ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria II: WIADOMOŚCI MATEMATYCZNE XVI (1973)

wmptm 1897

Ministerstwo Nauki i Szkolnictwa Wyższego

JAN MIKUSIŃSKI (Katowice)

O pochodnych cząstkowych i pełnych

1. Pochodne cząstkowe dotyczą funkcyj wielu zmiennych

$$f(x) = f(\xi_1, \ldots, \xi_q).$$

Symbol po lewej stronie jest nowocześniejszy i prostszy. Oznacza on, że f traktujemy jako funkcję punktu x w przestrzeni euklidesowej q-wymiarowej R^q, którego współrzędnymi są liczby rzeczywiste $\xi_1, \ldots, \xi_q$. Aby pochodną cząstkową $\partial f/\partial \xi_j$ zdefiniować używając symboliki zmiennej punktowej x, wygodnie jest wprowadzić oznaczenie e_j dla punktu przestrzeni R^q, którego j-ta współrzędna jest 1, a wszystkie pozostałe 0. Wtedy samą pochodną możemy oznaczyć przez $f^{(e_j)}$, co jest o tyle wygodne, że wartość tej pochodnej w jakimś punkcie x_0 daje się zapisać zwyczajnie jako $f^{(e_j)}(x_0)$, a używanie skomplikowanego zapisu $(\partial f/\partial \xi_j)_{x=x_0}$ staje się niepotrzebne. Definicję pochodnej cząstkowej można wtedy podać w postaci

$$f^{(e_j)}(x_0) = \lim_{\lambda \to 0} \frac{f(x_0+\lambda e_j) - f(x_0)}{\lambda}.$$

W tej definicji grają rolę tylko punkty leżące na prostej równoległej do j-tej osi współrzędnych i przechodzącej przez punkt x_0. Wskutek tego pochodna cząstkowa nie jest w istocie niczym innym, jak zwykłą pochodną z funkcji, której zakres określoności został zawężony do tej prostej. Nawet gdy znamy wszystkie pochodne cząstkowe $f^{(e_1)}(x_0), \ldots, f^{(e_q)}(x_0)$, możemy bardzo niewiele powiedzieć o zachowaniu się funkcji w otoczeniu punktu x_0. Na przykład nie możemy powiedzieć, czy ma różniczkę zupełną, a nawet czy jest ciągła w tym punkcie. Nieoperatywność pojęcia pochodnych cząstkowych przejawia się w praktyce w ten sposób, że we wszystkich twierdzeniach, gdzie pochodne te występują w założeniu, potrzebne są zawsze jeszcze inne założenia uzupełniające, przez co twierdzenia stają się bardziej skomplikowane.

Kłopotów tych można uniknąć, modyfikując nieznacznie pojęcie pochodnej cząstkowej. Dla wyraźniejszego uwidocznienia, na czym modyfikacja polega, sformułujemy najpierw pojęcie pochodnej cząstkowej w nieco inny, lecz równoważny sposób.

Mówimy, że u jest pochodną cząstkową funkcji f w punkcie x_0 i piszemy $u = f^{(e_j)}$, gdy dla każdego $\varepsilon > 0$ istnieje otoczenie V punktu x_0, takie że nierówność

$$\frac{|f(x)-f(x_0)-ue_j(x-x_0)|}{|x-x_0|} < \varepsilon$$

zachodzi dla wszystkich punktów x otoczenia V, różniących się od x_0 j-tą współrzędną i tylko tą współrzędną. Innymi słowy, punkt x ma leżeć na części wspólnej otoczenia V i prostej przechodzącej przez x_0 i równoległej do j-tej osi współrzędnych. (W nierówności powyższej iloczyn skalarny $e_j(x-x_0)$ jest równy różnicy j-tych współrzędnych punktów x i x_0. Zapis taki pozwala uniknąć wprowadzania oddzielnych symbolów dla współrzędnych rozważanych punktów.)

Zmodyfikowaną pochodną cząstkową, którą nazwiemy *pochodną pełną*, definiujemy w następujący sposób. Mówimy, że u jest pochodną pełną funkcji f w punkcie x_0, gdy dla każdego $\varepsilon > 0$ istnieje otoczenie V punktu x_0, takie że nierówność

$$\frac{|f(x)-f(y)-ue_j(x-y)|}{|x-y|} < \varepsilon \tag{1}$$

zachodzi dla wszelkich punktów x, y z otoczenia V, różniących się od siebie j-tą współrzędną i tylko tą współrzędną. Innymi słowy, punkty x i y mają leżeć na części wspólnej otoczenia V i dowolnej prostej (a więc nie koniecznie przechodzącej przez x_0), równoległej do j-tej osi współrzędnej. Przy tej definicji wszystkie punkty pełnego otoczenia V wchodzą w rachubę, co uzasadnia nazwę *pochodna pełna*. Jest widoczne, że jeżeli pochodna pełna istnieje, to istnieje też odpowiednia pochodna cząstkowa i obie są równe. Dzięki temu możemy używać tego samego symbolu $f^{(e_j)}(x_0)$ zarówno dla pochodnej cząstkowej, jak i pochodnej pełnej.

Mówiąc o pochodnej pełnej w pewnym punkcie będziemy zawsze milcząco zakładać, że dana funkcja jest określona w pewnym otoczeniu tego punktu (co nie jest konieczne w przypadku pochodnej cząstkowej, jeżeli wymiar przestrzeni jest większy niż 1).

2. Definicja pochodnej pełnej zachowuje sens także w przypadku jednej zmiennej rzeczywistej. Okazuje się, że i wtedy nie jest ona równoważna z pochodną zwykłą. Różnica pochodzi stąd, że w definicji pochodnej zwykłej oprócz punktu x_0 występuje jeszcze tylko jeden punkt zmienny x, a w definicji pochodnej pełnej oprócz x_0 występują dwa punkty zmienne x i y. Aby pokazać, że ta różnica jest istotna, rozważmy najpierw funkcję $r(x)$ pokazaną na załączonym wykresie (rys. 1).

