Ekstrema lokalne funkcji

11.V. 2021

Często stosuje się operator pochodnej cząstkowej nie bezpośrednio do funkcji, ale do jej pochodnej cząstkowej. Będziemy w zastosowaniach używali pochodnych rzędu nie większego, niż 2, do nich też ograniczymy wzór Taylora, będący głównym narzędziem przy badaniu warunków wystarczających da istnienia ekstremów. Na przykład, dla funkcji dwu zmiennych $f(x, y)$ pochodna cząstkowa z $\frac{\partial}{\partial x}$ w punkcie (x_0, y_0) stosowana do $\frac{\partial f}{\partial x}$ daje pochodną drugiego rzędu: $\frac{\partial^2 f}{\partial x^2}(x_0, y_0)$ (inny zapis: $f''_{xx}(x_0, y_0)$, lub niepreferowany przeze mnie: $f_{xx}(x_0, y_0)$. Jest to druga pochodna w punkcie x_0 funkcji zmiennej x oznaczmy ją np. G, gdzie

$$G(x) = f(x, y_0).$$

Podobnie, dla

$$H(y) := f(x_0, y)$$

mamy

$$\frac{d^2}{dy^2}H(y_0) = \frac{\partial^2 f}{\partial y^2}(x_0, y_0).$$

Ale można też rozważać tzw. pochodne mieszane, np. gdy najpierw liczymy pochodną względem x, potem względem y. Kolejność operacji różniczkowania zapisuje się tak, jak w przypadku składania odwzorowań. We wspomnianym przypadku

$$\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) = \frac{\partial}{\partial y}\{\frac{\partial}{\partial x}f(x_0, y_0)\}.$$

Można mieć pewne zastrzeżenia do precyzji ostatniego zapisu - bo w celu wyliczenia pochodnej "zewnętrznej" należy tę "wewnętrzną" wyliczyć w punktach (x_0, y), gdzie y przebiega pewne otoczenie y_0. Najlepiej zobaczmy to na paru prostych przykładach.

Przykład 1. Dla $f(x, y) = xy^4 + 3x^2y^2$ mamy $f'_x = y^4 + 6xy^2$, $f'_y = 4xy^3 + 6x^2y$, więc $f''_{xy} = \frac{d}{dx}(4xy^3 + 6x^2y) = 4y^3 + 12xy$. Natomiast $f''_{yx} = \frac{d}{dy}(y^3 + 6xy^2) = 4y^3 + 12xy$. Jak za chwilę zobaczymy, to nie jest przypadek, że pochodne mieszane sąmrówne czyli operatory różniczkowe $\frac{\partial f}{\partial x}$ oraz $\frac{\partial f}{\partial y}$ są na odpowiednich klasach funkcji przemienne. Jednak następny przykład pokaże, że na ogół wcale tak nie musi być.

Przykład 2,

Pochodne mieszane zależą od kolejności różniczkowań dla następującej funkcji

$$f(x, y) = \begin{cases} xy\frac{x^2-y^2}{x^2+y^2}, & x^2 + y^2 > 0; \\ 0, & \texttt{dla} \ \ x=y=0. \end{cases} \tag{1}$$

Proszę sprawdzić, że w punkcie P o współrzędnych $(0, 0)$ mamy $f''_{xy}(P) = -1$, natomiast $f''_{yx}(P) = 1$.

Pochodnych wyższych rzędów raczej nie będziemy używali (mają one znaczenie przy przybliżaniu wartości f przy użyciu wzoru Taylora). Natomiast pochodne cząstkowe rzędu 2 liczy się dość często podczas badania ekstremów lokalnych, do czego zresztą za niedługo przejdziemy. Najpierw poznamy warunki gwarantujące równość pochodnych mieszanych. Podał je Hermann Schwarz (1843-1921), autor m. inn. nierówności dla iloczynu skalarnego i ciekawych twierdzeń z analizy zespolonej. Zacznijmy od definicji macierzy drugiej różniczki.

Definicja. Dla funkcji n zmiennych określonej w otoczeniu punktu P macierzą Hessego $H_f(P)$ nazywamy macierz kwadratową $n \times n$ postaci

$$H_f(P) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(P) & \dots & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(P) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(P) & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(P) & \dots & \dots & \frac{\partial^2 f}{\partial x_n^2}(P) \end{pmatrix}$$

Twierdzenie Schwarza o symetrii. Jeśli drugie pochodne mieszane są ciągłe w danym punkcie P, to są one równe. Macierz Hessego jest więc symetryczna w przypadku funkcji klasy C^2, czyli takich, dla których pochodne cząstkowe $\frac{\partial^2 f}{\partial x_j \partial x_k}$ są w tym punkcie ciągłe.

