

Jak w poprzednich rozdziałach, badamy funkcję $f : \mathbb{R}^n \supset D_f \rightarrow \mathbb{R}$ zmiennych (x_1, \dots, x_n) . Zakładamy o niej, że jest dwukrotnie różniczkowalna (chyba, że jest napisane inaczej). Tak jak w przypadku funkcji jednej zmiennej, samo otrzymanie wzoru funkcji wielu zmiennych jako odpowiedzi na pytanie o związek pomiędzy jakimiś wielkościami (np. ekonomicznymi) nie zawsze daje od razu odpowiedzi na wszystkie pytania. Musimy się zawsze zastanowić, jakie informacje o funkcji możemy uzyskać na podstawie tego wzoru. Typowym zagadnieniem jest optymalizacja jakiejś wielkości, czyli poszukiwanie wartości argumentów dla których wartość funkcji jest najmniejsza lub największa (minimalizacja kosztu, maksymalizacja zysku itp.). Matematycznie sprowadza się to do szukania ekstremów funkcji. Umiemy to już robić dla funkcji jednej zmiennej, więc możemy się domyślić, że pochodne będą grały kluczową rolę w wyznaczaniu tych ekstremów. Jednakże, w przypadku wielu zmiennych, rozwiązanie będzie nieco bardziej skomplikowane.

I. „Monotoniczność”, gradient i pochodna kierunkowa

Pytanie o monotoniczność w przypadku funkcji wielu zmiennych nie ma sensu.

Przykład $f(x, y) = x + 2y$ w okolicy punktu $(0, 0)$. Nasza intuicja z funkcji jednej zmiennej mówi - jeśli wartości funkcji rosną, gdy „przesuwamy się w prawo” to funkcja jest rosnąca. Ale w tej sytuacji - jeśli na płaszczyźnie Oxy „przesuwamy się w prawo i w górę z tą samą prędkością” (tj. wzdłuż wektora $(1, 1)$) wartości funkcji rosną. A jeśli „przesuwamy się w prawo i w dół z tą samą prędkością” (tj. wzdłuż wektora $(1, -1)$) wartości funkcji maleją. Zatem nie możemy powiedzieć, że rozważana funkcja jest rosnąca lub malejąca - a na pewno już nie jest ona stała.

Możemy się naturalnie zastanawiać czy funkcja jest rosnąca lub malejąca wzdłuż pewnych prostych. Czasem też warto się zastanowić, w którą stronę funkcja najszybciej zmienia wartość. Do tego przydadzą nam się dwa nowe pojęcia - gradient i pochodna kierunkowa.

Definicja 1. Gradientem funkcji f w punkcie $a = (a_1, \dots, a_n)$ nazywamy wektor, którego współrzędnymi są pochodne cząstkowe funkcji f w tym punkcie. Zapisujemy:

$$\nabla_f(a) = (f'_{x_1}(a), f'_{x_2}(a), \dots, f'_{x_n}(a)).$$

Dopuszczalną notacją jest też $\text{grad } f(a)$.

Gradient to wektor wskazujący, w którą stronę (dokładnie, wzdłuż jakiej prostej) funkcja najszybciej zmienia wartość. Długość gradientu odpowiada wzrostowi wartości tej funkcji na jednostkę długości.

Przykład Załóżmy, że zyski firmy z nakładów na reklamę (x) i na dział obsługi klienta (y) wyrażają się wzorem $f(x, y) = \sqrt{x^2 + 5y}$. Obecne nakłady wynoszą $(2, 1)$. W jakich proporcjach firma powinna zwiększać nakłady na te dwie dziedziny działalności, by jej zysk z tego zwiększenia nakładów był jak największy (przy założeniu, że wzrost nakładów będzie niewielki)?

Za rozstrzygnięcie, czy funkcja rośnie, czy maleje w kierunku danej półprostej odpowiada pochodna kierunkowa.