Funkcję tę można określić analitycznie wzorem

$$r(x) = (-1)^k(x-k+\tfrac{1}{2}) \quad \text{dla } k-1 \leqslant x < k,\ k = 0, \pm 1, \pm 2, \ldots$$

Funkcja

$$f(x) = \begin{cases} x^2 r(x^{-2}) & \text{dla } x \neq 0, \\ 0 & \text{dla } x = 0, \end{cases}$$

ma w punkcie 0 pochodną 0 w zwykłym sensie, natomiast nie ma pochodnej pełnej. Istotnie z (1) wynika

$$\frac{|f(x)-f(y)|}{|x-y|} \leqslant |u|+\varepsilon,$$

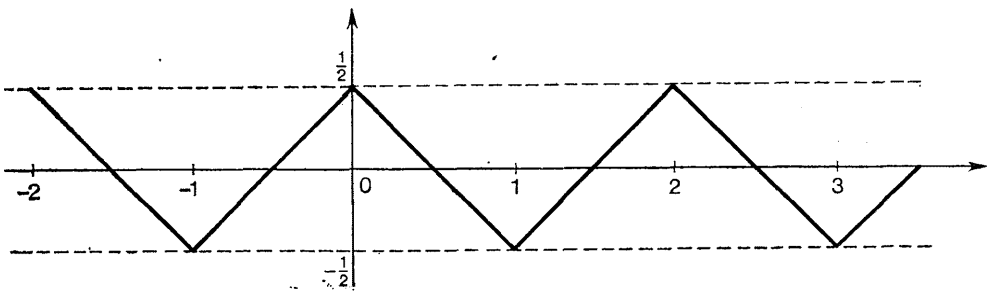
a tymczasem dla

$$x = \frac{1}{\sqrt{n}}, \quad y = \frac{1}{\sqrt{n+1}}$$

lewa strona powyższej nierówności wynosi, jak nietrudno obliczyć,

$$(2n+1)\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n+1}}\right),$$

nie jest więc ograniczona.



Rys. 1

Niewiele trudniej jest wykazać, że jeżeli w definicji funkcji $f(x)$ zastąpić $x^2 r(x^{-2})$ przez $x^3 r(x^{-2})$, to pochodna pełna również nie będzie istniała. Jeżeli natomiast zastąpić przez $x^4 r(x^{-2})$, to pochodna pełna w punkcie 0 będzie równa 0 (podobnie jak pochodna zwykła).

W zastosowaniach ważne jest następujące twierdzenie ogólne:

TWIERDZENIE 1. *Jeżeli X jest zbiorem punktów, w których pochodna pełna istnieje, to pochodna ta jest ciągła na X.*

W dowodzie tego twierdzenia wygodnie jest wprowadzić pojęcie kuli wielowymiarowej. W przestrzeni q-wymiarowej R^q przez kulę o promieniu ϱ i środku w punkcie x_0 rozumiemy zbiór punktów $x \in R^q$, spełniających nierówności $|x-x_0| < \varrho$. Będziemy ją oznaczać przez $\varrho K+x_0$. Kula taka jest otoczeniem punktu x_0. Ponadto dla dowolnego otoczenia punktu x_0 można tak dobrać ϱ, że kula $\varrho K+x_0$ znajdzie się w tym otoczeniu.

Dowód twierdzenia 1. Wiemy, że $f^{(e_j)}$ istnieje na X. Jeżeli więc $x_0 \in X$, to dla dowolnie ustalonego $\varepsilon > 0$ istnieje taka kula $\varrho_0 K+x_0$, że

$$|f(x)-f(y)-f^{(e_j)}(x_0)e_j(x-y)| < \frac{\varepsilon}{2}|x-y|$$

dla punktów x, y należących do kuli $\varrho_0 K+x_0$ i różniących się tylko j-tą współrzędną. Niech $V = \frac{1}{2}\varrho_0 K+x_0$. Wtedy dla każdego punktu $z \in V \cap X$ istnieje taka liczba dodatnia $\varrho < \varrho_0$, że

$$|f(x)-f(y)-f^{(e_j)}(z)e_j(x-y)| < \frac{\varepsilon}{2}|x-y|$$

dla punktów x, y należących do kuli $U = \varrho K+x_0$ i różniących się tylko j-tą współrzędną. Ponieważ z $U < \varrho_0 K+x_0$, więc w szczególności za y możemy przyjąć z, a za x dowolny punkt należący do U i różniący się od z tylko j-tą współrzędną. Wtedy z powyższych nierówności otrzymujemy

$$|f^{(e_j)}(z)(x-z)-f^{(e_j)}(x_0)(x-z)| < \varepsilon|x-z|,$$

a stąd wynika, że

$$|f^{(e_j)}(z)-f^{(e_j)}(x_0)| < \varepsilon. \tag{2}$$

Udowodniliśmy, że dla dowolnego $x_0 \in X$ i dowolnego $\varepsilon > 0$ istnieje takie otoczenie V punktu x_0, że dla każdego $z \in V \cap X$ zachodzi nierówność (2). Dowodzi to, że w zbiorze X funkcja $f^{(e_j)}$ jest ciągła w punkcie x_0. Ponieważ x_0 jest dowolnym punktem zbioru X, więc funkcja $f^{(e_j)}$ jest ciągła na całym zbiorze X.