Każda macierz kwadratowa A wymiaru $n \times n$ definiuje formę kwadratową $Q : \mathbb{R}^n \ni x \to x^T A x$, co w przypadku $A = (a_{jk})_{j,k \in \{1,2,\dots,n\}}$ daje wzór $Q(x) = \sum_{j=1}^n \sum_{j=1}^n a_{jk} x_j x_k$

Formę kwadratową określoną przez macierz Hessego nazywamy drugą różniczką w danym punkcie i oznaczamy ją symbolem $d_P^2 f$. Jest to więc wielomian jednorodny zależny od n zmiennych rzeczywistych $t_1, \dots, t_n$, określony wzorem

$$d_P^2 f(t_1, t_2, \dots t_n) = \sum_{j=1}^n \sum_{k=1}^n \frac{\partial^2 f}{\partial x_j \partial x_k}(P) t_j t_k.$$

W zapisie bezargumentowym pierwsza różniczka miała postać

$$d_P f = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(P)\, dx_j,$$

gdzie dx_j reprezentuje odwzorowanie liniowe "rzut na j-tą oś",

$$(dx_j)(t_1, 2_2, \dots t_n) = t_j, \;\; dx_j : \mathbb{R}^n \to \mathbb{R}.$$

Iloczyn pary takich odworowań, zapisywany jako $dx_j dx_k$ nie jest już odwzorowaniem liniowym, ale wielomianem jednorodnym stopnia 2 (czyli formą kwadratową), takie odwzorowania $dx_j dx_k$ tworzą bazę wektorową przestrzeni form kwadratowych.

W pewnym sensie uzasadnieniem takiej, a nie innej definicji będzie (stanowiący istotę dowodu wzoru Taylora) następujący fakt, który nietrudno jest sprawdzić stosując regułę łańcucha (i wzór na pochodną iloczynu dla funkcji 1 zmiennej).

Lemat. Dla funkcji f klasy C^2 w otoczeniu punktu $P_0 = (x_0, y_0)$ i dla niezerowego wektora $\vec{\Delta} = (h, k)$ niech $P_s := P_0 + s\vec{\Delta}$. Wówczas druga pochodna w punkcie s z funkcji $\Phi(s) := f(P_0 + s\vec{\Delta}) = f(x_0 + sh, y_0 + sk)$ dana jest wzorem

$$\Phi''(s) = f''_{xx}(P_s)h^2 + 2f''_{xy}(P_s)hk + f''_{yy}(P_s)k^2.$$

Jest to więc wartość drugiej różniczki w punkcie P_s na wektorze o współrzędnych (h, k).

Stosując teraz wzór Taylora z resztą postaci Lagrange'a dla funkcji Φ jednej zmiennej i powyższy lemat otrzymujemy następujące twierdzenie:

Twierdzenie (Wzór Taylora). Dla funkcji klasy C^2 w otoczeniu odcinka o końcach $P_0, P_0 + \Delta$ istnieje taki punkt pośredni $P_\lambda = P_0 + \lambda\Delta$, gdzie $0 < \lambda < 1$, że

$$f(P_0 + \Delta) = f(P_0) + d_{P_0} f(\Delta) + \frac{1}{2} d_{P_\lambda}^2 f(\Delta).$$

Ten "wzór Taylora z drugą różniczką" -da nam warunek wystarczający na istnienie ekstremum lokalnego w punkcie P_0

Definicja Mówimy, że funkcja $f : D \subset \mathbb{R}^n \to \mathbb{R}$ ma ekstremum lokalne w punkcie P_0, gdy jej przyrost $f(P_0 + \Delta) - f(P_0)$ ma stały znak w pewnym otoczeniu punktu P_0 (czyli dla $\|\Delta\|$ dostatecznie małych). Jest to np. silne maksimum lokalne, gdy $f(P_0)$ jest silnie większa od wartości f w pewnym sąsiedztwie punktu P_0, czyli gdy

$$\exists_{\delta>0}\forall_{\Delta\in\mathbb{R}^n}\,(0 < \|\Delta\| < \delta) \Rightarrow f(P_0) > f(P_0 + \Delta).$$

Minimum lokalne dla funkcji f odpowiada maksimum lokalnemu dla $-f$. Aby podać warunek konieczny na istnienie ekstremum lokalnego wystarczy zauważyć, że jeśli U jest otoczeniem punktu P_0, to $\{s \in \mathbb{R} : P_s \in U\}$ jest otoczeniem zera, czyli zawiera jakiś przedział $(-\epsilon, \epsilon)$. Ponadto jeśli f ma ekstremum lokalne w P_0, to używana powyżej funkcja $\Phi(s) := f(P_s)$ ma ekstremum lokalne w zerze. Dzięki twierdzeniu Fermata -warunkiem koniecznym jest tu $\Phi'(0) = 0$, czyli z powyższych razważań, wszystkie pochodne kierunkowe f w punkcie ekstremum lokalnego muszą być zerowe. Biorąc pochodne w kierunku wektorów z bazy kanonicznej (czyli w kierunku np, j- tej osi) otrzymamy następujący

Warunek konieczny. Jeśli funkcja f ma pochodne cząstkowe w punkcie P_0, w którym osiąga ona ekstremum lokalne, to te pochodne cząstkowe muszą być równe zero. Innymi słowy, szukając ekstremów lokalnych wystarczy zbadać punkty, w których gradient badanej funkcji jest wektorem zerowym.

