9 Różniczki i ich własności 5.maja 2021

Zapis bezargumentowy różniczki. Odwzorowania liniowe: $\mathbb{R}^2 \to \mathbb{R}$ przypisujące punktowi jego współrzędne na osiach $0X$, odp. $0Y$ oznaczamy symbolami dx, dy. Tak więc odwzorowanie L takie, że $L(h,k) = ah + bk$ możemy zapisać bezargumentowo: $L = a\,dx + b\,dy$. Różniczkę zupełną zapisujemy najczęściej w postaci

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy, \tag{1}$$

lub dokładniej,

$$d_{(x_0,y_0)}f = \frac{\partial}{\partial x}f(x_0,y_0)dx + \frac{\partial}{\partial y}f(x_0,y_0)dy.$$

Analogicznie jest w $\mathbb{R}^3$ („dochodzi dz"). W $\mathbb{R}^n$ mamy dx_j dla $j = 1,\ldots,n$.

Dla odwzorowań $F = (f_1,\ldots,f_k)$ jak wyżej, odwzorowanie L ma współrzędne $(L_1,\ldots,L_k)$ i norma z wektora różnicy $\Delta F - L(\mathbf{h})$ jest większa lub równa modułowi z jego dowolnej (i -tej) współrzędnej, czyli liczbie $|\Delta f_i - L_i(\mathbf{h})|$.
Z kolei, ta norma (euklidesowa) jest oszacowana przez sumę modułów i -tych współrzędnych względem $i = 1,\ldots,k$. Stąd wynika

Wniosek 1. *Różniczkowalność odwzorowania $F = (f_1,\ldots,f_k)$ o wartościach wektorowych jest równoważna różniczkowalności każdej z jego współrzędnych f_i $(i \leqslant k)$ z osobna.*

9.1 Gradient, macierz Jacobiego

Zacznijmy od prostszej sytuacji funkcji trzech zmiennych $f : D \subset \mathbb{R}^3 \to \mathbb{R}$.

Gradient funkcji $f(x,y,z)$ w punkcie $P_0 = (x_0,y_0,z_0)$ oznaczamy symbolem $\nabla f(P_0)$ lub $\mathrm{grad} f(P_0)$ definiujemy jako wektor pochodnych cząstkowych, czyli wektor $(f'_x(P_0), f'_y(P_0), f'_z(P_0))$. Analogicznie jest dla funkcji wielu zmiennych (czyli zniennej wektorowej $\mathbf{x} = (x_1,\ldots,x_d)$).

Odwzorowanie F o wartościach w $\mathbb{R}^k$ traktujemy jako zestawienie k funkcji $f_1,\ldots,f_k$ o wartościach skalarnych, czyli $F(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}),\ldots,f_k(\mathbf{x}))$.

Macierzą Jacobiego odwzorowania F (w punkcie $\mathbf{x}_0$) nazywamy macierz, której wierszami są kolejno (licząc od góry) gradienty funkcji $f_1,\ldots,f_k$, czyli macierz o wyrazach $\frac{\partial f_i}{\partial x_j}$. Wyznacznik tej macierzy (określony w przypadku $k = n$) nazywamy jakobianem i oznaczamy $\mathrm{Jac} f(\mathbf{x}_0)$, lub $\mathrm{Jac}_{\mathbf{x}_0} f$.

Podkreślmy, że macierz Jacobiego jest macierzą zależną od punktu $\mathbf{x}_0$, czyli tzw. macierzą funkcyjną. Podobnie, jakobian jest traktowany jako funkcja, przyjmująca w punkcie $\mathbf{x}$ wartość $\mathrm{Jac} f(\mathbf{x})$. Na przykład, dla $F(r,\phi) = (r\cos\phi, r\sin\phi)$, $F : \mathbb{R}^2 \to \mathbb{R}^2$, macierzą Jacobiego jest

$$\begin{pmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{pmatrix}$$

i jak łatwo przeliczyć (korzystając z jedynki trygonometrycznej), jakobian tego odwzorowania w punkcie (r_0,ϕ_0) wynosi r_0.