Z tego że funkcja ma pewną pochodną pełną w pewnym zbiorze nie wynika, że sama musi być ciągła. Niech na przykład $D(\xi)$ oznacza funkcję Dirichleta, tzn. funkcję jednej zmiennej rzeczywistej, równą 1 dla wymiernych ξ i równą 0 dla niewymiernych ξ. Wtedy funkcja $f(x) = D(\xi_q)$ ma pochodne pełne $f^{(e_1)}, \ldots, f^{(e_{q-1})}$ w każdym punkcie przestrzeni q-wymiarowej R^q, ale sama nie jest ciągła w żadnym punkcie. Inaczej rzecz się ma, gdy wszystkie pochodne pełne $f^{(e_1)}, \ldots, f^{(e_q)}$ istnieją na pewnym zbiorze X. Wtedy nie tylko one, ale także sama funkcja f jest ciągła na X, jak to zobaczymy w następnym paragrafie.

3. Mówimy, że funkcja f spełnia warunek Lipschitza ze stałą $k > 0$ w pewnym punkcie x_0, jeżeli istnieje takie otoczenie V punktu x_0, że nierówność

$$|f(x)-f(y)| < k|x-y|$$

zachodzi dla wszystkich punktów x i y z tego otoczenia.

Z definicji tej wynika, że jeżeli funkcja spełnia warunek Lipschitza ze stałą k w pewnym punkcie, to spełnia go z tą samą stałą we wszystkich punktach pewnego otoczenia tego punktu. Ponadto, jeżeli funkcja f spełnia warunek Lipschitza w pewnym punkcie, to jest ciągła w pewnym otoczeniu tego punktu.

TWIERDZENIE 2. *Jeżeli funkcja f ma w pewnym punkcie $x_0 \in R^q$ wszystkie pochodne pełne, ograniczone przez stałą $k > 0$, to dla każdego $\varepsilon > 0$ istnieje otoczenie tego punktu, w którym spełnia warunek Lipschitza ze stałą $\sqrt{q}k+\varepsilon$.*

Dowód twierdzenia 2. Z założenia istnienia pochodnych pełnych wynika,

że istnieje taka kostka $U = Q+x_0$ [1], że nierówność

$$\frac{|f(x)-f(y)-f^{(e_j)}(x_0)e_j(x-y)|}{|x-y|} < \frac{\varepsilon}{\sqrt{q}}$$

zachodzi dla punktów x, y, różniących się tylko j-tą zmienną. Niech teraz x i y będą dowolnymi różnymi punktami w U (różniącymi się nie koniecznie tylko jedną współrzędną). Wtedy istnieją punkty $y_0, y_1, \ldots, y_{p-1}, y_p$ $(1 \leqslant p \leqslant q)$ takie, że $y_0 = x$, $y_p = y$ i punkt $y_\varkappa$ różni się od $y_{\varkappa-1}$ tylko współrzędną o wskaźniku $j_\varkappa$ ($j_\varkappa \neq j_\mu$ dla $\varkappa \neq \mu$). Z założenia, że U jest kostką wynika, że wszystkie te punkty należą do U. Można napisać

$$\frac{f(x)-f(y)}{|x-y|} = \sum_{\varkappa=1}^{p} \frac{|y_{\varkappa-1}-y_\varkappa|}{|x-y|} \frac{f(y_{\varkappa-1})-f(y_\varkappa)-f^{(e_{j_\varkappa})}(x_0)e_j(y_{\varkappa-1}-y_\varkappa)}{|y_{\varkappa-1}-y_\varkappa|} +$$

$$+ \sum_{\varkappa=1}^{p} \frac{|y_{\varkappa-1}-y_\varkappa|}{|x-y|} f^{(e_{j_\varkappa})}(x_0).$$

Ponieważ $|y_{\varkappa-1}-y_\varkappa| : |x-y| \leqslant 1 : \sqrt{p}$, więc

$$\frac{|f(x)-f(y)|}{|x-y|} < \sqrt{p}\left(\frac{\varepsilon}{\sqrt{q}}+k\right) \leqslant \sqrt{q}\,k+\varepsilon.$$

Z twierdzenia 2 wynika, że jeżeli f ma wszystkie pochodne pełne w pewnym punkcie, to jest ciągła w pewnym otoczeniu tego punktu. Stąd dalej wynika, że jeżeli f ma wszystkie pochodne pełne na pewnym zbiorze X, to jest ciągła w pewnym zbiorze otwartym $\mathcal{O}$ zawierającym X (ale pochodne nie muszą być ciągłe w $\mathcal{O}$, a nawet mogą nie istnieć w żadnym punkcie poza X). Z uwag tych i z twierdzenia 1 wynika od razu

TWIERDZENIE 3. *Jeżeli f ma wszystkie pochodne pełne w zbiorze otwartym $X \subset R^q$, to jest klasy C^1 w X (tzn. zarówno sama funkcja f, jak i jej wszystkie pochodne pełne są ciągłe w X).*

Twierdzenie to pokazuje, że w przypadku zbiorów otwartych pojęcie pochodnych pełnych nie wnosi nic nowego. Wyraźna jednak różnica zarysowuje się, gdy zbiór X, na którym f ma pochodne pełne, nie jest otwarty. Różnica ta występuje już w przypadku jednej zmiennej rzeczywistej. Pokażemy to na odpowiednio skonstruowanym przykładzie.

Niech, jak w paragrafie 2,

$$r(x) = (-1)^k(x-k+\tfrac{1}{2}) \quad \text{dla } k-1 \leqslant x < k,\ k = 0, \pm 1, \pm 2, \ldots$$

---

[1] Tzn. zbiór punktów $x = (\xi_1, \ldots, \xi_q) \in R^q$ spełniających nierówności $|\xi_j-\xi_{0j}| \leqslant \varrho$, gdzie $\xi_{01}, \ldots, \xi_{0q}$ są współrzędnymi punktu x_0.

