24 Ekstrema -cd.

Omówimy jeszcze dwa ważne typy zagadnień: ekstrema lokalne funkcji uwikłanych oraz ekstrema warunkowe.

Zacznijmy od nieco prostszej sytuacji: przypuśćmy, ze funkcja dwu zmiennych $z = z(x, y)$ jest uwikłana w równaniu

$$F(x, y, z) = 0, \quad \texttt{gdzie} \quad \frac{\partial F}{\partial z}(P_0) \neq 0,\ F(P_0) = 0.$$

Z twierdzenia o funkcjach uwikłanych wiemy, że gdy F jest funkcją klasy C^1 w otoczeniu punktu $P_0 = (x_0, y_0, z_0)$ w $\mathbb{R}^3$, to w pewnym otoczeniu U punktu (x_0, y_0) w $\mathbb{R}^2$ istnieje dokładnie jedna funkcja ciagła $z(x, y)$ taka, że $z(x_0, y_0) = z_0$ oraz $F(x, y, z(x, y)) = 0$ w tym otoczeniu U. Ponadto z jest klasy C^1 oraz

$$z'_x(x, y) = -\frac{F'_x(x, y, z(x, y))}{F'_z(x, y, z(x, y))}. \tag{1}$$

Analogiczny jest wzór na z'_y, różniący się tylko licznikiem: F'_y zamiast F'_x. Zresztą -jeśli już wiemy (np. stosując ogólną metodę opartą na twierdzeniu o lokalnej odwracalności), że z jest różniczkowalna -to te wzory wynikną łatwo z reguły łańcucha stosowanej do funkcji złożonej $F(x, y, z(x, y))$ równej stale zero.

Przy badaniu ekstremów zakładamy jeszcze, że F jest klasy C^2, wówczas z twierdzenia o pochodnej ilorazu wynika, że $z(\cdot, \cdot)$ również będzie klasy C^2. Jeśli w punkcie $Q_0 = (x_0, y_0)$ funkcja z ma ekstremum lokalne, to jej gradient w tym punkcie jest zerowy. Licznik ułamka we wzorze (1) jest więc w badanym punkcie zerem, podobnie $F'_y(P_0) = 0$. Wzór na drugą różniczkę funkcji $z(x, y)$ będzie dość skomplikowany, ale w punkcie stacjonarnym (x_0, y_0) znacznie się uprości: np.

$$z''_{xx}(x_0, y_0) = -\frac{F''_{xx}(P_0)}{F'_z(P_0)},$$

$$z''_{xy}(x_0, y_0) = -\frac{F''_{xy}(P_0)}{F'_z(P_0)}, \tag{2}$$

analogicznie dla z''_{yy}. Sprawdźmy na przykład wzór na pochodną mieszaną. Różniczkowanie względem y wzoru $F(x, y, z(x, y)) = 0$, zapisując $P = (x, y, z(x, y))$ oraz $Q = (x, y)$ mamy $0 = F'_y(P) + F'_z(P) z'_y(Q)$. Teraz ostatnią równość różniczkujemy względem x, (zauważmy, że np. $P'_x = (1, 0, z'_x(x, y))$) otrzymując

$$0 = F''_{xy}(P) + F''_{xz}(P) z'_x(Q) + (F''_{xz}(P) + F''_{zz}) z'_y(Q) + F'_z(P) z''_{xy}.$$

Teraz uwzględniamy zerowanie się gradientu $\nabla_{P_0} z$ i otrzymujemy w punkcie stacjonarnym $Q_0 = (x_0, y_0)$ funkcji z równość

$$0 = F''_{xy}(P_0) + F'_z(P_0) z''_{xy}(Q_0).$$

W efekcie, macierz Hessego $H_{P_0} z$ otrzymana jest przez pomnożenie macierzy drugich pochodnych cząstkowych F w punkcie P_0 względem zmiennych x, y przez $-\frac{1}{F'_z(P_0)}$. W przypadku funkcji uwikłanej $z = z(x)$ poprzez równanie typu $F(x, z(x)) = 0$ otrzymamy bezpośredni wzór na drugą pochodną w punkcie stacjonarnym x_0 postaci

$$z''(x_0) = -\frac{F''_{zz}(P_0)}{F'_z(P_0)}, \qquad \texttt{gdzie} \quad P_0 = (x_0, z(x_0)),\ z'(x_0) = 0.$$

