3. Wykład. Reguły różniczkowania i twierdzenie o wartości średniej.

3.1. Interpretacja geometryczna gradientu funkcji.

Uwaga 9. Niech $f\colon G \subset \mathbb{R}^n \to \mathbb{R}$ różniczkowalna w x_0. Rozważmy pochodne kierunkowe $\frac{df}{dh}(x_0)$, gdzie $h \in \mathbb{R}^n$ jest dowolnym wersorem, tzn. $\|h\| = 1$.

Przy jakim wersorze h pochodna ma największą wartość? Możemy założyć że $Df(x_0) = \text{grad}\, f(x_0) \neq 0$. Z nierówności Schwarza:

$$\left|\frac{df}{dh}(x_0)\right| = |\text{grad}\, f(x_0) \cdot h| \leqslant \|\text{grad}\, f(x_0)\| \cdot \|h\| = ||\text{grad}\, f(x_0)||,$$

a równość otrzymujemy dla $h = \frac{\text{grad}\, f(x_0)}{\|\text{grad}\, f(x_0)\|}$.

Zatem

$$\max\{|\frac{df}{dh}(x_0)|\colon \|h\| = 1\} = \|\text{grad}\, f(x_0)\|,$$

gdzie równość otrzymujemy dla $h = \frac{\text{grad}\, f(x_0)}{\|\text{grad}\, f(x_0)\|}$. Zatem grad $f(x_0)$ jest kierunkiem najszybszego wzrostu funkcji f w punkcie x_0.

3.2. Funkcje klasy C^1.

Definicja 10. Mówimy że $f\colon G \overset{\text{otw}}{\subset} \mathbb{R}^n \to \mathbb{R}^k$ jest *różniczkowalna w sposób ciągły na* G (inaczej *klasy* C^1 *na* G), jeśli jest różniczkowalna na G i jeśli $DF\colon G \to L(\mathbb{R}^n, \mathbb{R}^k)$ jest ciągłym odwzorowaniem.

Twierdzenie 4. *Niech $f\colon G \overset{otw}{\subset} \mathbb{R}^n \to \mathbb{R}^k$. Wówczas f jest różniczkowalna w sposób ciągły na G wtedy i tylko wtedy gdy istnieją wszystkie pochodne cząstkowe $D_i f_j$ i są one ciągłe ($D_i f_j\colon G \overset{otw}{\subset} \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, n$, $j = 1, \dots, k$).*

Dowód:

($\Longrightarrow$)

$D_i f_j(x_0) = (Df(x_0)e_i)u_j$, gdzie $e_i \in \mathbb{R}^n$, $e_i = (0, \dots, 1, \dots, 0)$, jest i-tym wektorem bazy standardowej w $\mathbb{R}^n$, zaś $u_j \in \mathbb{R}^k$, $u_j = (0, \dots, 1, \dots, 0)$, jest j-tym wektorem bazy standardowej w $\mathbb{R}^k$.

Musimy pokazać, że

$$\forall_{\varepsilon>0}\exists_{\delta>0}\forall_{a,b\in G}\|b-a\| < \delta \Rightarrow |D_i f_j(b) - D_i f_j(a)| < \varepsilon.$$

Mamy

$$\begin{aligned}|D_i f_j(b) - D_i f_j(a)| &= |((Df(b) - Df(a))e_i)u_j| \leqslant \|(Df(b) - Df(a))e_i\| \\ &\leqslant |||Df(b) - Df(a)||| < \varepsilon,\end{aligned}$$

gdzie ostatnia nierówność wynika z ciągłości odwzorowania Df oraz norma odwzorowania $A \in L(\mathbb{R}^n, \mathbb{R}^k)$ jest zdefiniowana poprzez $|||A||| = \sup_{||x||\leqslant 1} ||Ax||$.

($\Longleftarrow$)

Wystarczy rozpatrzeć przypadek $k = 1$. Weźmy $a \in G$, $\varepsilon > 0$. Z otwartości G istnieje kula $B(a, r) \subseteq G$. Z ciągłości funkcji $D_i f$, możemy tak dobrać r, by

$$|D_i f(b) - D_i f(a)| < \frac{\varepsilon}{n} \forall_{b\in B(a,r)}\forall_{i=1,\dots,n}.$$