Wtedy dla funkcji

$$f(x) = \sum_{n=1}^{\infty} 6^{-n} r(2^n x)$$

pochodna istnieje prawie wszędzie na osi rzeczywistej, a mianowicie wszędzie poza zbiorem Y punktów x, dla których istnieje skończone rozwinięcie dwójkowe

$$x = a_p 2^p + \ldots + a_s 2^s,$$

gdzie p i s ($p > s$) są liczbami całkowitymi (dodatnimi lub ujemnymi), a współczynniki a_k są równe 1 lub 0. Zbiór Y jest gęsty na osi rzeczywistej, a więc w każdym przedziale istnieją punkty, w których pochodna pełna nie istnieje. Co więcej, jak okażemy, w żadnym punkcie zbioru Y pochodna pełna (która istnieje poza Y) nie da się przedłużyć do funkcji ciągłej. A więc punkty zbioru Y są istotnymi punktami nieciągłości tej pochodnej.

Zanim przystąpimy do dowodu opisanych własności, zwróćmy uwagę na to, że gdyby w definicji funkcji $f(x)$ zamienić role liczb 6 i 2, to otrzymalibyśmy funkcję, która nigdzie nie ma pochodnej (nawet zwykłej).

Niech teraz z będzie punktem, który nie rozwija się w skończony ułamek dwójkowy, i niech ε będzie dowolną liczbą dodatnią. Ustalamy liczbę naturalną p, taką że $3^{-p} < \varepsilon$ i następnie liczbę całkowitą m, taką że z należy do przedziału V, określonego nierównościami $(m-1)2^{-p} < x < m \cdot 2^{-p}$. Niech

$$r'(x) = (-1)^k \quad \text{dla } k-1 \leqslant x < k,\ k = 0, \pm 1, \pm 2, \ldots$$

oraz

$$a = \sum_{n=1}^{\infty} 3^{-n} r'(2^n z).$$

Wtedy dla $x, y \in V$, $x \neq y$, mamy

$$\frac{f(x)-f(y)-a(x-y)}{x-y} = \sum_{n=1}^{p} \left(3^{-n} \frac{r(2^n x)-r(2^n y)}{2^n x - 2^n y} - 3^{-n} r'(2^n z) \right) +$$

$$+ \sum_{n=1}^{\infty} 3^{-n} \frac{r(2^n x)-r(2^n y)}{2^n x - 2^n y} - \sum_{n=p+1}^{\infty} 3^{-n} r'(2^n z).$$

Ale w pierwszej sumie po prawej stronie równości wszystkie wyrazy są równe 0, a w drugiej i trzeciej sumie wyrazy są mniejsze od 3^{-n}. Wobec tego

$$\frac{|f(x)-f(y)-a(x-y)|}{|x-y|} \leqslant 0 + \tfrac{1}{2} 3^{-p} + \tfrac{1}{2} 3^{-p} < \varepsilon,$$

co dowodzi, że a jest pochodną pełną funkcji f w punkcie z.

Przypuśćmy teraz, że z rozwija się w skończony ułamek dwójkowy. Wtedy istnieje taka liczba nieparzysta m, że $z = m \cdot 2^{-p}$. Jeżeli $z < y < x < (m+1)2^{-p}$, to

$$\frac{f(x)-f(y)}{x-y} = S_{p-1} + 3^{-p}(-1)^{\frac{m+1}{2}} + \sum_{n=p+1}^{\infty} 3^{-n} \frac{r(2^n x)-r(2^n y)}{2^n x - 2^n y},$$

gdzie

$$S_{p-1} = \sum_{n=1}^{p-1} 3^{-n} \frac{r(2^n x)-r(2^n y)}{2^n x - 2^n y} = \sum_{n=1}^{p-1} 3^{-n} r'(2^n z).$$

Jeżeli $(m+1)/2$ jest liczbą parzystą, to

$$\frac{f(x)-f(y)}{x-y} \geqslant S_{p-1} + 3^{-p} - \sum_{n=p+1}^{\infty} 3^{-n} = S_{p-1} + \tfrac{1}{2} \cdot 3^{-p}.$$

Podobnie znajdujemy dla $(m-1)2^{-p} < y < x < z$:

$$\frac{f(x)-f(y)}{x-y} \leqslant S_{p-1} - \tfrac{1}{2} \cdot 3^{-p}.$$

Stąd wynika, że pochodna pełna nie istnieje w punkcie z. Wynika ponadto, że we wszystkich punktach przedziału $\left((m-1)2^{-p}, z\right)$, gdzie pochodna pełna istnieje, jest większa lub równa $S_{p-1} + 3^{-p}$, a w punktach przedziału $\left(z, (m+1)2^{-p}\right)$, gdzie istnieje, jest mniejsza lub równa $S_{p-1} - 3^{-p}$. Wobec tego pochodna pełna nie daje się przedłużyć na punkt z, tak, żeby była w nim ciągła. Do takiego samego rezultatu dochodzimy, gdy $(m+1)/2$ jest liczbą nieparzystą, należy tylko w rachunkach zmodyfikować odpowiednio nierówności.