W ten sposób (postępując analogicznie dla pozostałych pochodnych cząstkowych rzędu 2) wyprowadzimy następujący warunek wystarczający dla ekstremum lokalnego funkcji uwikłanej: (wypowiemy go w nieco ogólniejszej wersji dla funkcji $f : D \subset \mathbb{R}^n \to \mathbb{R}$ klasy C^1 w otoczeniu D punktu $P_0 := (\mathbf{x_0}, z_0)$. Tu dla uproszczenia zapisu niech $\mathbf{x} = (x_1, \ldots, x_{n-1})$ oznacza blok pierwszych $n-1$

zmiennych dla funkcji $f : D \subset \mathbb{R}^n \to \mathbb{R}$, ostatnią (n-tą) zmienną nazwijmy z, chociaż może niektórzy woleli by tu mieć y zamiast z. Chcę jednak zachować "ciągłość oznaczeń" z przypadkiem 3 zmiennych, gdzie $z = z(x, y)$.

Twierdzenie. Jeżeli pochodna cząstkowa $\frac{\partial F}{\partial z}(P_0)$ funkcji F jest dodatnia, zaś jej pochodne cząstkowe w tym punkcie względem poprzednich zmiennych są równe zero w punkcie P_0, to istnienie maksimum lokalnego funkcji $z(\cdot)$ uwikłanej w równaniu $F(\mathbf{x}, z(\mathbf{x})) = 0$ wynika z dodatniej określoności "częściowej macierzy Hessego" dla F względem bloku pierwszych $n-1$ zmiennych, czyli macierzy o wyrazach $a_{jk} = \frac{\partial^2 F}{\partial x_j \partial x_k}, j, k \in \{1, 2, \ldots, n-1\}$. Ujemna określoność daje minimum lokalne -wszystko ze względu na "minus" we wzorach (2).Oczywiście, gdy $\frac{\partial F}{\partial z}(P_0) < 0$, charakter ekstremów zmienia się na przeciwny.

Przykład. Zbadamy ekstrema $z(x, y)$ jako funkcji uwikłanej w równaniu

$$F(x, y, z) = x^4 + y^4 + z^4 - 2a^2(x^2 + y^2 + z^2) = 0, a > 0.$$

Od razu widać, że $F(x, y, z) = F(x, y, -z)$, więc mamy lokalnie parę funkcji rozwikłujących: $\pm z(x, y)$, gdzie wybór znaku dyktuje nam znak zadanej wartości z_0. Ponieważ $F'_x = 4x^3 - 4a^2x = 0$, mamy albo $x = 0$ albo $x = \pm a$, takie same możliwości-dla y. Ponieważ ma być $F'_z = 4z^3 - 4a^2z \neq 0$, dla $(x, y) = (0, 0)$ wartość $z = 0$ musimy odrzucić i pozostają 2 możliwości: $F(0, 0, z) = 0$ dla $z = \pm a\sqrt{2}$. Ponumerujmy te rozwiązania P_1, P_2. Dalej, gdy $|x| = a = |y|$, mamy rozwiązania $\pm z$ spełniające $|z| = a\sqrt{1 + \sqrt{3}}$, w kórych F'_z jest niezerowe. Dla niezerowych x, y mamy więc jeszcze 8 rozwiązań poprzez dowolność doboru znaków dla każdej z 3 zmiennych - oznaczmy je przez $P_j, 3 \leqslant j \leqslant 10$. Pochodne mieszane są zerowe, więc macierz Hessego dla z będzie diagonalna. Mamy $F''_{xx} = 12x^2 - 4a^2$, $F''_{yy} = 12y^2 - 4a^2$ i te wartości będą jednakowe, bo $|x| = |y|$ w naszych punktach stacjonarnych -więc stałego znaku. Określoność macierzy drugiej różniczki dla z bedzie więc dodatnia dla punktu $P_1 = (0, 0, a\sqrt{2})$ (minimum lokalne funkcji uwikłanej z) i maksimum lokalne w $(0, 0)$ dla $z = -a\sqrt{2}$ (punkt P_2). Podobnie w punktach $P_3, \ldots, P_{10}$ liczniki wzorów (2), czyli pochodne 2. rzędu "niemieszane" z F są takie same, zmienia się znak mianowników - w zależności od doboru znaku rozwiązania z. I tak, dla znaku "plus" wartość $z(x, y)$ równa $a\sqrt{1 + \sqrt{3}}$ jest maksimum lokalnym finkcji uwikłanej z, zaś wartość $-a\sqrt{1 + \sqrt{3}}$ jest jej minimum lokalnym.

