25 Ekstrema warunkowe-cd.

Przypomnijmy, że rozważamy następującą sytuację: $G_1, \ldots, G_k : D \subset \mathbb{R}^n \to \mathbb{R}$ są odwzorowaniami klasy C^1 "wyznaczającymi warunek $G_j = 0$", czyli definiującymi zbiór

$$M := \{P \in D : \forall_{j \leqslant k}\ G_j(P) = 0\}.$$

Wówczas funkcja $f : U \to \mathbb{R}$ ma maksimum (lokalne) warunkowe w punkcie P, jeśli po pierwsze $P \in M$, a po drugie istnieje takie otoczenie W punktu P, że wartości ze zbioru $f(W \cap M)$ są nie większe, niż $f(P)$. Mówimy, że f ma ekstremum warunkowe w P, jeśli jedna z funkcji: f lub $-f$ ma tam maksimum warunkowe.

Aby zbiór M miał strukturę $n-k$-wymiarowej rozmaitości zakładamy dodatkowo liniową niezależność gradientów funkcji G_j (przynajmniej w pewnym otoczeniu badanego punktu). Z taką sytuacją mamy do czynienia w przeważającej większości zastosowań.

Twierdzenie, które (przynajmniej częściowo) chcemy uzasadnić mówi, że warunkiem koniecznym istnienia ekstremum warunkowego w punkcie P dla funkcji klasy C^1 jest liniowa zależność układu wektorów: $\nabla_P f, \nabla_P G_1, \ldots, \nabla_P G_k$, czyli stacjonarność punktu P dla funkcji Lagrange'a

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

w której liczby $\lambda_j \in \mathbb{R}$ są pewnymi stałymi, zwanymi mnożnikami Lagrange'a.

Zacznijmy od najprostszej sytuacji funkcji f zmiennych $(x, y) \in D \subset \mathbb{R}^2$ -gdy chcemy sprawdzić, w jakim punkcie $P_0 = (x_0, y_0)$ może ona mieć ekstremum warunkowe przy warunku $g(x, y) = 0$. Warunek ten, gdy $g'_y(P_0) \neq 0$ -wyznacza y jako funkcję uwikłaną zmiennej x, czyli w pewnym otoczeniu W punktu P_0 mamy zbiór $M \cap W := g^{-1}(\{0\}) \cap W$ równy wykresowi pewnej funkcji $y(x)$, określonej dla $|x - x_0| < \delta$. W tym otoczeniu W wartości $f(x, y)$ w punktach zbioru M są postaci $f(x, y(x)), |x - x_0| < \delta$. Ekstremum warunkowe f w punkcie P_0 w tej sytuacji odpowiada zwykłemu ekstremum lokalnemu funkcji złożonej $\psi(x) := f(x, y(x))$ w punkcie x_0, czyli warunek konieczny, to $\psi'(x_0) = 0$. Ostatnia pochodna wyliczana będzie z reguły łańcucha i ze wzoru na pochodną funkcji uwikłanej, z uwzglednieniem równości $y(x_0) = y_0$:

$$0 = \psi'(x_0) = f'_x(P_0) + f'_y(P_0) y'(x_0) = f'_x(P_0) - f'_y(P_0) \frac{g'_x(P_0)}{g'_y(P_0)}. \tag{1}$$

Pomnożenie stronami tej równości przez $g'_y(P_0)$ i uwzględnienie interpretacji zerowania się wyznacznika macierzy 2×2 implikuje, że relacja (1) onacza liniową zależność gradientów funkjcji f oraz g w punkcie P_0. W przypadku dwu wektorów (przy czym $\nabla_{P_0} G \neq 0$) oznacza to dokładnie, jak w tezie twierdzenia -istnienie stałej λ, dla której gradient z funkcji Lagrange'a $f + \lambda g$ jest zerowy w punkcie P_0.

Na tym szczególnym przypadku przerwę uzasadnianie twierdzenia o mnożnikach Lagrange'a. Pozostaje jeszcze problem typu ekstremum (i ewentualności jego braku w punkcie stacjonarnym.) Można tu przedstawić dwa podejścia. Jedno wiąże się z zachowaniem się restrykcji macierzy Hessego do przestrzeni stycznej w punkcie P_0 do rozmaitości $M = \{P : g(P) = 0\}$ Podejście to jest dość uniwersalne -jedyny "drobny problem", to brak w naszym kursie definicji rozmaitości i jej przestrzeni stycznej w danym punkcie P (dalej piszę P, żeby nie pisać P_0).