Jako zastosowanie Wniosku 1. oraz definicji różniczki -dla wektorów przyrostu postaci te_j gdzie $t \in \mathbb{R}, t \to 0$ otrzymujemy następujący opis różniczki:

Twierdzenie 1. *Różniczka zupełna jest odwzorowaniem liniowym, którego macierzą w bazach kanonicznych $\mathbb{R}^d$ oraz $\mathbb{R}^k$ jest macierz Jacobiego.*

Jak już wiemy, samo istnienie wszystkich pochodnych cząstkowych nie gwarantuje nawet ciągłości, a tym bardziej różniczkowalności funkcji. Załóżmy w dalszym ciągu, że funkcja f jest określona w otoczeniu D punktu $\mathbf{x}$. Bardzo ułatwi nam badanie różniczkowalności następujący warunek wystarczający:

Twierdzenie. *Jeśli pochodne cząstkowe $\frac{\partial f_i}{\partial x_j}$ istnieją $(\forall_{i\leqslant k}, \forall_{j\leqslant d})$ w pewnym otoczeniu punktu i są ciągłe w tym punkcie, to odwzorowanie $F = (f_1,\ldots,f_k)$ jest w tym punkcie różniczkowalne.*

Uzasadnienie w przypadku funkcji 2 zmiennych: Dla f zależnej od dwu zmiennych $(x, y) \in \mathbb{R}^2$, o wartościach w $\mathbb{R}$, przyrost argumentu (czyli zmiennej niezależnej) niech będzie wektorem $(h, k) \in \mathbb{R}^2 \setminus \{(0, 0)\}$. Odpowiadający mu przyrost wartości f, zwany przyrostem całkowitym i oznaczany tu przez Δf , czyli

$$\Delta f := f(x + h, y + k) - f(x, y).$$

Mamy rozkład

$$\Delta f = f(x + h, y + k) - f(x, y + k) + f(x, y + k) - f(x, y) = \Delta_x f + \Delta_y f$$

na sumę przyrostów w kierunkach osi OX, OY. Dla tych przyrostów, jak wiemy z poprzedniego wykładu, można stosować tw. Lagrange'a , znajdując punkty pośrednie $x_\alpha = x + \alpha h$, $y_\beta = y + \beta k$ dla pewnych $\alpha, \beta \in (0, 1)$ takie, że

$$\Delta_x f = hf'_x(x_\alpha, y + k),\ \ \Delta_y f = kf'_y(x, y_\beta).$$

Różniczka $d_(x, y)f$, o ile istnieje, na wektorze (h, k) musi przyjmować wartości $hf'_x(x, y)+kf'_y(x, y)$, więc warunek różniczkowalności oznacza, że wartość $\Delta f - d_(x, y)f(h, k)$ po podzieleniu przez $\|(h, k)\| = \sqrt{h^2 + k^2}$ ma nadal dążyć do zera gdy $\|(h, k)\| \to 0$. Ten iloraz faktycznie zmierza do zera, ponieważ jest on równy, jak łatwo sprawdzić, sumie

$$\frac{h}{\|(h, k)\|}[f'_x(x_\alpha, y + k) - f'_x(x, y)] + \frac{k}{\|(h, k)\|}[f'_y(x, y_\beta) - f'_y(x, y)].$$

Ułamki $\frac{h}{\|(h,k)\|}$ oraz $\frac{k}{\|(h,k)\|}$ mają moduły $\leqslant 1$, zaś różnice pochodnych cząstkowych w nawiasach kwadratowych zmierzają do zera, dzięki ciągłości tych pochodnych. Faktycznie, odległości między punktami $(x_\alpha, y+k)$ oraz (x, y) i między (x, y_β) oraz (x, y) nie przekraczają $\|(h, k)\|$, więc zmierzają do zera. (Przypadek większej ilości zmiennych nie stwarza żadnych dodatkowych trudności, może poza notacją przy rozbiciu przyrostu całkowitego na sumę przyrostów w kierunku poszczególnych osi.)$\square$

Powyższe twierdzenie daje tylko warunek wystarczający. Jest on daleki od koniecznego -bo już nawet dla funkcji 1 zmiennej różniczkowalność w punkcie (np. w $t = 0$) nie implikuje ciągłości w żadnym innym punkcie, np. gdy ψ jest funkcją Dirichleta (lub jakąkolwiek inną funkcją ograniczoną), to funkcja $t \mapsto t^2\psi(t)$ ma pochodną równą 0 w punkcie $t = 0$, który może być jedynym punktem ciągłości f i jedynym, w którym f jest różniczkowalna.

