15 Odwzorowania liniowe, różniczkowalność

Zanim podamy ogólną definicję różniczkowalności przypomnijmy sobie podstawowe fakty o odwzorowaniach liniowych:

Niech X, Y będą przestrzeniami wektorowymi unormowanymi, nad ciałem $\mathbb{R}$. Przez $\mathcal{L}(X,Y)$ oznaczmy zbiór wszystkich odwzorowań liniowych $T : X \to Y$ i niech $\mathcal{B}(X,Y) = \{T \in \mathcal{L}(X,Y) : T \texttt{ jest ciągłe } \}$.

1. Liniowość T oznacza, że $\forall_{u,v\in X}\forall_{s,t\in\mathbb{R}}\, T(su+tv) = sT(u) + tT(v)$
2. Zbiór $\mathcal{L}(X,Y)$ z naturalnie określonymi działaniami jest przestrzenią wektorową (np. suma $(T+S)(v) := T(v) + S(v)$ dwu odwzorowań liniowych jest liniowa. Iloczyn operatora T przez skalar $\alpha \in \mathbb{R}$, to operator $X \ni v \to \alpha T(v) \in Y$ -oczywiście, liniowy.)
3. Gdy dodatkowo $X = Y$, to przestrzeń $\mathcal{L}(X,X)$ oznaczamy symbolem $\mathcal{L}(X)$ i mamy tu działanie mnożenia zdefiniowane jako składanie:
$$(ST)(v) := (S \circ T)(v) = S(T(v)), \qquad v \in X.$$
Elementem neutralnym tego działania jest operator identyczności $I = I_X$, gdzie $I(v) = v, v \in X$.
4. $\mathcal{L}(X)$ ma strukturę tzw. algebry (nieprzemiennej, gdy $\dim(X) > 1$), czyli przestrzeni wektorowej z dodztkowym działaniem mnożenia, które jest rozdzielne wzgl. dodawania, łączne, łączne wzgl. mnożenia przez skalary
5. Gdy $\dim(X) < \infty$, to każde odwzorowanie $T \in \mathcal{L}(X,Y)$ jest ciągłe, czyli wówczas $\mathcal{L}(X,Y) = \mathcal{B}(X,Y)$.
6. Odwzorowanie przypisujące operatorowi $T \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^m)$ macierz o wyrazach $a_{jk} = \langle T(e_k), e_j\rangle \in M_{d\times m}$ jest bijekcją liniową, składaniu operatorów odpowiada przez ten izomorfizm -mnożenie macierzy. Kolumna o numerze k, to współczynniki wektora $T(e_k)$.
7. Każde odwzorowanie liniowe $A \in \mathcal{B}(\mathbb{R}^d, Y)$ jest postaci $A(v_1, v_2, \dots v_d) = v_1y_1 + \cdots + v_dy_d$ dla pewnych wektorów $y_j \in Y$. Na przykład, dla $d = 3$, gdy $Y = \mathbb{R}$, odwzorowanie T wyznacza trójka liczb: α, β, γ, gdzie $T(x,y,z) = \alpha x + \beta y + \gamma z$, zaś $\alpha = T(1,0,0), \beta = T(0,1,0), \gamma = T(0,0,1)$.

W zasadzie, wszystkie tezy z wyjątkiem 5. –są prostymi faktami z algebry liniowej. Gdy $X = \mathbb{R}^d$, ciągłość $T \in \mathcal{L}(X,Y)$ wynika z postaci takich odwzorowań (=teza 7.), co powinno wystarczyć dla naszych zastosowań. Podam tylko szkic dowodu w sytuacji ogólnej. Mając ustaloną bazę $f_1, f_2, \dots, f_d$ przestrzeni X, tworzymy izomorfizm algebraiczny
$\Phi : X \ni x = a_1f_1 + a_2f_2 + \cdots + a_df_d \to \Phi(x) := (a_1, a_2, \dots, a_d) \in \mathbb{R}^d$.
Wówczas $\Phi^{-1}(a_1, \dots, a_d) = a_1f_1 + \cdots + a_df_d$. Sprawdzamy, że odwzorowanie $\mathbb{R}^d \ni (a_1, \dots, a_d) \to \|\Phi^{-1}(a_1, \dots, a_d)\|$ jest również normą w $\mathbb{R}^d$, co jest bardzo łatwe. Korzystamy następnie z pierwszego z twierdzeń z poprzedniego wykładu (9) o równoważności norm. Dzięki niemu (i na podstawie poniższego lematu) uzyskujemy ciągłość odwzorowań Φ oraz Φ^{-1}. Z ciągłości odwzorowania liniowego $T_1 := T \circ \Phi^{-1} : \mathbb{R}^d \to Y$ wnioskujemy teraz ciągłość $T = T_1 \circ \Phi$.