Definicja 2. Pochodną kierunkową funkcji f w punkcie $a = (a_1, \dots, a_n)$ w kierunku wektora $v = (v_1, \dots, v_n)$ nazywamy granicę (o ile istnieje i jest skończona):

$$f'_v(a) = \lim_{h \rightarrow 0^+} \frac{f(a + h \cdot v) - f(a)}{h}.$$

Jak widać, definicja jest podobna do zwykłej pochodnej, ale wzrost wartości mierzymy tylko w zadanym kierunku. Nic dziwnego, że zachodzi poniższe twierdzenie:

Twierdzenie 1. Jeśli funkcja f jest różniczkowalna w kuli otwartej o promieniu r i $f'_v(x) > 0$ dla każdego x z tej kuli, to wewnątrz tej kuli f rośnie w kierunku wektora v . Jeśli funkcja f jest różniczkowalna w kuli otwartej o promieniu r i $f'_v(x) < 0$ dla każdego x z tej kuli, to wewnątrz tej kuli f maleje w kierunku wektora v .

Pochodną kierunkową najłatwiej obliczyć z następującego twierdzenia:

Twierdzenie 2. *Jeśli w punkcie a funkcja f ma ciągłe pochodne cząstkowe to:*

$$f'_v(a) = \langle \nabla f(a), v \rangle.$$

Przykład Dla tego samego przykładu, co poprzednio - jeśli nakłady na reklamę zmniejszymy i połowę z tak pozyskanych oszczędności przeznaczymy na dział obsługi klienta, to zyski firmy wzrosną, czy zmaleją?

Zwróćmy uwagę, że pochodne cząstkowe są też pochodnymi kierunkowymi - w kierunkach zadanych przez osie układu współrzędnych.

Podobnie można wzdłuż prostych badać wklęsłość i wypukłość.

Przykład Jeśli U jest funkcją użyteczności koszyka dóbr (x_1, \dots, x_n) dla konsumenta, to spełnia ona prawo Gossena (malejącej użyteczności krańcowej) wtedy i tylko wtedy gdy jej drugie pochodne jednorodne (tj. „niemieszane”) są ujemne (tj. $U''_{x_1x_1} < 0, U''_{x_2x_2} < 0, \dots$)

II. Ekstrema lokalne

Definicja 3. *Funkcja f ma w punkcie $a = (a_1, a_2, \dots, a_n)$ maksimum lokalne, jeżeli istnieje takie otoczenie punktu a , że każde $x \neq a$ z tego otoczenia spełnia zależność $f(x) < f(a)$. Dla $a \in \mathbb{R}^n$ możemy ten warunek formalnie zapisać:*

$$\exists \epsilon > 0 \forall x \in U_\epsilon(a) \setminus \{a\} f(x) < f(a).$$

Funkcja f ma w punkcie $a = (a_1, a_2, \dots, a_n)$ minimum lokalne, jeżeli istnieje takie otoczenie punktu a , że każde $x \neq a$ z tego otoczenia spełnia zależność $f(x) > f(a)$. Dla $a \in \mathbb{R}^n$ możemy ten warunek formalnie zapisać:

$$\exists \epsilon > 0 \forall x \in U_\epsilon(a) \setminus \{a\} f(x) > f(a).$$

Jak w przypadku jednej zmiennej, wszystkie minima i maksima nazywamy *ekstremami* funkcji.

Jeśli w powyższych zdaniach możemy uzyskać tylko słabe nierówności to mówimy o *słabym minimum/maksimum lokalnym*.

Wobec tego, że w otoczeniu ekstremum, wzdłuż żadnej prostej przechodzącej przez ekstremum, funkcja nie będzie rosnąca ani malejąca, nie powinno zaskakiwać poniższe twierdzenie:

Twierdzenie 3. *(Warunek konieczny istnienia ekstremum lokalnego) Jeśli funkcja f ma ekstremum w punkcie a i wszystkie jej pochodne w tym punkcie istnieją to $\nabla f(a) = (0, 0, \dots, 0)$. Punkty spełniające to równanie nazywamy stacjonarnymi lub krytycznymi.*

Funkcja zatem może mieć ekstrema tylko w punktach stacjonarnych i w punktach, w których choć jedna pochodna cząstkowa nie istnieje. Nie każdy punkt stacjonarny jest jednak ekstremum lokalnym.

