightarrow P_t := P_0 + t(P_1 - P_0)$. Punkty postaci P_t , gdzie $0 < t < 1$ będziemy nazywali punktami wewnętrznymi tego odcinka. (Odwzorowanie to określone jest, oczywiście, nawet dla dowolnych $t \in \mathbb{R}$.) Różniczką jest tu część liniowa tego odwzorowania, czyli odwzorowanie $\mathbb{R} \ni t \rightarrow t(P_1 - P_0)$. Jego wartość w punkcie $t = 1$, to pochodna tego odwzorowania, jest nią (w każdym punkcie) wektor (stały) $P_1 - P_0$. Stosując więc regułę łańcucha do funkcji złożonej $t \mapsto f(P_t)$, dzięki Twierdzeniu Lagrange'a otrzymujemy jego wersję dla funkcji wielu miennych:

Twierdzenie (o wartości średniej). Jeżeli funkcja $f : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ jest różniczkowalna w punktach wewnętrznych odcinka $[P_0 : P_1]$ i ciągła na tym odcinku, to dla pewnego $\theta \in (0, 1)$ jej przyrost wyraża się wzorem:

$$f(P_1) - f(P_0) = d_{P_\theta}f(P_1 - P_0).$$

Gdy więc wektor $P_1 - P_0$ jest wektorem $h = (h_1, h_2, \dots, h_d)$, to przyrost ten możemy wyrazić przy użyciu pochodnych cząstkowych:

$$f(P_1) - f(P_0) = \sum_{j=1}^d \frac{\partial f}{\partial x_j}(P_0) h_j.$$

Podobnie, jak w przypadku odwzorowań $\gamma : [a, b] \rightarrow \mathbb{R}^k$, analogiczna teza przestaje być prawdziwa dla odwzorowań o wartościach wektorowych (dla $k > 1$). Mamy wówczas tylko możliwość oszacowania normy przyrostu. Dowód (polegający na pomnożeniu skalarnym F przez wektor w będący wektorem przyrostu: $F(P_1) - F(P_0)$ podzielonym przez jego normę) -pomińmy.

Twierdzenie (o przyrostach). Jeżeli odwzorowanie $F : D \subset \mathbb{R}^d \rightarrow \mathbb{R}^k$ jest różniczkowalne w punktach wewnętrznych odcinka $[P_0 : P_1]$ i ciągle na tym odcinku, to

$$\|F(P_1) - F(P_0)\| \leq \|P_1 - P_0\| \sup_{t \in (0,1)} \|d_{P_t}\|.$$

Tu norma z różniczki, to norma odwzorowania liniowego. Przypomnijmy, że dla $T : \mathbb{R}^d \rightarrow \mathbb{R}^k$ taka norma, $\|T\|$, to najmniejsza stała M , dla której na każdym z wektorów $v \in \mathbb{R}^d$ zachodzą nierówności: $\|T(v)\| \leq M\|v\|$. Ponadto $\|T\| = \sup\{\|T(w)\| : \|w\| = 1\}$. W przypadku macierzy symetrycznych $d \times d$ taka norma (zwana normą spektralną), to maksimum z modułów jej wartości własnych. Zbiór wartości własnych -to widmo (spectrum) macierzy.

9.3 Pochodna kierunkowa, gradient

Innym zastosowaniem reguły łańcucha jest wzór dla tzw. pochodnej kierunkowej. W niektórych źródłach przyjmuje się jej definicję "obustronną" =a w innych "jednostronną" -która jest nieco bardziej "elastyczna" i z niej skorzystamy.

Definicja. Jeżeli $w \in \mathbb{R}^d$ jest wektorem niezerowym, $f : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ jest pewną funkcją oraz $P_0 \in D$ jest takim punktem, że zbiór $\{t \in \mathbb{R} : P_0 + tw \in D\}$ jest prawostronnym otoczeniem zera, to przez pochodną $\partial^w f(P_0)$ w kierunku wektora w w punkcie P_0 rozumiemy granicę

$$\partial^w f(P_0) := \lim_{h \rightarrow 0^+} \frac{f(P_0 + hw) - f(P_0)}{h}.$$

Ścisłej- można mówić o pochodnej kierunkowej jednostronnej- stronę wyznacza to zwrot wektora, bo kierunek - to cała prosta $\{P_0 + tw : t \in \mathbb{R}\}$. W przypadku $P_0 = 0$ odwzorowanie $P \mapsto \|P\|$ ma w każdym kierunku pochodną 1-stronną równą 1, ale brak tu granicy obustronnej. Jest to więc pochodna 1-stronna w punkcie $t = 0$ z funkcji złożonej $f(P_0 + tw)$ i dzięki regule łańcucha (pochodną odzyskujemy z różniczki licząc jej wartość na 1-wymiarowym wektorze 1), mamy następujący

Wniosek. Funkcja różniczkowalna w danym punkcie ma również pochodne kierunkowe w kierunku każdego wektora $w \in \mathbb{R}^d$, przy czym

$$\partial^w f(P_0) = d_{P_0} f(w) = \langle \nabla f(P_0), w \rangle.$$

W ostatnim wzorze użyłem symbolu gradientu ∇f , gdzie

$$\nabla f(P_0) = (f'_x(P_0), f'_y(P_0), f'_z(P_0)).$$

Ponadto iloczyn skalarny wektorów u, w oznaczam tu symbolem $\langle u, w \rangle$ i z geometrii analitycznej wiemy, że jest to liczba równa $\|u\|\|w\| \cos \angle(u, w)$. Analogicznie jest dla funkcji zmiennej wektorowej $\mathbf{x} = (x_1, \dots, x_d)$. Różniczka funkcji o wartościach skalarnych jest odwzorowaniem liniowym, a każde odwzorowanie $T \in \mathcal{L}(\mathbb{R}^d, \mathbb{R})$ (czyli funkcjonal liniowy) jest mnożeniem skalarnym przez pewien wektor -w tym przypadku przez gradient.

