23 Wzór Taylora, ekstrema lokalne

Przypomnijmy, że dla funkcji $f : D \subset \mathbb{R}^n \to \mathbb{R}$ klasy C^2 zdefiniowaliśmy jej drugą różniczkę w punkcie $P \in D$ jako formę kwadratową:

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

Macierz $[\frac{\partial^2 f}{\partial x_k \partial x_j}(P)]_{j,k \leqslant n}$ nazywamy macierzą Hessego i oznaczamy ją $H_P f$.

Ta forma kwadratowa pojawi się jako odpowiednik drugiej pochodnej funkcji 1 zmiennej we wzorze Taylora. Wzór ten uzyskamy zawężając przebieg zmiennej niezależnej do prostej przechodzącej przez zadany punkt P_0, leżącej w kierunku pewnego ustalonego wektora $\Delta = (h_1, \ldots, h_n) \in \mathbb{R}^n$. (Nie piszę już strzałki nad Δ). Każdy punkt z tej prostej możemy opisać parametrem s przebiegającym zbiór $\mathbb{R}$ w następujący sposób:

$$P_s := P_0 + s\Delta.$$

Wówczas zbiór $\{P_s : 0 \leqslant s \leqslant 1\}$ będzie odcinkiem o końcach P_0 oraz $P_0 + \Delta$. Będziemy jedynie zakładali, że ten odcinek zawarty jest w dziedzinie D naszej funkcji i w punktach tego odcinka istnieją pochodne cząstkowe rzędu 1 oraz rzędu 2 z funkcji f i są one ciągłe. Z Reguły Łańcucha wiemy, że ponieważ odwzorowanie liniowe $L : \mathbb{R} \ni s \to s\Delta \in \mathbb{R}^n$ jest liniową częścią (a więc i różniczką) odwzorowania afinicznego $s \mapsto P_s$, to pierwsza różniczka z funkcji złożonej $\Phi(s) := f(P_s)$ jest złożeniem $d_{P_s} f \circ L$. Pochodna $\Phi'(s)$ tej funkcji złożonej, to wartość jej różniczki na liczbie 1 (tu 1 jest jedno-elementową bazą przestrzeni wektorowej $\mathbb{R}$), czyli iloczyn skalarny gradientu f w punkcie P_s oraz wektora Δ, jest to też pochodna kierunkowa f w punkcie $P(s)$, równa $d_{P_s} f(\Delta)$. Współrzędnymi wektora Δ są $(h_1, \ldots, h_n)$, więc

$$\Phi'(s) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(P_s) h_j. \tag{1}$$

Licząc teraz drugą pochodną, czyli $\Phi''(s)$, liczymy sumę pochodnych poszczególnych składników w (1), gdzie h_j -są to pewne stałe (nie zależą od s). Pochodne z $\Psi_j(s) := \frac{\partial f}{\partial x_j}(P_s)$ wyrażają się on podobnym do (1) wzorem, tylko dla już ustalonego j musimy wybrać inną nazwę na indeksy sumowania- niech to np. będzie k. tak więc

$$\frac{d}{ds} \left\{ \frac{\partial f}{\partial x_j}(P_s) h_j \right\} = \sum_{k=1}^{n} \frac{\partial^2 f}{\partial x_k \partial x_j}(P_s) h_j h_k.$$

Sumując teraz prawe strony ostatniej równości po $1 \leqslant j \leqslant n$ otrzymujemy wzór na drugą pochodną $\Phi^{''}(s)$, jak w poniższym lemacie:

Lemat. Dla funkcji f klasy C^2 w otoczeniu odcinka $\{P_s : s \in [0, 1]\}$, zdefiniowanego powyżej przez wektor $\Delta = (h_1, \ldots, h_n)$ mamy

$$\frac{d^2 \Phi}{ds^2} = \sum_{k=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_k \partial x_j}(P_s) h_j h_k = d_{P_s}^2 f(\Delta). \tag{2}$$

.

W szczególności, dla funkcji dwu zmiennych x, y i dla wektora $\Delta = (h, k)$ otrzymujemy podany na ostatnim wykładzie wzór

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

W przypadku ogólnym n zmiennych możemy teraz bez trudu zapisać wzór Taylora z resztą postaci Lagrange'a dla naszej funkcji złożonej Φ: Istnieje punkt λ leżący na odcinku otwartym $(0, 1)$, dla którego

$$\Phi(1) = \Phi(0) + \Phi'(0) \cdot (1 - 0) + \Phi^{''}(\lambda) \frac{(1-0)^2}{2!} \tag{3}$$

Stosując teraz wzory (1), (2) 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