Niech $h = \sum_{i=1}^n h_i e_i$, $\|h\| < r$, $v_0 = 0$, $v_k = h_1 e_1 + ... + h_k e_k$ dla $k = 1, \dots, n$. Wtedy

$$f(a+h) - f(a) = \sum_{i=1}^n [f(a+v_i) - f(a+v_{i-1})].$$

Ponieważ $v_i = v_{i-1} + h_i e_i$, to z twierdzenia o wartości średniej dla funkcji $t \mapsto f(a + v_{i-1} + th_i e_i)$ na $[0, 1]$ dostajemy, że

$$f(a + v_i) - f(a + v_{i-1}) = h_i D_i f(a + v_{i-1} + \theta_i h_i e_i).$$

Ponieważ $|D_i f(a + v_{i-1} + \theta_i h_i e_i) - D_i f(a)| < \frac{\varepsilon}{n}$, to

$$|f(a + h) - f(a) - \sum_{i=1}^{n} h_i D_i f(a)| \leqslant \sum_{i=1}^{n} |h_i| \frac{\varepsilon}{n} \leqslant \varepsilon \|h\|.$$

Zatem f jest różniczkowalna w punkcie a i $Df(a) = (D_1 f(a), ..., D_n f(a))$. Z ciągłości $D_i f(x)$ wynika ciągłość $Df(x)$, zatem f jest klasy C^1 na G. □

3.3. **Różniczka złożenia odwzorowań.**

Twierdzenie 5 (o pochodnej złożenia). *Niech* $f\colon G_1 \overset{otw}{\subseteq} \mathbb{R}^n \to \mathbb{R}^m$ *różniczkowalna w punkcie* $x_0 \in G_1$, $f(G_1) \subseteq G_2$ *i* $g\colon G_2 \overset{otw}{\subseteq} \mathbb{R}^m \to \mathbb{R}^k$ *różniczkowalna w punkcie* $f(x_0) \in G_2$. *Wtedy odwzorowanie* $F(x) = g(f(x))$ *jest różniczkowalne w punkcie* x_0 *i zachodzi wzór:*

$$DF(x_0) = Dg(f(x_0)) \circ Df(x_0), \quad \textit{gdzie} \quad Dg(f(x_0)) \in L(\mathbb{R}^m, \mathbb{R}^k), \quad Df(x_0) \in L(\mathbb{R}^n, \mathbb{R}^m).$$

Dowód: Niech $y_0 = f(x_0)$, $A = Df(x_0)$ i $B = Dg(y_0)$. Z różniczkowalności odwzorowań f i g mamy

$$u(h) = f(x_0 + h) - f(x_0) - Ah, \quad \lim_{\|h\| \to 0} \frac{u(h)}{\|h\|} = 0, \quad \|u(h)\| = \varepsilon(h)\|h\|$$
$$v(k) = g(y_0 + k) - g(y_0) - Bk, \quad \lim_{\|k\| \to 0} \frac{v(k)}{\|k\|} = 0, \quad \|v(k)\| = \eta(k)||k||.$$

Stąd dla $k = f(x_0 + h) - f(x_0)$ otrzymujemy

$$F(x_0 + h) - F(x_0) - BAh = g(y_0 + k) - g(y_0) - BAh = B(k - Ah) + v(k) = Bu(h) + v(k).$$

Zatem

$$\frac{\|F(x_0 + h) - F(x_0) - BAh\|}{\|h\|} \leqslant |||B|||\varepsilon(h) + \eta(k)\frac{\|k\|}{\|h\|} \leqslant |||B|||\varepsilon(h) + \eta(k)(|||A||| + \varepsilon(h)) \to 0 \text{ przy } h \to 0,$$

bo $\|k\| = \|Ah + u(h)\| \leqslant |||A||| \cdot \|h\| + \varepsilon(h)\|h\|$. □

Przykład 6 (reguła łańcuchowa). Niech $F(x, y) = f(u(x, y), v(x, y))$. Wówczas

$$\frac{\partial F}{\partial x} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x}$$
$$\frac{\partial F}{\partial y} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial y}.$$

3.4. **Twierdzenia o wartości średniej i ekstrema.**

Twierdzenie 6 (Warunek konieczny istnienia ekstremum). *Niech* $f\colon G \overset{otw}{\subset} \mathbb{R}^n \to \mathbb{R}$ *ma ekstremum lokalne w punkcie* $x_0 \in G$. *Jeżeli* f *jest różniczkowalna w kierunku* $v \in \mathbb{R}^n$ *w punkcie* x_0, *to* $D_v f(x_0) = 0$.