Na przestrzeniach nieskończenie wymiarowych większość odwzorowań liniowych jest nieciągła. Np. na przestrzeni wielomianów $\mathbb{R}[x]$ z (dość nietypową) normą $\|p\| := \sup\{|p(x)| : x \in [0,1]\}$ operator różniczkowania $D : p \mapsto \frac{dp}{dx}$ nie jest ciągły, bo chociaż ciąg $p_n(t) := \frac{1}{\sqrt{n}}t^n$ dąży do zera, to ciąg pochodnych, $p_n'(t) = \sqrt{n}t^{n-1}$ –ma normy nieograniczone.

Lemat. Dla operatora $T \in \mathcal{L}(X,Y)$ ciągłość jest równoważna istinieniu stałej M takiej, że
$$\forall_{x\in X}\ \|T(x)\| \leqslant M\|x\|. \tag{1}$$

Najmniejsza ze stałych M o tej własności oznaczana jest symbolem $\|T\|$. Można wykazać, że

$$\|T\| = \sup\{\|T(x)\| : \|x\| \leqslant 1\}. \tag{2}$$

Dowód. Oczywiście, z nierówności $\|T(x)\| \leqslant M\|x\|$ wynika nawet warunek Lipschitza, bo $\|T(u) - T(v)\| = \|T(u - v)\| \leqslant M\|u - v\|$ ($\Rightarrow$ ciąłość T). Na odwrót, z ciągłości T w punkcie $x = 0$ wynika, że dla $\epsilon > 0$ istnieje $\delta > 0$ taka, że $\|v\| \leqslant \delta \Rightarrow \|T(v)\| < \epsilon$. Teraz dla $x \in X \setminus \{0\}$ niech $v := \frac{\delta}{\|x\|}x$. Dzięki jednorodności normy, $\|v\| = \frac{\delta}{\|x\|}\|x\| = \delta$, więc $\|T(v)\| < \epsilon$. Ale $x = \frac{\|x\|}{\delta}v$, więc z jednorodności normy i z liniowości T otrzymyjemy $\|T(x)\| \leqslant \frac{\|x\|}{\delta}\epsilon$ (również dla $x = 0$, bo $T(0) = 0$), czyli (1) zachodzi dla $M = \frac{\epsilon}{\delta}$. Ostatnią równość zostawmy jako proste ćwiczenie (wykorzystujące dzielenie wektora niezerowego przez jego normę). $\square$

W szczególności, będziemy często korzystali z nierówności

$$\|T(x)\| \leqslant \|T\|\|x\|. \tag{3}$$

Proszę zwrócić uwagę, że zarówno tu, jak i w lemacie, występują normy w przestrzeni Y (po lewej stronie nierówności), jak i w przestrzeni X (po prawej). W dodatku mamy jeszcze trzecią normę: $\|T\|$ określoną w przestrzeni $\mathcal{B}(X, Y)$. To, że jest to norma wynika z równości (2), bo jest to norma supremum po kuli jednostkowej. Zbieżność ciągu operatorów T_n do operatora T, czyli warunek $\|T_n - T\| \to 0$ jest zbieżnością jednostajna na kuli (i na każdym podzbiorze ograniczonym w przestrzeni X), ale nie na całej przestrzeni X. Mimo tego, używa się określenia "zbieżność jednostajna" -dla odróżnienia od innych również intensywnie badanych typów zbieżnosci w $\mathcal{B}(X, Y)$. Na kursie analizy funkcjonalnej wykażemy zupełność tak unormowanej przestrzeni $\mathcal{B}(X, Y)$, o ile Y jest zupełna. W przypadku skończenie wymiarowym zupełność możemy łatwo wywnioskować z twierdzenia o równoważności norm (wymiarem $\mathcal{B}(X, Y)$ jest iloczyn wymiarów przestrzeni X oraz przestrzeni Y - co już wiemy z teorii macierzy.)