24.1 Ekstrema warunkowe.

Jak już wiemy, w pewnych punktach zbiór M rozwiązań równania $G(x) = 0$, gdzie $G : D \subset \mathbb{R}^d \to \mathbb{R}$ lokalnie wygląda jak wykres funkcji. Ale na ogół jest to jednak pewien zwarty podzbiór $\mathbb{R}^n$ -np. sfera, elipsoida. Tak jest zwłaszcza, gdy przy $\|x\| \to +\infty$ jest $|G(x)| \to \infty$ - bo wówczas przeciwobrazy zbiorów zwartych są zwarte. Dla regularnych w odpowiednim sensie funkcji G takie zbiory stanowią pewne "gładkie powierzchnie zakrzywione" zwane rozmaitościami (ang. *manifolds*). Można formalnie wprowadzić ich definicję i np. w przypadku $d = 3$ wektorem normalnym (prostopadłym do płaszczyzny stycznej) dla takiej powierzchni M będzie gradient G. Nas interesują ekstrema funkcji określonych na M, gdzie wartości w punktach spoza M nie uwzględniamy. Takie funkcje f mogą być określone na znacznie większej,niż M dziedzinie- np. na jakimś zbiorze otwartym $U \subset \mathbb{R}^d$, jednak wartości w punktach $x \in U \setminus M$ nas nie interesują. Tak jest np. gdy liczymy kres górny wartości f po sferze jednostkowej -a więc jest to dość często występujące zagadnienie. W takim punkcie $P_0 \in M$ gradient f **nie musi być równy zero**, bo w otoczeniu tego punktu znajdą się punkty spoza M, gdzie f może mieć wartości istotnie większe, niż $\sup\{f(x) : x \in M\}$. Dobrym przykładem jest $f(x, y, z) = x + y + z$- funkcja ciągła osiągająca na sferze swoją wartość największą w pewnym punkcie P_*, ale jej gradient w każdym punkcie $\mathbb{R}^3$ jest taki sam, różny od zera. Pokażemy, jak taki punkt znaleźć, nie odwołując się do teorii rozmaitości. Zacznijmy jednak od definicji.

Definicja. Jeżeli $G_1, \ldots, G_k : D \subset \mathbb{R}^n \to \mathbb{R}$ są odwzorowaniami "wyznaczającymi warunek $G_j = 0$", to mówimy, że przy tym warunku funkcja $f : U \to \mathbb{R}$ ma maksimum lokalne warunkowe w punkcie P_0, jeśli po pierwsze $P_0 \in M := \{P \in D : \forall_{j \leqslant k} G_j(P) = 0\}$ a po drugie istnieje otoczenie W punktu P_0 takie, że gdy $P \in W \cap M$, to $f(P_0) \geqslant f(P)$. Analogicznie definiujemy minimum warunkowe, ścisłe (= silne) maksimum warunkowe.
Funkcją Lagrange'a dla powyższego zagadnienia nazwiemy funkcję

$$\Phi(x) := f(x) + \lambda_1 G_1(x) + \lambda_2 G_2(x) + \cdots + \lambda_k G_k(x).$$

Tu liczby $\lambda_j \in \mathbb{R}$ są pewnymi stałymi, zwanymi mnożnikami Lagrange'a (i na początku traktujemy je jako niewiadome).