W pewnym uproszczeniu -możemy sobie wyobrazić wektory styczne jako wektory wyznaczone przez krzywe położone na danej powierzchni (np. na sferze) -przechodzące przez ten punkt, np. niech $\gamma : (-\epsilon, \epsilon) \to M \subset \mathbb{R}^n$, będzie taką krzywą klasy C^1, że $\gamma(0) = P$. Więc $\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$. Wówczas wektor $\gamma'(t)$ ma współrzędne będące pochodnymi w punkcie t z odpowiednich składowych $\gamma_j(t)$ i jest to wektor "odpowiadający krzywej γ". Możemy wyznaczyć w zbiorze takich krzywych γ, dla których $\gamma(0) = P$ relację równoważności polegającą na porównaniu jedynie wektorów $\gamma'(0)$ i traktowaniu otrzymanych klas równoważności właśnie jako wektorów stycznych. Ten trochę niecodzienny sposób jest używany w geometrii różniczkowej do definiowania przestrzeni stycznej. Gdy $M \subset \mathbb{R}^n$ jest zadana równaniem $g(x) = 0$, w którym g ma niezerowy gradient w punkcie P, to tak otrzymany zbiór wektorów stycznych jest podprzestrzenią wektorową (oznaczaną T_pM) wymiaru $n-1$ i gdy $n=3$ -to T_PM jest przechodzącą przez 0 płaszczyną równoległą do płaszczyzny geometrycznie stycznej do M w punkcie P. Można wykazać, że T_PM jest to zbiór wszystkich wektorów prostopadłych do gradientu funkcji g. W pewnym sensie podobna jest sytuacja, gdy g zastąpimy odwzorowaniem $G : D \subset \mathbb{R}^n \to \mathbb{R}^k$, czyli zestawieniem k funkcji: $G = (g_1, g_2, \dots g_k)$ i wówczas zakładamy, że gradienty tych składowych są liniowo niezależne w punkcie P (jak w twierdzeniu o odwzorowaniach uwikłanych). Innymi słowy, rząd macierzy Jacobiego dla G jest w tym punkcie (a więc i w jego otoczeniu) maksymalny, równy k. Pewien minor rzędu k tej macierzy (już niekoniecznie główny) jest niezrowy. Wówczas przestrzeń styczna T_PM dla $M = G^{-1}(\{0\})$ ma wymiar $n-k$ i jest, jak można wykazać, zbiorem wektorów prostopadłych do każdego z gradientów $\nabla_P g_j,\ j = 1, \dots, k$. Warto podkreślić, że wbrew swojej nazwie, przestrzeń styczna T_PM do rozmaitości M nie "dotyka zbioru M w punkcie P" , wraz z przemieszczaniem się punktu P po powierzchni M raczej obraca się wokół "początku układu współrzędnych" -czyli wektora 0. Gdy M jest powierzchnią boczną walca lub stożka - to wzdłuż pewnych prostych leżących na powierzchni M taka płaszczyzna styczna może być stała, wówczas mamy "obrót o kąt zero stopni". Taką strukturę, w której punktom P powierzchni M przyporządkujemy "zmieniającą się w sposób ciągły przestrzeń wektorową V_P o stałym wymiarze nazywamy wiązką wektorową nad M, jej przekrojami nazywamy odwzorowania $v(\cdot)$ przypisujące punktom $P \in M$ pewne wektory $v(P) \in V_P$. Przykładem jest wiązka styczna. Słynne "Twierdzenie Borsuka[1] o uczesaniu sfery" mówi, że nie istnieje ciągły przekrój wiązki stycznej do sfery w R^3 taki, by dla każdego jej punktu P było $v(P) \neq 0$. Jeśli wektory to włosy rosnące na powierzchni sfery i wszystkie mają niezerową długość, a zależą w sposób ciągły od położenia punktu na sferze, z którego wyrastają, to któryś (przynajmniej jeden) musi nie mieścić się na płaszczyźnie stycznej -czyli musi np. sterczeć pionowo. W tym sensie nie da się "idealnie uczesać sfery" - gdzieś musi się pojawić "kręciołek". O dowód nie należy jednak pytać fryzjerów, tylko specjalistów od topologii algebraicznej.