A cały ten wstęp zaraz nam się przyda, bo różniczka, to będzie pewne odwzorowanie liniowe.

15.1 Różniczka zupełna

Definicja Mówimy, że odwzorowanie $F : D \subset \mathbb{R}^d \to \mathbb{R}^k$ jest różniczkowalne w punkcie $\mathbf{x}$, jeżeli jego dziedzina, D jest otoczeniem punktu $\mathbf{x}$ w $\mathbb{R}^d$ oraz istnieje takie odwzorowanie liniowe[1] $L : \mathbb{R}^d \to \mathbb{R}^k$, że

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - L(\mathbf{h})\|}{\|h\|} = 0. \tag{4}$$

Wówczas odwzorowanie L nazywamy różniczką zupełną F w punkcie $\mathbf{x}$, oznaczając $L = d_{\mathbf{x}}F$.

Przyrostowi argumentu o wektor $\mathbf{h}$ odpowiada przyrost wartości F o wektor $\Delta F := F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x})$. Różniczkowalność oznacza, że ten przyrost można „z dokładnością do $o(\|\mathbf{h}\|)$ przy $\mathbf{h} \to \mathbf{0}$" przybliżyć przez wartość pewnego liniowego odwzorowania na wektorze $\mathbf{h}$. To sformułowanie oznacza, że pozostała w wyniku tego przybliżenia reszta: $r(\mathbf{h}) := \Delta F - L(\mathbf{h})$ (błąd przybliżenia) podzielona przez $\|\mathbf{h}\|$ nadal zmierza do zera (czyli $r(\mathbf{h}) = o(\|\mathbf{h}\|)$ gdy $\mathbf{h} \to \mathbf{0}$).

Następujace (równoważne) sformułowanie powyższej definicji będzie zwłaszcza przydatne przy badaniu różniczki złożenia:

Odwzorowanie F jest różniczkowalne w punkcie $\mathbf{x}$ należącym do wnętrza dziedziny D, jeśli istnieją: takie odwzorowanie liniowe $L : \mathbb{R}^d \to \mathbb{R}^k$, oraz odwzorowanie $\alpha : (\texttt{otoczenie zera w } \mathbb{R}^d) \to \mathbb{R}^k$, że

$$F(\mathbf{x} + \mathbf{h}) = F(\mathbf{x}) + L(\mathbf{h}) + \|h\|\alpha(h), \quad \texttt{gdzie} \quad \lim_{h\to 0} \|\alpha(h)\| = 0. \tag{5}$$

[1] W ogólnej sytuacji w przestrzeniach unormowanych dodatkowo zakłada się ciągłość L

Oczywiście, gdy zarówno iloraz, jak i jego mianownik w (4) (tu $\|\mathbf{h}\|$) zmierzają do zera, to licznik też musi zmierzać do zera. Ale odwzorowania liniowe na $\mathbb{R}^d$ są zawsze ciągłe, więc $L(\mathbf{h}) \to 0$, co implikuje zmierzanie do zera samego przyrostu: ΔF, gdy $\mathbf{h} \to \mathbf{0}$ (=ciągłość F).

Wniosek. *W punktach, w których funkcja jest różniczkowalna, jest ona również ciągła.*

Przykład 1. Odwzorowanie liniowe $L : \mathbb{R}^d \to \mathbb{R}^k$ jest różniczkowalne i w każdym punkcie $\mathbf{x} \in \mathbb{R}^d$ jego różniczka jest równa L. Wynika to z równości $L(\mathbf{x}+\mathbf{h}) - L(\mathbf{x}) = L(\mathbf{h})$, więc licznik ze wzoru (4) jest tu stale równy zero.

Różniczkowalne są też odwzorowania afiniczne, czyli postaci $F(\mathbf{x}) = L(\mathbf{x}) + \mathbf{C}$, gdzie $\mathbf{C} \in \mathbb{R}^k$ jest stałym wektorem, zaś L jest liniowe. Częścią liniową przyrostu, czyli różniczką, jest tu L.

W ogólnym przypadku $L(\mathbf{h})$ jest „liniową częścią przyrostu" wartości F odpowiadającą przyrostowi zmiennej niezależnej (argumentu) o wektor $\mathbf{h}$ w tym sensie, że błąd względny takiego liniowego przybliżenia tego przyrostu zmierza do zera. Gdy F jest liniowe, wspomniany błąd wynosi zero.