4. Wszyscy matematycy, którym pokazywałem definicję pochodnej pełnej, zawsze od razu mnie pytali, czy przy tej definicji pochodne mieszane są przemienne, innymi słowy, czy zachodzi dla nich równość

$$(f^{(e_i)})^{(e_j)} = (f^{(e_j)})^{(e_i)}.$$

Pytanie to można stawiać w stosunku do jednego określonego punktu, ale żeby miało ono sens, pochodne wewnętrzne muszą być określone w pewnym otoczeniu tego punktu. Przy tym naturalnym założeniu, odpowiedź na pytanie jest pozytywna. Mamy mianowicie

TWIERDZENIE 4. *Jeżeli pochodne pełne $f_i = f^{(e_i)}$ i $f_j = f^{(e_j)}$ istnieją w otoczeniu pewnego punktu x_0, a ponadto istnieją w tym punkcie pochodne pełne $f_i^{(e_j)}(x_0)$ i $f_j^{(e_i)}(x_0)$, to są sobie równe.*

Dowód. Z twierdzenia 1 wynika, że istnieje otoczenie U punktu x_0, w którym funkcje f_i i f_j są ciągłe. Niech ε będzie dowolną liczbą dodatnią. Z definicji pochodnej pełnej $f_i^{(e_j)}$ wynika, że istnieje otoczenie V punktu x_0, zawarte w U,

takie że

$$\frac{|f_i(y)-f_i(z)-f_i^{(e_j)}(x_0)e_j(y-z)|}{|y-z|} < \frac{\varepsilon}{2} \tag{3}$$

zachodzi dla punktu $y, z \in V$, różniących się tylko j-tą współrzędną. Niech

$$A(x, \lambda, \varkappa) = \frac{f(x+e_i\lambda+e_j\varkappa)-f(x+e_i\lambda)-f(x+e_j\varkappa)+f(x)}{\lambda\varkappa}.$$

Jeżeli punkty $x+e_i\lambda+e_j\varkappa$, $x+e_i\lambda$, $x+e_j\varkappa$ i x należą do V, to z ciągłości f_i wynika, że

$$A(x, \lambda, \varkappa) = \frac{f_i(x+e_i\theta\lambda+e_j\varkappa)-f_i(x+e_i\theta\lambda)}{\varkappa},$$

gdzie $0 < \theta < 1$. Ponieważ argumenty funkcji w liczniku po prawej równości należą do V i różnią się tylko j-tą pochodną, z (3) wynika, że

$$|A(x, \lambda, \varkappa)-f_i^{(e_j)}(x_0)| < \varepsilon/2.$$

Podobnie, zamieniając role i i j, znajdujemy

$$|A(x, \lambda, \varkappa)-f_j^{(e_i)}(x_0)| < \varepsilon/2.$$

Z ostatnich dwóch nierówności wynika

$$|f_i^{(e_j)}(x_0)-f_j^{(e_i)}(x_0)| < \varepsilon.$$

Ponieważ za ε możemy wziąć dowolną liczbę dodatnią, wynika stąd, że

$$f_i^{(e_j)}(x_0) = f_j^{(e_i)}(x_0).$$

4. Ostatnie twierdzenie pokazuje harmonijność i prostotę teorii pochodnych pełnych w porównaniu do teorii pochodnych cząstkowych tradycyjnych (które są nieprzemienne). Prostota ta bardziej jeszcze uwydatnia się przy rozważaniu różniczki zupełnej. Gdy

$$f^{(e_1)}(x_0) = a_1, \quad \ldots, \quad f^{(e_q)}(x_0) = a_q,$$

to funkcja liniowa

$$a_1\varkappa_1+\ldots+a_q\varkappa_q \tag{4}$$

zmiennych $\varkappa_1, \ldots, \varkappa_q$ bywa tradycyjnie nazywana różniczką zupełną funkcji f w punkcie x_0, ale tylko wówczas, gdy pewien dodatkowy warunek jest spełniony. Ten dodatkowy warunek nazwijmy *warunkiem Stolza*, nie wchodząc na razie w to, na czym on polega. Zauważmy natomiast, że wprowadzając oznaczenia wektorowe $A = (a_1, \ldots, a_q)$, $h = (\varkappa_1, \ldots, \varkappa_q)$, różniczkę (4) można zapisać krótko w postaci Ah. Ponieważ wektor h jest zmienną niezależną, różniczka jest w zupełności określona przez wektor A i matematycznie wyjdzie na to samo, jeżeli za różniczkę będziemy uważać po prostu wektor A, o ile oczywiście warunek Stolza jest spełniony. Mamy więc sytuację taką: wektor A istnieje zawsze, gdy istnieją pochodne cząstkowe, ale nazywamy go różniczką tylko wtedy, gdy warunek Stolza jest spełniony. To

subtelne rozróżnienie jest trochę przysłowiowym dzieleniem włosa na dwoje. Sytuacja upraszcza się, gdy zamiast tradycyjnych pochodnych cząstkowych rozważamy pochodne pełne, ponieważ istnienie ich pociąga za sobą automatycznie spełnienie warunku Stolza. Mamy mianowicie

TWIERDZENIE 5. *Funkcja f ma w punkcie $x_0 \in R^q$ wszystkie pochodne pełne wtedy i tylko wtedy, gdy istnieje wektor $A = (a_1, \ldots, a_q)$ o tej własności, że dla każdego $\varepsilon > 0$ istnieje otoczenie V punktu x_0, takie że*

$$\frac{|f(x)-f(y)-A(x-y)|}{|x-y|} < \varepsilon \tag{5}$$

dla wszystkich par różnych punktów $x, y \in V$. Wtedy zachodzą też równości $f^{(e_j)}(x_0) = a_j$ $(j = 1, \ldots, q)$.

Zanim udowodnimy to twierdzenie, zauważmy jeszcze, że zawarty w nim warunek staje się warunkiem Stolza, gdy przyjmiemy $y = x_0$. Warunek Stolza jest więc trochę słabszy; gdy istnieją wszystkie pochodne pełne, jest tym bardziej spełniony.

Dowód twierdzenia 5. Jeżeli x i y różnią się tylko j-tą współrzędną, to nierówność (5) redukuje się do (1) (wtedy $u = a_j$), skąd wynika, że a_j jest pochodną pełną. Odwrotnie, jeżeli $a_1, \ldots, a_q$ są pochodnymi pełnymi, to istnieje taka kostka $V = \varrho Q + x_0$, że

$$\frac{|f(x)-f(y)-a_j e_j(x-y)|}{|x-y|} < \frac{\varepsilon}{q}$$

dla punktów $x, y \in V$ różniących się tylko j-tą współrzędną.