Teraz jeśli f osiąga ekstremum lokalne na zbiorze M w punkcie P, i γ jest krzywą na M określoną w otoczeniu zera, wyznaczającą wektor styczny $\gamma'(0)$, to $f \circ \gamma$ jest funkcją określoną w otoczeniu zera, osiągającą (zwykłe, "jednowymiarowe") ekstremum lokalne w punkcie $t = 0$, więc jej pochodna w tym punkcie jest zerem. Ale z reguły łańcucha wiemy, że ta pochodna, to iloczyn skalarny wektora $\nabla_P f$ przez wektor $\gamma'(0)$.

Gradient f w punkcie P ekstremum warunkowego jest więc prostopadły do każdego z wektorów przestrzeni stycznej T_PM.

Można też podać interpretacje w terminach różniczki: $d_P f$, ponieważ jej wartość na wektorze v jest równa iloczynowi skalarnemu gradientu f przez ten wektor v, zaś zerowanie się iloczynu skalarnego -to prostopadłość danych wektorów:

Warunek konieczny dla ekstremum warunkowego w punkcie P, to zerowanie się różniczki f na przestrzeni stycznej: $d_P f|_{T_PM} \equiv 0$.

Porównując wymiary widzimy, że wektory gradientów $\nabla_P g_j$, $j = 1, \dots k$ stanowią bazę dopełnienia ortogonalnego podprzestrzeni stycznej T_PM w przestrzeni $\mathbb{R}^n$, czyli bazę przestrzeni wszystkich wektorów normalnych do M w punkcie P. Wynika stąd, że $\nabla_P f$ musi być kombinacją liniową tych wektorów bazowych. W ten sposób uzasadniliśmy twierdzenie o warunku koniecznym podane w poprzednim wykładzie (o ile przyjmiemy bez dowodu fakt opisu przestrzeni stycznej do M przy użyciu gradientów z funkcji zadających warunek).

[1] Ciekawy artykuł o tym polskim matematyku znalazłem tutaj: https://steemit.com/polish/@romualdd/karol-borsuk-naukowiec-dzialacz-ak-tworca-gier-planszowych

Na tym nie kończą się zalety tego podejścia. Okazuje się, że dodatniość formy kwadratowej wyznaczonej przez macierz Hessego $H_P f$ (czyli drugiej różniczki) - ale po zawężeniu do niezerowych wektorów z przestrzeni stycznej $T_P M$ wystarcza do osiągania minimum warunkowego w punkcie P -o ile w tym punkcie zachodzi warunek konieczny (w postaci $d_P f|_{T_P M} \equiv 0$ lub równoważnej). Widać, że zastąpienie "pełnej przestrzeni $\mathbb{R}^n$ przez przestrzeń styczną do M pozwala przenieść niemal dosłownie teorię ekstremów lokalnych na grunt teorii ekstremów warunkowych.

Drugie, bardziej bezpośrednie podejście do warunków wystarczajacych na ekstrema warunkowe dostępne jest dla funkcji dwu zmiennych ($n = 2$). Mamy wtedy jeden warunek postaci $g(x, y) = 0$. Przypuśćmy, że P jest punktem stacjonarnym dla funkcji Lagrange'a $\Phi := f + \lambda g$. Zamiast wyznaczania przestrzeni stycznej i badania na niej określoności formy kwadratowej Hessego - możemy utworzyć "Hessian z brzegiem $|H^b_P(f)|$" (w wersji angielskiej -*"bordered Hessian"*) czyli wyznacznik postaci:

$$\begin{vmatrix} 0 & -g'_x(P) & -g'_y(P) \\ -g'_x(P) & \Phi''_{xx}(P) & \Phi''_{xy}(P) \\ -g'_y(P) & \Phi''_{yx}(P) & \Phi''_{yy}(P) \end{vmatrix}$$

Można wykazać, że gdy w punkcie spełniającym warunek konieczny mamy $|H^b_P(f)| > 0$, to w punkcie tym f osiąga maksimum warunkowe. W przypadku przeciwnej ostrej nierówności mamy minimum. Gdy $|H^b_P(f)| = 0$, sytuacja pozostaje nierozstrzeygnieta.