Przykład 2. Gdy $n = k = 1$, odwzorowanie liniowe $L : \mathbb{R} \to \mathbb{R}$ musi być postaci „stała razy identyczność", czyli dla pewnej stałej c ma być $L(h) = ch$. Tą stałą musi być pochodna, $f'(x)$. Faktycznie, wówczas zamiast norm w liczniku i mianowniku mamy moduł, więc wystepujący w definicji ułamek jest równy $|\frac{f(x+h)-f(x)-ch}{h}| = |\frac{f(x+h)-f(x)}{h} - c|$, co zmierza do zera wedy i tylko wtedy, gdy $c = f'(x)$.

Przykład 3. Dla $n = 2, k = 1$ zamiast $(x_1, x_2) = \mathbf{x}$ używamy pary zmiennych (x, y). Zamiast **h**- wektor przyrostu oznaczmy przez (h, k). Ogólną postać liniowego L mamy daną wzorem $L(h, k) = ah + bk$. Wykażemy, że z różniczkowalności wyniknie istnienie pochodnych cząstkowych i równość

$$a = f'_x(x, y), b = f'_y(x, y).$$

Faktycznie, wystarczy ograniczyć się do (h, k) zmierzających do $(0, 0)$ wzdłuż osi układu współrzędnych. Wzdłuż osi OX zmierzamy, gdy $k = 0$, $h \to 0$. Ponieważ w tym przypadku $\|(h, 0)\| = |h|$, zaś $\Delta f = f(x + h, y) - f(x, y), L(h, 0) = ah$, więc do zera zmierza iloraz $\frac{1}{|h|}|\Delta f - ah| = |\frac{\Delta f}{h} - a|$, stąd $a = f'_x(x, y)$. Podobnie jest dla b (jak również dla funkcji większej ilości zmiennych).

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

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy, \tag{6}$$

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.*

15.2 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 oznaczaną symbolem

$$\frac{\partial(f_1, \ldots, f_k)}{\partial(x_1, \ldots, x_n)} \quad \text{lub, dokładniej: } \frac{\partial(f_1, \ldots, f_k)}{\partial(x_1, \ldots, x_n)}(\mathbf{x}_0)$$

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 uogólnienie rozważań z Przykładu 3. oraz zastosowanie Wniosku 1. 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 2. *Jeśli pochodne cząstkowe $\frac{\partial f_i}{\partial x_j}$ istnieją $(\forall_{i \leqslant k}, \forall_{j \leqslant d})$ w pewnym otoczeniu punktu $\boldsymbol{x}$ 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 $(x, y) \in \mathbb{R}^2$, we wzorze (7) z poprzedniego wykladu, wstawiając $x, x + h$ w miejsce x_0, x_1, $x_\alpha := x + \alpha h$ punkt pomiędzy x oraz $x + h$, (i analogicznie dla zmiennej y) otrzymamy $f(x+h, y+k) - f(x, y) =: \Delta f = f(x+h, y+k) - f(x, y+k) + f(x, y+k) - f(x, y) = \Delta_x f + \Delta_y f = h f'_x(x_\alpha, y+k) + k f'_y(x, y_\beta)$, więc

$$\Delta f - ah - bk = h[f'_x(x_\alpha, y+k) - f'_x(x, y)] + k[f'_y(x, y_\beta) - f'_y(x, y)].$$

Tu korzystamy z równości $a = f'_x(x, y), \ldots$. Po podzieleniu przez $\|(h, k)\|$ każdej z liczb $|h|, |k|$ otrzymamy wartości ograniczone przez 1, więc iloraz (występujący w warunku różniczkowalności) będzie zmierzał do zera. Faktycznie, do zera zmierzają wyrażenia w nawiasach kwadratowych (z zakładanej ciągłości f'_x, f'_y w punkcie (x, y)). Przypadek większej ilości zmiennych nie stwarza żadnych dodatkowych trudności, może poza notacją przy rozbiciu przyrostu całkowitego na sumę przyrostów w kierunku poszczególnych osi.

Jak łatwo sprawdzić, *Suma funkcji różniczkowalnych w danym punkcie jest różniczkowalna, o różniczce równej sumie różniczek:*

$$d_{\mathbf{x}}(f + g) = d_{\mathbf{x}} f + d_{\mathbf{x}} g.$$