Uwaga: Można wykazać (używając tw. o lokalnej odwracalności) twierdzenie o funkcyjnej zależności, które mówi, że gdy gradienty funkcji $G_1, \ldots, G_k$ są liniowo zależne, to w otoczeniu danego punktu którąś z tych funkcji da się wyrazić jako funkcję od pozostałych funkcji. W pewnym sensie, można ją wyeliminować z zestawu definiującego zbiór M. Będziemy więc zakładać liniową niezależność gradientów ∇G_j w interesującym nas punkcie.

Twierdzenie (Warunek konieczny na ekstremum warunkowe). Jeśli w punkcie $P_0 \in M$ funkcja f osiąga lokalne ekstremum warunkowe, to w tym punkcie gradienty funkcji f oraz G_j są liniowo zależne. (Ze względu na niezależność samych $\nabla_{P_0} G_j$ oznacza to, że $\nabla_{P_0} F$ jest liniową kombinacją typu $\sum_{j=1}^{k} (-\lambda_j) \nabla_{P_0} G_j$, czyli P_0 **jest punktem stacjonarnym dla funkcji Lagrange'a Φ.**

Zauważmy, że jeśli M jest zbiorem zwartym, to często już znajomość punktów stacjonarnych wystarcza, by znaleźć punkt, w którym f osiąga któryś ze swoich kresów (górny lub dolny) na zbiorze M. Jeśli jest tylko skończenie wiele punktów stacjonarnych dla Φ, możemy porównać wartości f w tych punktach i wybrać tę największą (odp. najmniejszą).
Przykład. Niech S będzie sferą jednostkową w $\mathbb{R}^d$ i rozważmy jako f formę kwadratową wyznaczoną przez macierz symetryczną $A = (a_{mj})_{m,j \leqslant d}$. Szukamy jej kresów na sferze S, w tym celu badamy ekstrema warunkowe

$$f(x) = \sum_{m=1}^{d} \sum_{j=1}^{d} a_{mj} x_m x_j, \quad \texttt{warunek to} \quad G(x) := \sum_j x_j^2 - 1 = 0.$$

Ponieważ w punkcie ekstremum warunkowego będzie

$$\forall_{p \in \{1,\ldots,d\}} \ 0 = \Phi_p'(x) = \sum_{j=1, j \neq p}^{d} (a_{jp} x_j + a_{pj} x_j) + 2a_{pp} x_p - 2\lambda x_p,$$

oznacza to, że λ jest taką liczbą, że dla pewnego wektora $x \in S$ (a więc niezerowego) jest $(A - \lambda I)x = 0$. To nic innego, jak fakt, że λ jest wartością własną macierzy A, zaś x -związanym z nią wektorem własnym. Czyli $Ax = \lambda x$. Natomiast (w zapisie macierzowym) $f(x) = x^T A x = \lambda x^T x = \lambda$, gdy $\|x\| = 1$. Stąd maksimum formy f na sferze jednostkowej, to jej największa wartość własna. Te wartości możemy znaleźć rozwiązując równanie charakterystyczne: $\det(A - \lambda I) = 0$. Następnie układ równań liniowych pozwolli na wyznaczenie podprzestrzeni własnej. Wektory własne są bowiem wyznaczone z dokładnością do stałego współczynnika. A ten współczynnik tak dobieramy, by umieścić wektor na sferze jednostkowej. Mamy przynajmniej dwa takie wektory na sferze S -bo $f(-x) = (-1)^2 f(x) = f(x)$. Gdy krotność wartości własnej (czyli wymiar jądra homomorfizmu $A - \lambda I$) jest większa od 1, takich wektorów na sferze jest... trochę więcej niż 2.

(dowód twierdzenia będzie naszkicowany w następnym wykładzie)