Przykład. Dla $n \in \mathbb{N}$ niech $f(x, y) = (x - y)^n$ i szukamy ekstremów tej f przy warunku $g(x, y) = x^2 + y^2 - 1 = 0$. Dla funkcji Lagrange'a $f - \lambda g$ punkty stacjonarne, to rozwiązania układu równań:

$$\begin{aligned} n(x-y)^{n-1} + 2\lambda x &= 0 \\ -n(x-y)^{n-1} + 2\lambda y &= 0 \\ x^2 + y^2 - 1 &= 0 \end{aligned}$$

Dodając stronami dwa pierwsze równania widzimy, że $\lambda(x + y) = 0$, więc albo $\lambda = 0$ i wówczas $x = y = \pm\frac{\sqrt{2}}{2}$, albo $x = -y$ Gdy $x = y$, to $f(x, y) = 0$. W przypadku $x = -y$ o module $\frac{\sqrt{2}}{2}$ wartości f w tych punktach, to $(\pm\sqrt{2})^n$ i są one równe dla n parzystych , zaś dla n nieparzystych są przeciwnych znaków (równe odpowiednio $\pm 2^{\frac{n}{2}}$). Możemy te punkty ponumerowac jako $P_1, \ldots P_4$ -są to kolejne wierzchołki kwadratu wpisanego w okrąg $x^2 + y^2 = 1$ zaczynając numerację (w kierunku wskazówek zegara) od pierwszej ćwiartki układu współrzędnych. Reasumując -mamy 4 punkty krytyczne funkcji Lagrange'a. Zobaczmy, czy te dane uuzyskamy też patrząc na hessian z brzegiem. Jego "brzeg" obramowujący z góry i z lewej strony zwykłą macierz Hessego dla Φ, to wektor $(0, -g'_x, -g'_y)$, czyli w naszym przypadku wektor $(0, -2x, -2y)$. Wyrazy przekątniowe macierzy Hessego, czyli Φ''_{xx}, Φ''_{yy} są w tym przypadku równe i wynoszą $n(n-1)(x-y)^{n-2} + 2\lambda$. Pochodne mieszane dla Φ, to $-n(n-1)(x-y)^{n-2}$. Stąd cały wyznacznik, jak nietrudno przeliczyć, jest równy

$$|H^b_P(f)| = -4n(n-1)(x-y)^{n-2}(x+y)^2 + 8\lambda(x^2 - y^2).$$

W naszych 4 punktach drugi składnik znika. Dla $n = 1$ lub $n \geqslant 3$ cała wartość $|H^b_P(f)|$ jest równa zero we wszystkich czterech punktach P_j. Jedynie dla $n = 2$ i to tylko w przypadku minimum warunkowego nasz test będzie skuteczny. Wówczas (w punktach $P \in \{P_1, P_3\}$ mamy $|H^b_P(f)| = -16$ i wtedy f osiąga wartość najmniejszą. Ale dla P_2, P_4 ten test pozostaje nierozstrzygający, bo nasz wyznacznik się zeruje, gdy tylko $x + y = 0$.

W prostszym przypadku -gdy szukamy ekstremów $f(x, y) = xy$ przy warunku $x + y = 6$, funkcja Lagrange'a $f + \lambda(6 - x - y)$ ma jedyny punkt stacjonarny $x = y = 3 = \lambda$. W tym przypadku mamy wyznacznik

$$|H^b_P(f)| = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 2 > 0,$$

więc w tym punkcie stacjonarnym jest maksimum warunkowe. Można się o tym przekonać i w nieco inny sposób: na prostej $y = 6 - x$ funkcja f przyjmuje dla $0 < x < 6$ wartości dodatnie, zaś poza domknięciem tego przedziału -już tylko ujemne. W jakimś punkcie musi więc przyjąć wartość największą. Oczywiście, w punkcie stacjonarnym dla Φ. Równie łątwo możemy tu badać $f(x, 6 - x)$.

Powyższa metoda ma odpowiednik dla większej ilości zmiennych -lecz oprócz wyznacznika całej macierzy z brzegiem postaci $(0, -g'_{x_1}, \cdots - g'_{x_n})$ trzeba badać znaki minorów kątowych głównych rzędu nie mniejszego, niż 3. Gdy wszystkie są ujemne - mamy w punkcie stacjonarnym dla Φ mininum warunkowe f. W przypadku maksimum -zmiany znaków muszą być alternujące -lepiej podmienić wtedy f na $-f$, zamiast pytać, od jakiego znaku ta alternacja sie zaczyna. Szczegóły można znaleźć w paru monografiach -np. C.Caratheodory, *Calculus of Variations and Partial Differentioal Equations,* Holden-Day, San Francisco, 1965.