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

Odpowiednik tego wzoru zawierający różniczki wyższych rzędów sformułujemy wkrótce, lecz nie będziemy go raczej stosować, w odróżnieniu od powyższego "wzoru Taylora z drugą różniczką" -który da nam warunek wystarczający na istnienie ekstremum lokalnego w punkcie P_0

23.1 Ekstrema lokalne

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* (inna nazwa: *ścisłe*) *maksimum lokalne*, gdy $f(P_0)$ jest silnie większa od wartości f w pewnym sąsiedztwie punktu P_0, czyli gdy

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

Oczywiście, minimum lokalne dla funkcji f odpowiada maksimum lokalnemu dla funkcji $-f$. Aby podać warunek konieczny na istnienie ekstremum lokalnego wystarczy zauważyć, że jeśli U jest otoczeniem punktu P_0, to zbiór $\{s\in\mathbb{R}: P_s\in U\}$ 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 $\Phi'(0)=0$, więc wszystkie pochodne kierunkowe f w punkcie ekstremum lokalnego muszą być zerowe. Biorąc pochodne w kierunku wektorów z bazy kanonicznej (= w kierunku 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 $(1,\alpha)$". Na prostej pionowej o równaniu $x=0$, też $f(0,y)=y^2$ ma minimum w zerze. Ale w każdym otoczeniu zera w $\mathbb{R}^2$ są punkty, w których f jest ujemna -to punkty z obszaru między parabolami $y=x^2$, $y=4x^2$, czyli zbiór $\{(x,y): x^2<y<4x^2\}$. Ta funkcja nie ma więc 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

$$\forall_{x=(x_1,\dots,x_n)\neq 0}\; Q_A(x) := x^TAx = \sum_{j,k=1}^n a_{jk}x_jx_k > 0.$$

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. Albo macierz Hessego $H_{P_0} f$ albo macierz $-H_{P_0} f$ jest dodatnio określona.

Przy tym gdy $H_{P_0} f >> 0$, to funkcja f ma ścisłe minimum lokalne w P_0, zaś gdy $(-H_{P_0}) f >> 0$, to jest tam ścisłe maksimum lokalne.

Dowód. Ścisłe ekstremum lokalne wystąpi w punkcie $P_0 \Leftrightarrow$ gdy znak przyrostu $f(P_0 + \Delta) - f(P_0)$ będzie stały i niezerowy w pewnym sąsiedztwie tego punktu. Ze względu na zerowanie się $d_{P_0} f$, wzór Taylora mówi nam, że ten przyrost jest równy $\frac{1}{2} d^2_{P_\lambda}(\Delta)$, gdzie dla pewnego $\lambda \in (0, 1)$ punkt P_λ leży na otwartym odcinku o końcach $P_0, P_0 + \Delta$. Wystarczy więc wiedzieć, że znak wartości tej drugiej różniczki na wektorze "przyrostu argumentu": Δ jest stały w punktach z pewnego sąsiedztwa P_0. Zbadajmy przypadek gdy $H_{P_0} f >> 0$, czyli warunek wystarczający dla minimum.

Przykład 1. pokazuje, że nie wystarczy zrobić tego przy ustalonym Δ, choć to by było łatwiejsze. Faktycznie, nierówność ostra, która zachodzi dla funkcji ciągłej w P_0 -propaguje się na pewne otoczenie P_0. Natomiast (w przypadku C=1) wystarczy wykazać dodatnią określoność macierzy Hessego $H_P f$ dla wszystkich punktów P z pewnego otoczenia punktu P_0. Tu można działać na dwa sposoby. W sytuacji przestrzeni Banacha stosuje się sposób z infimum formy po sferze jednostkowej (które musi być dołączone do założeń jako wzmocnienie warunku 2, ale dla $n < \infty$ wynika to ze zwartości sfery). Moim zdaniem, łatwiej jest skorzystać z kryterium Sylvestera. Dotyczy ono tzw. minorów kątowych głównych $m_k(A)$ macierzy A . Definiujemy te minory jako wyznaczniki:

$$m_k(A) := \det([a_{ij}]_{1 \leqslant i,j \leqslant k}),$$

gdzie $k = 1, 2, \ldots, n$.

Kryterium Sylvestera mówi, że dla macierzy symetrycznej $n \times n$ mamy równoważność:

$$A >> 0 \Leftrightarrow \forall_{k \in \{1,2,\ldots,n\}}\, m_k(A) > 0.$$

Ze względu na definicję macierzy Hessego, definicję wyznacznika i założenie o klasie C^2, mamy ciągłą zależność minorów $m_k(H_P f)$ od punktu P w pewnym otoczeniu punktu P_0. Jeśli więc $H_{P_0} >> 0$, to z dodatniości jej wszystkich minorów kątowych głównych wynika, że dla każdego $k \leqslant n$ znajdziemy takie otoczenie U_k punktu P_0, w którym dla $P \in U_k$ mamy $m_k(H_P f) > 0$. Przecięcie tych otoczeń nadal jest otoczeniem, nazwijmy je U. Dla $P \in U = U_1 \cap U_2 \cap \cdots \cap U_n$ mamy $\forall_k\, m_k(H_P f) > 0$- czyli dodatnią okresloność macierzy Hessego $H_P f$. To dowodzi twierdzenia w przpadku minimum lokalnego.