Niech teraz x i y będą dowolnymi różnymi punktami w V (różniącymi się nie koniecznie tylko jedną współrzędną). Wtedy istnieją punkty $y_0, y_1, \ldots, y_{p-1}, y_p$ $(1 \leqslant p \leqslant q)$, takie że $y_0 = x$, $y_p = y$ i punkt $y_\varkappa$ różni się od $y_{\varkappa-1}$ tylko współrzędną o wskaźniku $j_\varkappa$ $(j_\varkappa \neq j_\mu$ dla $\varkappa \neq \mu)$. Z założenia, że V jest kostką wynika, że wszystkie te punkty należą do V. Wobec tego

$$I_\varkappa = \frac{|f(y_{\varkappa-1})-f(y_\varkappa)-a_{j_\varkappa} e_{j_\varkappa}(y_{\varkappa-1}-y_\varkappa)|}{|y_{\varkappa-1}-y_\varkappa|} < \frac{\varepsilon}{q} \quad \text{dla } \varkappa = 1, \ldots, p,$$

a stąd

$$\frac{|f(x)-f(y)-A(x-y)|}{|x-y|} \leqslant I_1 + \ldots + I_p < \frac{p}{q}\varepsilon \leqslant \varepsilon,$$

ponieważ $|y_{\varkappa-1}-y_\varkappa| \leqslant |x-y|$.

W przypadku $q = 1$ wektor A w twierdzeniu 5 redukuje się do jedynej współrzędnej a_1, która jest równa pochodnej $f^{(e_1)}(x_0)$. W tym przypadku pochodną tę można oznaczać po prostu $f'(x_0)$ i pisać $A = f'(x_0)$. Oznaczenie to można przenieść także na dowolną liczbę wymiarów; wtedy $f'(x_0)$ jest wektorem o współrzędnych $f^{(e_1)}(x_0), \ldots, f^{(e_q)}(x_0)$. Zamiast mówić, że funkcja f ma wszystkie pochodne pełne, można mówić, że f jest w pełni różniczkowalna, a $f'(x_0)$ jest jej różniczką.

5. Do tej pory nie powiedzieliśmy, jakie wartości przyjmują rozważane w tym artykule funkcje. Ponieważ jednak używaliśmy wielokrotnie znaku bezwzględnej wartości czy modułu, czytelnik mógł się domyślać, że chodzi o funkcje o wartościach rzeczywistych lub zespolonych. Jedna i druga interpretacja jest dobra. Ale jest też możliwa ogólniejsza interpretacja. Wartości rozważanych funkcyj mogą być elementami dowolnie ustalonej przestrzeni Banacha (przestrzeń liczb rzeczywistych i przestrzeń liczb zespolonych są szczególnymi przestrzeniami Banacha). Wtedy występujące we wzorach moduły należy uważać za normy rozważanych elementów. Ponieważ norma ma takie same algebraiczne własności jak moduł, wszystkie przeprowadzone dotąd rachunki i rozumowania pozostają w mocy. Trzeba tylko zwrócić uwagę na to, że jeżeli rozważana funkcja ma wartości w pewnej przestrzeni Banacha $\mathscr{X}$, to także wszystkie pochodne cząstkowe i pochodne pełne mają wartości w tej przestrzeni. Wskutek tego rozważany w ostatnim paragrafie wektor A nie jest wektorem w przestrzeni euklidesowej, a jego współrzędne $a_1, \ldots, a_q$ nie są liczbami, lecz elementami przestrzeni $\mathscr{X}$.

Przestrzeń euklidesowa R^q jest szczególną przestrzenią Banacha, w której norma jest równa odległości punktu od początku układu. Argumenty rozważanych w poprzednich paragrafach funkcyj należą do R^q (lub w szczególności do R^1), a wartości mogą należeć do dowolnie ustalonej przestrzeni Banacha $\mathscr{X}$. Znak $|\ |$ w odniesieniu do punktów z R^q oznacza normę w R^q, a w odniesieniu do elementów w $\mathscr{X}$ — normę w $\mathscr{X}$, co nie prowadzi do nieporozumień.

W dalszym ciągu tego paragrafu będziemy rozważali funkcje f, których zarówno wartości, jak i argumenty należą do tej samej przestrzeni R^q. Wtedy funkcja $f(x)$ może być interpretowana jako układ q funkcyj q zmiennych

$$\varphi_1(\xi_1, \ldots, \xi_q), \ldots, \varphi_q(\xi_1, \ldots, \xi_q),$$

gdzie $f = (\varphi_1, \ldots, \varphi_q)$ i $x = (\xi_1, \ldots, \xi_q)$. Różniczka $f'(x_0)$ jest wtedy wektorem

$$f'(x_0) = \left(f^{(e_1)}(x_0), \ldots, f^{(e_q)}(x_0)\right),$$

którego każda współrzędna $f^{(e_j)}(x_0)$ jest znowu wektorem

$$f^{(e_j)}(x_0) = \left(\varphi_1^{(e_j)}(x_0), \ldots, \varphi_q^{(e_j)}(x_0)\right).$$

Wobec tego różniczkę można w tym przypadku interpretować jako macierz kwadratową

$$f'(x) = \begin{pmatrix} \varphi_1^{(e_1)}(x_0), \ldots, \varphi_1^{(e_q)}(x_0) \\ \cdots\cdots\cdots\cdots \\ \varphi_q^{(e_1)}(x_0), \ldots, \varphi_q^{(e_q)}(x_0) \end{pmatrix}.$$

W twierdzeniu 5 macierz ta jest oznaczona literą A, a iloczyn $A(x-y)$ jest iloczynem macierzy przez wektor.