Są jeszcze zagadnienia na ekstrema zupełnie odmiennego typu -związane z tzw. rachunkiem wariacyjnym. Są to np. zagadnienia izoperymetryczne- gdy wśród krzywych zamkniętych o stałej długości (i ustalonego typu) poszukujemy tej określającej obszar o największym polu. Albo słynne zagadnienie brachistochrony. Brachistochrona to krzywa, po której czas staczania się masy punktowej od punktu A do punktu B pod wpływem stałej siły (np. siły ciężkości) jest najkrótszy. Nazwa pochodzi od złożenia greckich słów brachistos - " najkrótszy" i chronos - " czas". Problem ten postawiony w 1696 przez Jakuba Bernoulliego został rozwiązany niezależnie przez Leibniza, Newtona, Jana Bernoulliego oraz de L'Hospitala. Okazało się, że brachistochroną jest fragment cykloidy. Ponieważ te zagadnienia prowadzą do równań różniczkowych (równanie Eulera-Lagrange'a), metody rozwiązania poznacie Państwo nieco później.

Na koniec tego wykładu przyjrzyjmy się warunkom opisanym nie przez równania (bądź ich układy), lecz przez słabe nierówności. Na przykład mamy jakiś zbiór zwarty K opisany przez układ nierówności: $G_1(x) \leqslant 0, \ldots, G_k(x) \leqslant 0$ oraz określoną na nim funkcję klasy C^1, powiedzmy funkcję f. Wiemy, że już z samej jej ciągłości wynika, że w pewnych punktach P_1, P_2 funkcja f osiąga swoje minimum i maksimum na zbiorze K. Dla każdej z wartości $j = 1, 2$ mamy dwie możliwości: albo P_j leży na brzegu zbioru K, albo w jego wnętrzu. W tym drugim przypadku mamy zwykłe ekstremum lokalne. Na brzegu (np. dla $k = 1$ oznacza to , że $G_1(P_j) = 0$) mamy do czynienia z ekstremum warunkowym. Może się zdarzyć w wyższych wymiarach, że w tym punkcie częś funkjcji G_m przyjmuje wartości silnie mniejsze od zera, a tylko część -równe zero. Wtedy ostre nierówności wyznaczają zbiory otwarte (ich skończone przecięcie też jest otwarte), a te pozostałe (np. $G_{m+1}, \ldots, G_k$) przyjmują w badanym punkcie wartość zero. Tylko te funkcje bierzemy jako układ wyznaczający warunek w sensie ekstremum warunkowego. Pozostałe -mogą jedynie zawężać otwarty zbiór, na którym zdefiniowane jest f.

Przykład: Wyznaczmy maksimum i minimum z funkcji $f(x, y) = x^2 - y^2 + 2R^2$ w kole $K = \{(x, y) : x^2 + y^2 \leqslant R^2\}$. Punkty stacjonarne dla f, to tylko początek układu, czyli punkt $P_0 = (0, 0)$, co sprawdzamy przyrównując gradient f do zera. Mamy $f(0, 0) = 2R^2$. Następnie badamy kresy f na okręgu $\partial K = \{(x, y) : x^2 + y^2 = R^2\}$. Można to sprawdzić rozwikłując równanie okręgu, czyli wstawiając $y(x) = \pm\sqrt{R^2 - x^2}$. Wtedy mamy $f(x, y(x)) = 2x^2 + R^2$ dla $|x| \leqslant R$. Pochodna z tej funkcji w przedziale otwartym ma jedyne miejsce zetowe dla $x = 0$, wtedy $f(0, R) = R^2 = f(0, -R)$. Nie możemy też zapomnieć o końcach przedziału, czyli dla $x = \pm R$. Wtedy $y(x) = 0$ oraz $f(\pm R, 0) = 3R^2$. Obydwa ekstrema osiągnięte są więc na brzegu.

W przypadku funkcji wypukłych, czyli takich funkcji określonych na dzidzinie wypukłej D , ze $\forall_{t \in [o,1]}\, f(P_0 + t(P_1 - P_0)) \leqslant f(P_0) + t(f(P_1) - f(P_0))$ ich maksimum moze być osiągnięte jedynie na brzegu zbioru D.

O tym, że minimum modułu wielomianu zmiennej zespolonej w punkcie wewnętrznym koła nie może być silnie dodatnie przekonaliśmy się dowodząc zasadniczego twierdzenia algebry. Zasada Maksimum Modułu mówi, że żadna funkcja analityczna zmiennej zespolonej nie może mieć max $|f(z)|$ osiągniętego wewnątrz jakiegoś obszaru -jedynie na brzegu.