Przykład $f(x, y) = xy$, $f(x, y) = |x| + |y|$.

Zanim sformułujemy warunek wystarczający istnienia ekstremum lokalnego, potrzebujemy sformalizowania pojęcia przywoływanej już macierzy Hessego, czyli hesjanu.

Definicja 4 (Hesjan). *Dla funkcji f zmiennych x_1, \dots, x_n , dwukrotnie różniczkowalnej w punkcie $a = (a_1, \dots, a_n)$ macierzą Hessego lub hesjanem w punkcie a nazywamy macierz złożoną z jej drugich pochodnych cząstkowych, czyli:*

$$H_f(a) = \begin{bmatrix} f''_{x_1x_1}(a) & f''_{x_1x_2}(a) & \dots & f''_{x_1x_n}(a) \\ f''_{x_2x_1}(a) & f''_{x_2x_2}(a) & \dots & f''_{x_2x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ f''_{x_nx_1}(a) & f''_{x_nx_2}(a) & \dots & f''_{x_nx_n}(a) \end{bmatrix}.$$

Warto zauważyć, że zgodnie z twierdzeniem o równości pochodnych mieszanych, jeśli tylko te pochodne są ciągłe w a to hesjan jest macierzą symetryczną.

Twierdzenie 4 (Warunek wystarczający istnienia ekstremum lokalnego). *Niech f będzie funkcją dwukrotnie różniczkowalną, której drugie pochodne są ciągłe w otoczeniu U punktu a . Załóżmy, że w punkcie a spełniony jest warunek konieczny istnienia ekstremum ($\nabla_f(a) = (0, 0, \dots, 0)$).*

a) *Jeśli hesjan $H_f(a)$ jest dodatnio określony to funkcja f osiąga minimum lokalne w punkcie a .*

b) *Jeśli hesjan $H_f(a)$ jest ujemnie określony to funkcja f osiąga maksimum lokalne w punkcie a .*

c) *Jeśli hesjan $H_f(a)$ jest nieokreślony, funkcja f nie osiąga ekstremum w punkcie a .*

Jak widać, twierdzenie to nie rozstrzyga zagadnienia istnienia ekstremum we wszystkich przypadkach (np. półokreśloności). W takich wypadkach trzeba badać punkt a jako „kandydata na ekstremum” innymi metodami - jednak zdarza się to na tyle rzadko, że nie będziemy tego typu sytuacji rozważać na wykładzie (ani na egzaminie).

Przykład $f(x, y, z) = x^3 + xy + y^2 - 2zx + 2z^2 + 3y - 1$.

III. Metoda najmniejszych kwadratów

Podstawowym zastosowaniem umiejętności znajdowania ekstremów wielu zmiennych, użytecznym we wszelkich badaniach statystycznych jest tzw. metoda najmniejszych kwadratów.

Założmy, że dla jakiegoś procesu (fizycznego, ekonomicznego itp.) mamy model matematyczny, z którego wynika, że zależność między dwiema wielkościami jest funkcją pewnego typu (liniową, kwadratową, logarytmiczną itp.), jednak nie jesteśmy pewni, jakie dokładnie parametry ma taka funkcja. Można próbować wyznaczyć je doświadczalnie, poprzez pewną liczbę pomiarów danych wielkości. W idealnym świecie, takie pomiary powinny się rozłożyć wzdłuż wykresu poszukiwanej funkcji i rozwiązanie układu kilku równań liniowych wystarczyłoby, by znaleźć dokładny wzór szukanej zależności. Niestety, praktycznie nigdy się tak nie zdarza - ze względu na idealizację konieczną do konstrukcji modelu, a także brak pełnej informacji, występowanie „szumu”, czyli niemożliwość zachowania stałych warunków doświadczenia oraz zwykłe błędy i niedokładności pomiarowe - wyniki doświadczeń będą tworzyły raczej nieregularną „chmurę” punktów niż przybliżenie jakiegoś prostego wykresu. Metoda najmniejszych kwadratów pozwala nam dobrać parametry naszej funkcji tak, by w najlepszy sposób przybliżała ona wyniki doświadczalne. Taką funkcję możemy zastosować do badania modelu lub też do przewidywania wyników przyszłych pomiarów.