Powstaje przy okazji pytanie, czy z istnienia minimum lokalnego wzdłuż każdej prostej przechodzącej przez dany punkt wynika, że jest tam minimum lokalne? Dla minimu absolutnego (=wartości najmniejszej na całej dziedzinie przeciętej z prostymi) odpowiedź by była pozytywna. Ale nie dla ekstremów lokalnych -warunku wystarczającego dla 1 zmiennej nie możemy więc tą metodą przenieść na sytuację wielowymiarową. Oto przykład:

Przykład 1. Niech $n = 2, P_0 = (0, 0), f(x, y) = (x^2-y)(4x^2-y)$. Na prostych o równaniu $y = \alpha x$ mamy $f(x, \alpha x) = 4x^4-5\alpha x^3+\alpha x^2 = x^2(x^2-5\alpha x+\alpha^2)$ i przy $x \to 0$ granica wyrażenia w nawiasie jest dodatnia, więc w pewnym (o rozmiarze zależnym od α) sąsiedztwie zera jest $f(x, \alpha x) > 0$, jest więc minimum lokalne "w kierunku wektora $(1, \alpha)$". Pozostaje jeszcze prosta pionowa o równaniu $x = 0$, tam $f(0, y) = y^2$ też ma minimum w zerze. Ale w każdym otoczeniu zera w $\mathbb{R}^2$ są punkty, w których f jest ujemna -to punkty z obszaru nad parabolą $y = x^2$ a poniżej $y = 4x^2$, czyli zbiór $\{(x, y) : x^2 < y < 4x^2\}$, więc ta funkcja f nie ma ekstremum lokalnego w zerze.

Aby podać warunek wystarczający musimy więc odwołać się do metod wielowymiarowych (choć wzór Taylora, z którego skorzystamy był uzyskany "metodą 1-wymiarową" -brania wartości f w punktach P_s, czyli wzdłuż prostej).

Definicja. Mówimy, że macierz kwadratowa $A \in M_{n\times n}$ o wyrazach a_{jk} jest dodatnio określona, co zapisujemy symbolem $A >> 0$ gdy dodatnia na $\mathbb{R}^n \setminus \{0\}$ jest związana z nią forma kwadratowa (oznaczana Q_A), czyli gdy

$$\forall_{x=(x_1,\dots,x_n)\neq 0}\ Q_A(x) := x^T A x = \sum_{j,k=1}^{n} a_{jk}x_j x_k > 0.$$

Twierdzenie (Warunek wystarczający). Załóżmy, $f : D \subset \mathbb{R}^n \to \mathbb{R}$ jest klasy C^2 w pewnym otoczeniu punktu $P_0 \in D$. Wówczas f ma ekstremum lokalne w punkcie P_0 jeśli zachodzi koniunkcja obydwu warunków:

1. W punkcie P_0 gradient f jest zerowy
2. Macierz Hessego $H_{P_0}f$ pomnożona przez jedną z wartości $C \in \{-1, 1\}$ jest dodatnio określona.

Przy tym, dla $C = 1$, czyli gdy $H_{P_0}f >> 0$ -funkcja f ma ściśle minimum lokalne w P_0, zaś gdy $(-H_{P_0})f >> 0$, to jest tam ściśle maksimum lokalne.

10.1 Przykład ekstremów

Szukamy ekstremów lokalnych funkcji określonej dla (x, y, z) takich, że $x > 0, y > 0, z > 0$

$$f(x,y,z) = x + \frac{y^2}{4x} + \frac{z^2}{y} + \frac{2}{z}.$$

Najpierw pochodne cząstkowe 1. rzędu:

$$f'_x = 1 - \frac{y^2}{(2x)^2},\ f'_y = \frac{2y}{4x} - \frac{z^2}{y^2},\ f'_z = \frac{2z}{y} - \frac{2}{z^2}.$$

Przyrównując gradient do zera widzimy, że $y = 2x, z = y, z = 1$, więc jedynym punktem z tak określonej dziedziny, w którym może zaistnieć ekstremum lokalne jest punkt stacjonarny $P_0 = (\frac{1}{2}, 1, 1)$. Aby zbadać warunek wystarczający, liczymy współczynniki macierzy Hessego.

$$f"_{xx} = \frac{2y^2}{4x^3}, f"_{yy} = \frac{2}{4x} + \frac{2z^2}{y^3}, f"_{zz} = \frac{2}{y} + \frac{4}{z^2}, f"_{yz} = -\frac{2z}{y^2}, f"_{xz} = 0, f"_{xy} = -\frac{y}{2x^2},$$

resztę uzyskujemy dzięki symetrii (bo nawet $f \in C^\infty$. Macierz Hessego w punkcie stacjonarym, to

$$H = \begin{pmatrix} 4 & -2 & 0 \\ -2 & 3 & -2 \\ 0 & -2 & 6 \end{pmatrix}$$

Wszystkie minory kątowe główne są tu dodatnie: $m_1(H) = 4, m_2(H) = 8, m_3 = 6 \cdot 8 - 2 \cdot 8 > 0$. W punkcie P_0 mamy więc minimum lokalne.