Zaletą (zwłaszcza zapisanej bezargumentowo) różniczki w stosunku do pochodnych cząstkowych jest unikanie żmudnych (często -wielowskaźnikowych) sumowań. Następne twierdzenie jest tego wyraźnym przykładem. Chodzi o różniczkę złożenia.

Twierdzenie (Reguła łańcucha). Jeżeli odwzorowanie $f : D \subset \mathbb{R}^d \to \mathbb{R}^k$ jest różniczkowalne w punkcie wewnętrznym x zbioru D, $y = f(x)$ oraz $g : D_1 \subset \mathbb{R}^k \to \mathbb{R}^m$ 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:

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

`Dowód. (szczegóły pominiemy)` Zauważmy, że z definicji różniczkowalności (w jej drugiej wersji) istnieją odwzorowania α : (otoczenie 0 w $\mathbb{R}^d$) $\to \mathbb{R}^k$ oraz β : (otoczenie 0 w $\mathbb{R}^k$) $\to \mathbb{R}^m$ -odpowiednio, zmierzajace w tych punktach do zera, dla których

$$\begin{aligned} f(x + h) &= f(x) + d_x f(h) + \|h\|\alpha(h), \quad \textit{gdzie} \quad \|\alpha(h)\| \to 0 \quad \textit{przy } h \to 0 \\ g(y + k) &= g(y) + d_y g(k) + \|k\|\beta(k), \quad \textit{gdzie} \quad \|\beta(k)\| \to 0 \quad \textit{przy } k \to 0 \end{aligned} \tag{3}$$

Aby poskładać te wzory, definiujemy $k := f(x + h) - f(x)$ -wyrażone przez pierwszy z tych wzorów i niech $y = f(x)$. Z ciągłości w punktach, w których f jest różniczkowalna widzimy, że $k \to 0$ gdy $h \to 0$. Takie k możemy więc wstawić do drugiego wzoru(3), pamiętając, że $y+k = f(x)+f(x+h)-f(x) = f(x+h)$, a stąd wynika, że $g(y + k) = (g \circ f)(x + h)$.

Z własności odwzorowań liniowych i z równości $k = d_x f(h) + \|h\|\alpha(h)$ wnioskujemy ograniczenie ilorazu $\frac{\|k\|}{\|h\|}$ przez pewną stałą.

Możemy też wykorzystać liniowość $d_y g$, dzięki czemu

$$d_y g(k) = d_y g(d_x f(h) + \|h\|\alpha(h)) = \big((d_y g) \circ (d_x f)\big)(h) + \|h\| d_y g(\alpha(h)).$$

Tak więc, po podstawieniach, drugi ze wzorów (3) przyjmie postać

$$(g \circ f)(x+h) = (g \circ f)(x) + \big((d_y g) \circ (d_x f)\big)(h) + \|h\| d_y g(\alpha(h)) + \big\|k\big\|\beta(k) \quad (4)$$

Funkcja $h \mapsto \|\alpha(h)\|$ zmierzająca do zera gdy $h \to 0$ –jest w pewnym otoczeniu zera ograniczona. Zastępując w ostatnim wzorze $\|k\|$ przez $\|h\|\frac{\|k\|}{\|h\|}$, możemy więc skorzystać z ograniczoności jej lewej strony w pewnym otoczeniu zera, zapisując wzór (4) w postaci

$$(g \circ f)(x+h) = (g \circ f)(x) + \big((d_y g) \circ (d_x f)\big)(h) + \|h\| \big\{ d_y g(\alpha(h)) + \frac{\|k\|}{\|h\|}\beta(k) \big\},$$

w którym wyrażenie w nawiasie { } zmierza do zera przy $\|h\| \to 0$. Dowodzi to naszej tezy. □

Zauważmy, że wynika stąd, w szczególności, zaanonsowany przed tygodniem wzór na pochodną ϕ' z funkcji postaci $t \mapsto \phi(t) = f(x(t), y(t))$. Faktycznie, pochodna $\phi'(t)$ jest wartością $d_t\phi(1)$.