Metoda opiera się na tzw. postulacie Legendre’a: w wypadku badania serii tak samo (w sensie rachunku prawdopodobieństwa) niedokładnych pomiarów, najbardziej prawdopodobną „prawdziwą wartością” poszukiwanej wielkości, jest taka, od której obliczone odchylenia tych wyników, po podniesieniu do kwadratu i zsumowaniu dają wielkość najmniejszą z możliwych. Stwierdzenie to jest poparte wieloma statystycznymi i praktycznymi „dowodami”, jednak nie jest twierdzeniem w sensie matematycznym ani nie wynika z żadnej matematycznej teorii - raczej z matematycznej intuicji. Jest to do dziś główna metoda modelowania statystycznego i ekonometrycznego - głównie dlatego, że bardziej dokładne metody nie dają wiele lepszych rezultatów, a działają zawsze kosztem o wiele bardziej skomplikowanych obliczeń.

Mimo dyskusyjnych podstaw, metoda daje bardzo potężne wyniki. Pierwsze znane jej spektakularne zastosowanie jest dziełem jednego z najwybitniejszych matematyków wszech czasów Carla Friedricha Gaussa (znanego współczesnym jako *książę matematyków*). W 1801 roku astronomowie odkryli asteroidę Ceres. Zbadanie jej trajektorii było bardzo ważne dla badań Układu Słonecznego. Niestety, można ją było obserwować tylko przez 40 dni, po których zniknęła po drugiej stronie Słońca. By móc ją śledzić ponownie (przy bardzo niedoskonałych ówczesnych narzędziach) trzeba było przewidzieć z dużą dokładnością, w którym miejscu i kiedy wynurzy się zza Słońca. Wiadomo było, że się porusza zgodnie z równaniami Keplera ruchów planet, ale konkretne parametry tego ruchu były tajemnicą. Jedynym, któremu udało się poprawnie rozwiązać zagadnienie był właśnie

24-letni wówczas Gauss. Użył do tego celu właśnie metody najmniejszych kwadratów. Asteroida pojawiła się dokładnie tam, gdzie wskazał.

Po reklamie, czas na wyjaśnienie, jak ta metoda działa.

Dane jest n punktów $(x_n, y_n) \in \mathbb{R}^2$ (dane pomiarowe, wyniki doświadczeń - metodę da się uogólnić na powiązanie ze sobą większej liczby zmiennych objaśniających, ale to nie zmienia znacząco sposobu postępowania). Szukamy funkcji różniczkowalnej postaci $f_{a_1, \dots, a_n}(x)$ zależnej od szukanych parametrów a_1, \dots, a_n , której wykres leży jak najbliżej danych pomiarowych (może to być np. $f_{a,b,c,d}(x) = ax^2 + bx + x^c + \sin(x+d)$). Rozważamy wyrażenie:

$$S(a_1, \dots, a_n) = \sum_{i=1}^n (y_i - f_{a_1, \dots, a_n}(x_i))^2.$$

Zagadnienie nasze sprowadza się do znalezienia minimum funkcji S . Generalnie, minimum takie istnieje, jest jedno i jest rozwiązaniem układu równań liniowych pochodzącego z przyrównania pochodnych cząstkowych S do 0 (zauważmy, że S jest funkcją a_1, \dots, a_n , a nie x, y).

W statystyce używa się tej metody przede wszystkim do przybliżania wyniku eksperymentu prostą o równaniu $y = ax + b$ i sprawdzaniu korelacji zmiennych za pomocą współczynnika a . Dlatego na kursie statystyki prawdopodobnie otrzymają Państwo gotowy wzór na obliczanie współczynników a i b w takiej sytuacji. Jednakże, po pierwsze, szkoda czasu na pamiętanie takiego wzoru, jeśli można to w miarę szybko policzyć, a po drugie, nasza metoda jest bardziej elastyczna i umożliwia przybliżanie wyników innymi krzywymi.

Przykład $f(x) = a \log_2 x + b$. Wyniki doświadczeń: $(\frac{1}{2}, 0), (1, 1), (2, 3)$.

Uwaga