Gdy zaś $-H_{P_0} f >> 0$, to minory kątowe główne $m_k := m_k(H_P f)$ zmieniają znak w sposób alternujący, poczynając od ujemnego m_1, dodatniego m_2, ujemne m_3 itd. Wiąże się to ze wzorem Cauchy'ego który mówi,że dla macierzy $k \times k$ wyznacznik jest funkcją jednorodną stopnia k: $\forall_{C \in \mathbb{R}} \det(CA) = C^k \det(A)$. Ponadto $-H_P f$ jest macierzą Hessego dla $-f$, więc otrzymane tu minimum lokalne dla $-f$ oznacza maksimim lokalne dla f. $\square$.

Uwaga: Jest to warunek wystarczający, ale nie konieczny. Może się zdarzyć, jak dla funkcji 1 zmiennej, że np. druga różniczka jest równa zero. Wtedy można próbować różniczek wyższego rzędu, można też wykazać istnienie (słabego) ekstremum lokalnego dowodząc, że forma kwadratowa, jaką jest druga różniczka- pozostaje stałego znaku (np.$\geqslant 0$) w pewnym otoczeniu punktu P_0. Kryterium Sylvestera nie stosuje się do słabych nierówności- to widać na przykladzie macierzy diagonalnych; są one nieujemnie określone (mówi się też "półokreślone dodatnio", gdy wyrazy z przekątnej są $\geqslant 0$. Dodatnio - gdy są > 0. Tu $m_k(A) = a_{11} \cdot a_{22} \cdot \ldots \cdot a_{kk}$, bo tym razem $a_{ij} = 0$ dla $i \neq j$. Pojawienie się zera dla $k = m < n$ w pewnym sensie "kasuje" już wszelkie informacje o znaku

a_{jj} dla $j > m$. Jest jednak twierdzenie -wkw na dodatnią półokreśloność: Tym warunkiem jest nieujemność wszystkich tzw. minorów głównych. Zainteresowanych odsyłam np. do skryptu A.Turowicza (Teoria Macierzy, Wyd. AGH). Dalszy problem - to brak propagacji na otoczenia dla słabych nierówności - tam trzeba to sprawdzać "ręcznie" w całym otoczeniu. Czasami da się to zrobić. Na przykład, dla wielomianów jednorodnych stopnia 2.

Zauważmy najpierw, że dla odwzorowań liniowych L mieliśmy w każdym punkcie P pierwszą różniczkę równą L , niezależnie od P, więc druga róznniczka byłą równa zero. Dla wielomianu jednorodnego $W(x_1, x_2, \ldots, x_n)$ stopnia 2 mamy z kolei resztę związaną z 3. różniczką - zerową we wzorze Taylora, czyli $W = \frac{1}{2} d_P^2 W$ -znów niezależnie od punktu P. Półokreśloność będzie (a właściwie, może) wtedy oznaczać stałość w kierunku jakiejś prostej, badź na jakiejś niezerowej podprzestrzeni (patrz rozkład Sylvestera formy kwadratowej). Stosując taki rozkład moglibyśmy zresztą propagować taką półokreśloność na otoczenie punktu P_0 -ale pod jednym dodatkowym warunkiem: że wymiar przestrzeni zerowej $d_P^2 f$ -czyli zbioru tych wektorów $v \in \mathbb{R}^n$, dla których $(d_P^2 f)(v) = 0$ będzie stały. Ponieważ wymiar części dodatniej (powiedzmy, równy d dla $P = P_0$) -pozostanie większy lub równy d w pewnym sąsiedztwie, bo restrykcja formy do tej podprzestrzeni dodatniej pozostanie dodatnio określona w otoczeniu. W przypadku. gdyby wymiar podprzestrzeni zerowej się zmniejszył- może się on zmniejszyć -poprzez wpadnięcie jakichś z wektorów z przestrzeni zerowej (przy $P = P_0$) do "podprzestrzeni ujemnej" dla tej formy liczonej w pobliskich punktach P. Inną metodę radzenia sobie z półokreślonością drugiej różniczki poznamy przy okazji ekstremów warunkowych.

23.2 Przykład szukania ekstremów

Szukamy ekstremów lokalnych funkcji określonej dla $x > 0, y > 0, z > 0$ wzorem

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

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.