Odwracalność macierzy $f'(x_0)$ jest związana z odwracalnością funkcji f w otoczeniu punktu x_0. Mamy mianowicie

TWIERDZENIE 6. *Jeżeli funkcja f o wartościach w R^q jest w pełni różniczkowalna w punkcie $x_0 \in R^q$, a macierz $f'(x_0)$ jest odwracalna, to funkcja f jest odwracalna w otoczeniu punktu x_0. Funkcja odwrotna g jest w pełni różniczkowalna w punkcie $y_0 = f(x_0)$, a macierz $g'(y_0)$ jest odwrotna do $f'(x_0)$.*

Dowód tego twierdzenia poprzedzimy kilkoma uwagami. Pełna różniczkowalność funkcji w pewnym punkcie jest własnością bardzo mocną. Wynika z niej w szczególności, że funkcja jest określona w pewnym otoczeniu i jest w nim ciągła (zob. rozdział 3). Funkcja f jest więc ciągła w otoczeniu punktu x_0, a funkcja g jest ciągła wobec twierdzenia 6 w otoczeniu punktu y_0. Wobec odwracalności f oznacza to, że f jest homeomorfizmem w otoczeniu x_0. Najdelikatniejszym miejscem w dowodzie jest wykazanie, że funkcja g jest określona w pewnym otoczeniu punktu y_0. Dopiero mając zapewnioną tę określoność, możemy dowodzić jej pełnej różniczkowalności. Poza tym dowód nasz będzie się opierał na następującym twierdzeniu o macierzach (które podajemy bez dowodu jako znane):

TWIERDZENIE O MACIERZACH. *Dla każdej macierzy A (o q wierszach i q kolumnach) istnieje taka liczba M, że $|Ax| \leqslant M|x|$ dla wszelkich $x \in R^q$. Macierz A jest odwracalna wtedy i tylko wtedy, gdy ponadto istnieje taka liczba $m > 0$, że $|Ax| \geqslant m|x|$ dla wszelkich $x \in R^q$.*

Dowód twierdzenia 6. Z odwracalności macierzy $A = f'(x_0)$ wynika istnienie takiej liczby $m > 0$, że $|Ax| \geqslant m|x|$ dla wszelkich $x \in R^q$. Niech W będzie otoczeniem punktu x_0, w którym funkcja f jest ciągła. Z twierdzenia 5 wynika, że dla każdego $\varepsilon > 0$ istnieje taka kula domknięta $U_\varepsilon = \varrho_\varepsilon K + x_0$ zawarta w W, że

$$\frac{|f(x)-f(y)-A(x-y)|}{|x-y|} \leqslant \varepsilon \quad \text{dla } x, y \in U_\varepsilon,\ x \neq y.$$

Możemy założyć, że $U_{\varepsilon_1} \subset U_{\varepsilon_2}$ dla $\varepsilon_1 < \varepsilon_2$. Z własności liczby m wynika, że

$$|f(x)-f(y)-A(x-y)| \leqslant \frac{\varepsilon}{m}|A(x-y)| \quad \text{dla } x, y \in U_\varepsilon \tag{6}$$

oraz

$$|f(x)-f(y)| \geqslant \left(1-\frac{\varepsilon}{m}\right)|A(x-y)| \geqslant (m-\varepsilon)|x-y| \quad \text{dla } x, y \in U. \tag{7}$$

Dla $\varepsilon < m$ wynika z ostatniej nierówności, że funkcja f jest różnowartościowa, a więc odwracalna na kuli U_ε.

Udowodnimy, że dla każdego ustalonego $\varepsilon < m/2$ obraz $f(U_\varepsilon^0)$ wnętrza kuli U_ε jest zbiorem otwartym. Istotnie, niech $a \in U_\varepsilon^0$ i niech $S = rK + a$ będzie kulą domkniętą zawartą w U_ε^0. Pokażemy, że obraz $f(S)$ zawiera kulę $pK + f(a)$, gdzie $p = \frac{1}{2}(m-\varepsilon)r$. W tym celu ustalamy takie b, że $|b-f(a)| < p$ i definiujemy

$$F(x) = |b-f(x)| \quad \text{dla } x \in S.$$

Jeżeli $|x-a| = r$, to z (7) otrzymujemy

$$2p \leqslant |f(x)-f(a)| \leqslant F(x)+F(a) < F(x)+p.$$

Zatem

$$F(a) < p < F(x) \quad \text{dla } |x-a| = r. \tag{8}$$

Ponieważ funkcja f jest ciągła na kuli domkniętej S, istnieje punkt $c \in S$, taki że $F(c) \leqslant F(x)$ dla wszystkich $x \in S$. Wobec (8) c należy do wnętrza S^0 kuli S, bo równość $|x-a| = r$ oznacza, że x leży na brzegu tej kuli.

Niech $w = b-f(c)$. Ponieważ A jest macierzą odwracalną, istnieje takie $h \in R^q$, że $Ah = w$. Wybieramy $\lambda, 0 < \lambda < 1$, tak małe, że $c+\lambda h \in S$. Wtedy

$$|f(c)-b+A\lambda h| = (1-\lambda)|w|.$$

Ale wobec (6) mamy

$$|f(c+\lambda h)-f(c)-A\lambda h| \leqslant \frac{\lambda}{2}|w|.$$

Ponieważ $F(c+\lambda h)$ jest długością sumy wektorów, po lewej stronie dwóch ostatnich wzorów, mamy

$$F(c+\lambda h) \leqslant (1-\tfrac{1}{2}\lambda)F(c),$$

gdyż $|w| = F(c)$. Gdyby było $F(c) > 0$, to mielibyśmy stąd $F(c+\lambda h) < F(c)$, co przeczy definicji punktu c. Zatem $F(c) = 0$, czyli $f(c) = b$.