W szczególności, gdy pochodne kierunkowe istnieją, ale nie zależą w sposób liniowy od wektora kierunku, to dana funkcja nie jest różniczkowalna. Natomiast maksymalną wartość (spośród wektorów w o ustalonej normie) taka pochodna kierunkowa $\partial^w f(P_0)$ osiąga w kierunku wektora $\nabla f(P_0)$, wartość zero - osiąga ona w kierunkach prostopadłych do gradientu f w danym punkcie.

9.4 Płaszczyzna styczna

W przypadku $d = 2$ wektory prostopadłe do wektora normalnego, zaczepione (=mające początek) w punkcie $(P_0, f(P_0)) \in \mathbb{R}^3$, tworzą płaszczyznę styczną do wykresu f . Dokładniej, wykres funkcji dwu zmiennych $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, to w tej sytuacji zbiór

$$\Gamma(f) = \{(x, y, f(x, y)) \in \mathbb{R}^3 : (x, y) \in D\}$$

-nazywany też płatem powierzchniowym wyznaczonym przez f . Na przykład, dla D będącego kołem jednostkowym na płaszczyźnie, gdy

$$f(x, y) = \sqrt{1 - (x^2 + y^2)},$$

to wykresem f jest górna połowa sfery jednostkowej w $\mathbb{R}^3$.

Nie wchodząc w szczegóły definicji płaszczyzny stycznej, sformułujmy jedynie następujący jej opis:

Twierdzenie. Wykres odwzorowania $f : D \rightarrow \mathbb{R}$ określonego w otoczeniu D punktu P_0 ma płaszczyznę styczną w tym punkcie wtedy i tylko wtedy, gdy f jest różniczkowalna. W takim przypadku, równanie płaszczyzny stycznej ma postać

$$z - f(P_0) = (x - x_0) \frac{\partial f}{\partial x}(P_0) + (y - y_0) \frac{\partial f}{\partial y}(P_0) \quad (4)$$

Ogólniej, dla $F : \mathbb{R}^3 \rightarrow \mathbb{R}$ różniczkowalnej, płat o równaniu $F(x, y, z) = 0$ ma płaszczyznę styczną w punkcie $Q_0 = (x_0, y_0, z_0)$ leżącym na tym płacie, jest to płaszczyzna prostopadła do gradientu F w tym punkcie. Innymi słowy, płaszczyzna styczna ma wtedy równanie

$$(x - x_0) \frac{\partial F}{\partial x}(Q_0) + (y - y_0) \frac{\partial F}{\partial y}(Q_0) + (z - z_0) \frac{\partial F}{\partial z}(Q_0) = 0. \quad (5)$$

Ogólne równanie płaszczyzny w $\mathbb{R}^3$ przechodzącej przez punkt (x_0, y_0, z_0) ma postać

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0.$$

Jeśli Q oznacza punkt o współrzędnych (x, y, z) , zaś $Q_0 = (x_0, y_0, z_0)$, to $(x - x_0, y - y_0, z - z_0) = Q - Q_0$ jest wektorem o początku Q_0 i końcu Q , a równanie opisuje zbiór $\{Q \in \mathbb{R}^3 : Q - Q_0 \perp \vec{n}\}$, gdzie $\vec{n}$ jest tzw. **wektorem normalnym** o współrzędnych (A, B, C) . Widzieliśmy to przy opisie płaszczyzny stycznej do wykresu funkcji $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$. W przypadku f różniczkowalnej w punkcie $P_0 = (x_0, y_0)$ w równaniu płaszczyzny stycznej mieliśmy $z_0 = f(x_0, y_0)$, $A = \frac{\partial f}{\partial x}(P_0)$, $B = \frac{\partial f}{\partial y}(P_0)$, zaś $C = -1$. W przypadku funkcji różniczkowalnej $F : D \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ w punkcie $Q_0 = (x_0, y_0, z_0)$ mogliśmy opisać płaszczyznę styczną do powierzchni

$$S = \{(x, y, z) \in D : F(x, y, z) = 0\} \quad (6)$$

i wówczas $(A, B, C) = \nabla F(Q_0)$, czyli wektor normalny jest gradientem F . Na przykład, jeśli $F(x, y, z) = f(x, y) - z$, to równanie $F(x, y, z) = 0$ opisuje dokładnie wykres f i faktycznie, $\nabla F(x_0, y_0, z_0) = (\frac{\partial f}{\partial x}(P_0), \frac{\partial f}{\partial y}(P_0), -1)$.