Dowód: Bez zmniejszenia ogólności możemy założyć, że $f(x_0)$ — minimum lokalne, czyli istnieje $B(x_0, r) \subset G$ taka, że $\forall_{x \in B(x_0, r)} f(x) \geqslant f(x_0)$. Rozpatrzmy funkcje $g(t) = f(x_0 + tv) \geqslant f(x_0) = g(0) \forall_{\|tv\| < r}$. Ponieważ f jest różniczkowalna w kierunku v, to g różniczkowalna w zerze i ma tam z ostatniej nierówności minimum. Zatem $f'(0) = 0$. □

Twierdzenie 7 (O wartości średniej dla funkcji)**.** *Niech* $f\colon G \overset{otw}{\subset} \mathbb{R}^n \to \mathbb{R}$*. jeśli* f *jest różniczkowalna w każdym punkcie odcinka* $[a,b] \subset G$*, to istnieje taki punkt* $c \in [a,b]$*, że*

$$f(b) - f(a) = Df(c)(b-a).$$

Dowód: Oznaczmy $h = b - a$ i rozważmy funkcję $g(t) = f(a+th)$ dla $t \in [0,1]$. Z Twierdzenia 5 o pochodnej funkcji złożonej mamy, że

$$g'(t) = Df(a+th)h.$$

Z twierdzenia Lagrange'a dla funkcji jednej zmiennej g istnieje $t_0 \in (0,1)$, że $g(1) - g(0) = g'(t_0)$, czyli $f(b) - f(a) = Df(a+t_0h)h$. Stąd dla $c = a + t_0h$ i $h = b - a$ mamy $f(b) - f(a) = Df(c)(b-a)$. □

Twierdzenie 8. *Jeżeli* $f\colon G \overset{obszar}{\subset} \mathbb{R}^n \to \mathbb{R}$ *jest różniczkowalna oraz* $Df(x) = 0\ \forall_x \in G$*, to* f *stała.*

Dowód: Niech $a, b \in G$. Ponieważ G — obszar, to istnieje łamana $L \subset G$ łącząca punkty a i b. Na mocy Twierdzenia 7 funkcja f przyjmuje tę samą wartość na końcach każdego z odcinków z łamanej L. Stąd $f(b) = f(a)$. □

Twierdzenie 9 (O wartości średniej dla odwzorowań)**.** *Jeżeli* $f\colon G \overset{otw.,\ wypukły}{\subset} \mathbb{R}^n \to \mathbb{R}^m$ *jest różniczkowalna w* G *i istnieje stała* M *taka, że* $\forall_{x\in G}\ |||Df(x)||| \leqslant M$*, to wówczas* $\forall_{a,b\in G}\ \|f(b) - f(a)\| \leqslant M\|b-a\|$*.*

Dowód: Niech $g(t) = f(a + t(b-a)) - f(a)$ i $F(t) = \langle g(1), g(t)\rangle$ dla $t \in [0,1]$. Zauważmy, że $g(t)$ i $F(t)$ są funkcjami różniczkowalnymi. Ponadto

$$F(1) = \|g(1)\|^2 = \|f(b) - f(a)\|^2 \quad \text{i} \quad F(0) = 0.$$

Dalej, stosując twierdzenie Lagrange'a dla funkcji jednej zmiennej dostajemy

$$\|f(b) - f(a)\|^2 = F(1) - F(0) = F'(\theta) \quad \text{dla pewnego} \quad \theta \in (0,1).$$

Z Twierdzenia 5 o pochodnej złożenia i z nierówności Schwarza (Twierdzenia 1) dostajemy oszacowanie

$$F'(\theta) = \langle g(1), g'(\theta)\rangle = \langle f(b) - f(a), Df(a + \theta(b-a))\cdot(b-a)\rangle$$
$$\leqslant \|f(b) - f(a)\| \cdot \|Df(a+\theta(b-a))\cdot(b-a)\| \leqslant \|f(b)-f(a)\| \cdot |||Df(a+\theta(b-a))||| \cdot \|b-a\|.$$

Stąd

$$\|f(b) - f(a)\| \leqslant |||Df(a+\theta(b-a))||| \cdot \|b-a\| \leqslant M\|b-a\|.$$

□

Stąd podobnie jak w przypadku funkcji dostajemy:

Wniosek 2. *Jeżeli* $f\colon G \overset{obszar}{\subset} \mathbb{R}^n \to \mathbb{R}^m$ *jest różniczkowalna oraz* $Df(x) = 0\ \forall_{x\in G}$*, to* f *stała.*