Udowodniliśmy, że dla każdego punktu $f(a)$ obrazu $f(U^0)$ istnieje kula o środku w $f(a)$ zawarta w $f(U^0)$. Zatem $f(U_\varepsilon^0)$ jest zbiorem otwartym zawierającym $y_0 = f(x_0)$. W konsekwencji $f(U_\varepsilon)$ jest otoczeniem punktu y_0. Napiszmy

$$R(x,y) = \frac{f(x)-f(y)-A(x-y)}{|x-y|}. \tag{9}$$

Wtedy $|R(x,y)| \leqslant \varepsilon$ dla $x, y \in U_\varepsilon$, $x \neq y$, czyli dla $f(x), f(y) \in f(U_\varepsilon)$, $f(x) \neq f(y)$ ze względu na wzajemną jednoznaczność odwzorowania U_ε w $f(U_\varepsilon)$. Podstawiając $x = g(\bar{x})$, $y = g(\bar{y})$, mamy

$$\left|R\big(g(\bar{x}), g(\bar{y})\big)\right| \leqslant \varepsilon \quad \text{dla } \bar{x}, \bar{y} \in f(U_\varepsilon),\ \bar{x} \neq \bar{y}.$$

Jeżeli B oznacza macierz odwrotną do A, to z (9) wynika

$$BR\big(g(\bar{x}), g(\bar{y})\big) = \frac{B(\bar{x}-\bar{y})-\big(g(\bar{x})-g(\bar{y})\big)}{|g(\bar{x})-g(\bar{y})|}.$$

Ponadto, gdy założymy, że $0 < \varepsilon < m/2$, to z (7) mamy

$$\frac{|g(\bar{x})-g(\bar{y})|}{|\bar{x}-\bar{y}|} \leqslant \frac{2}{m}.$$

Na podstawie twierdzenia o macierzach istnieje taka liczba M, że $|Bx| \leqslant M|x|$ dla $x \in R^q$. Wobec tego z ostatniej równości wynika, że wyrażenie

$$\frac{|g(\bar{x})-g(\bar{y})-B(\bar{x}-\bar{y})|}{|\bar{x}-\bar{y}|} \tag{10}$$

jest nie większe niż

$$\frac{|g(\bar{x})-g(\bar{y})|}{|\bar{x}-\bar{y}|}\,M\,|R(g(\bar{x}),g(\bar{y}))| \leqslant 2\frac{M}{m}\varepsilon.$$

dla $\bar{x},\bar{y} \in f(U_\varepsilon)$, $\bar{x} \neq \bar{y}$.

Udowodniliśmy, że dla każdego ε spełniającego nierówności $0 < \varepsilon < m/2$ istnieje takie otoczenie V punktu y_0, że wyrażenie (10) jest nie większe od $2\frac{M}{m}\varepsilon$ dla $\bar{x},\bar{y} \in V$, $\bar{x} \neq \bar{y}$. Stąd łatwo wynika, że dla dowolnego $\bar{\varepsilon} > 0$ istnieje otoczenie V punktu y_0, takie że wyrażenie (10) jest nie większe od $\bar{\varepsilon}$ dla $\bar{x},\bar{y} \in V$, $\bar{x} \neq \bar{y}$. A to dowodzi wobec twierdzenia 5, że g ma wszystkie pochodne pełne w punkcie y_0, co kończy dowód twierdzenia 6.

6. Na zakończenie podamy bez dowodów pewne twierdzenia o zbiorach mierzalnych i o całce Lebesgue'a. Dowodów nie podajemy ze względu na elementarność tego artykułu. Z drugiej strony wzmianka o tych twierdzeniach jest konieczna dla całości informacji, czym jest pochodna pełna i jakie daje korzyści. W twierdzeniach poniższych $\det f'(x)$ oznacza wyznacznik macierzy $f'(x)$, czyli jakobian.

TWIERDZENIE 7. *Jeżeli funkcja f jest w pełni różniczkowalna na zbiorze mierzalnym $X \subset R^q$, to*

$$\mu f(X) \leqslant \int_X |\det f'(x)|\,dx,$$

gdzie μ oznacza miarę Lebesgue'a, a całka jest rozumiana także w sensie Lebesgue'a. Gdy funkcja f jest ponadto odwracalna na X, to zamiast powyższej nierówności zachodzi równość.

TWIERDZENIE 8. *Jeżeli f jest funkcją odwracalną i w pełni różniczkowalną na zbiorze mierzalnym $X \subset R^q$, a F jest funkcją całkowalną w sensie Lebesgue'a na obrazie $f(X)$ tego zbioru, to*

$$\int_{f(X)} F(x)\,dx = \int_X |\det f'(x)|F[f(x)]\,dx.$$

Zauważmy, że twierdzenia powyższe, jak i wszystkie twierdzenia podane w tym artykule stają się fałszywe, gdy w nich pochodne pełne zastąpić przez tradycyjne pochodne cząstkowe. Gdyby zaś pochodne pełne zastąpić przez pochodne cząstkowe ciągłe, to twierdzenie 3 staje się tautologią, a wszystkie inne tracą sens, ponieważ pochodne cząstkowe ciągłe zawsze muszą być określone na zbiorach otwartych. Wprowadzenie pochodnych pełnych daje uogólnienie i jednocześnie uproszczenie twierdzeń o pochodnych cząstkowych ciągłych.

Wszystkie twierdzenia podane w tym artykule zostały opublikowane bez dowodów w Biuletynie Polskiej Akademii Nauk (*On partial derivatives*, XX, 11 (1972), str. 941-944). Dowody podane w tym artykule są publikowane po raz pierwszy.