Dla funkcji f o wartościach skalarnych otrzymamy następujące twierdzenie o wartości średniej: 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

$$[0,1] \ni t \to P_t := P_0 + t(P_1 - P_0).$$

Punkty postaci P_t, gdzie $0 < t < 1$ będziemy nazywali punktami wewnętrznymi tego odcinka. (Odwzorowanie to określone jest, oczywiśce, nawet dla dowolnych $t \in \mathbb{R}$.) Różniczką jest tu część liniowa tego odwzorowana, czyli odwzorowanie $\mathbb{R} \ni t \to t(P_1 - P_0)$. Jego wartość w punkcie $t = 1$, to pochodna tego odwzorowania, jet 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 $f : D \subset \mathbb{R}^d \to \mathbb{R}$ jest różniczkowalna w punktch 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 wektor $P_1 - P_0$ jest wektorem $h = (h_1, h_2, \ldots, h_d)$, to przyrost ten możemy wyrazić przy użyciu pochodnych cząstkowych: $f(P_1) - f(P_0) = \sum_{j=1}^d \frac{\partial f}{\partial x_j}(P_\theta) h_j$.

Podobnie, jak w przypadku odwzorowań $\gamma : [a,b] \to \mathbb{R}^k$, analogiczna teza przestaje być prawdziwa dla odwzorowań o wartościach wektorowych (dla $k > 1$). Mamy wówczas tylko możliwość oszacowania normy przyrostu:

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

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

Tu norma z różniczki, to norma odwzorowania liniowego. Przypomnnijmy, że dla $T : \mathbb{R}^d \to \mathbb{R}^k$ taka norma, $\|T\|$, to najmniejsza stała M, dla której na każdym z wektorów $v \in \mathbb{R}^d$ zachodzą nierówności: $\|T(v)\| \leqslant M\|v\|$. 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.

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

Definicja. Jeżeli $w \in \mathbb{R}^d$ jest wektorem niezerowym, $f : D \subset \mathbb{R}^d \to \mathbb{R}$ 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

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

Ściślej- 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

$$\partial^w f(P_0) = d_{P_0} f(w).$$

W szczególności, gdy pochodne kierunkowe istnieją, ale nie zależą w sposób liniowy od wektora kierunku, to dana funkcja nie jest różniczkowalna. Następujący przykład pokazuje, że z istnienia pochodnych kierunkwych $\partial^w f$ zależnych w sposób liniowy od w jeszcze nie wynika różniczkowalność f. Niech mianowicie

$$f(x, y) = \begin{cases} \frac{x^3 y}{x^4 + y^2}, & \text{gdy } x^2 + y^2 > 0 \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Wówczas dla wektora $w = (s, t)$ i dla $h > 0$ mamy $hw = (hs, ht)$ oraz $\frac{1}{h}(f(0 + hw) - f(0)) = \frac{1}{h} \frac{(hs)^3 ht}{h^2(h^2 s^4 + t^2)}$. To wyrażenie w przypadku $t = 0, s \neq 0$ jest stale równe 0 , a gdy $t \neq 0$ -zmierza do zera, więc $\partial^w f(0, 0) = 0$ dla każdego wektora w. Jednak różniczkowalności w zerze nie mamy. Pochodne w kierunkach osi, to pochodne cząstkowe i są one równe zero. Gdyby zachodziła różniczkowalność, mielibyśmy zmierzanie do zera przy $(x, y) \to (0, 0)$ ułamka, którego licznikiem jest $f(x, y) - f(0, 0) - d_0 f(x, y) = f(x, y)$, a mianownikiem $\sqrt{x^2 + y^2}$ Ze względu na wyższe potęgi, szybciej do zera zmierza mianownik ze wzoru na $f(x, y)$ i to ten mianownik spóbujmy "ujednorodnić" -czyli wyrównać potęgi jego składników. Podstawmy $y = x^2$ Zerem powinna być więc granica z wyrażenia

$$\frac{x^3 x^2}{\sqrt{x^2 + x^4}(2x^4)} = \frac{x^5}{2|x| x^4 \sqrt{1 + x^2}} = \frac{signum(x)}{2\sqrt{1 + x^2}}.$$

Ale dla $x > 0$ przy $x \to 0^+$ to wyrażenie zmierza do $\frac{1}{2}$, zamiast do zera.