

Warunki optymalności dla ZPN bez ograniczeń

Maciej Patan

Instytut Sterowania i Systemów Informatycznych
Uniwersytet Zielonogórski

Zadanie optymalizacji bez ograniczeń

Szukamy takiego x^* , dla którego

$$f(x^*) = \min_{x \in \mathbb{R}^n} [f(x)] = \max_{x \in \mathbb{R}^n} [-f(x)]$$

gdzie $f : \mathbb{R}^n \rightarrow \mathbb{R}$ – dowolny funkcjonal osiągnący skończony kres dolny

- ▶ mimo braku ograniczeń duże znaczenie praktyczne:
 - ▶ istniejące ograniczenia często można zaniedbać lub uwzględnić w analizie pooptymalizacyjnej,
 - ▶ osiągnane rozwiązania mogą być nadal satysfakcjonujące w problemach z ograniczeniami,
- ▶ metody i narzędzia analityczne stanowią punkt wyjścia do bardziej skomplikowanych problemów z ograniczeniami.

Przykład:

Pocisk kalibru 7,62 mm o masie 15 g wystrzelono z AK-47 z prędkością $v_0 = 715$ m/s w powietrze pod kątem α_0 względem poziomu. Siła oporu powietrza wynosi $F_{op} = kv^2$, gdzie współczynnik oporu k dla kuli karabinowej wynosi około $k = 9 \cdot 10^{-6}$ kg/m. Znaleźć wartości α_0 , dla których pocisk osiągnie (a) maksymalną wysokość, (b) maksymalny zasięg?

Równania ruchu:

$$m \frac{d^2x}{dt^2} = -F_{op} \cos \alpha, \quad (\text{w poziomie w kierunku strzału})$$

$$m \frac{d^2y}{dt^2} = -mg - F_{op} \sin \alpha, \quad (\text{w pionie do góry w kierunku strzału})$$

warunki początkowe:

$$x(0) = 0, \quad y(0) = 0, \quad v_x(0) = v_0 \cos(\alpha_0), \quad v_y(0) = v_0 \sin(\alpha_0)$$

gdzie: $g = 9,81$ m/s², $\tan \alpha = v_y/v_x$.

Zakładając, że współczynnik $k = 0$ (pomijając opory ruchu) równania po scałkowaniu dają trywialne rozwiązania:

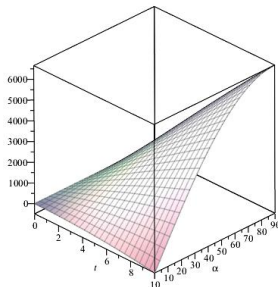
$$x(t) = v_0 t \cos \alpha_0,$$

$$y(t) = v_0 t \sin \alpha_0 - gt^2/2,$$

(a) wysokość $y(t; \alpha_0)$ jest opisana parabolą, z maksimum w chwili

$$t_1 = \frac{v_0 \sin \alpha_0}{g}, \text{ stąd}$$

$$y_{\max} = \frac{v_0^2 \sin^2 \alpha_0}{2g} \Rightarrow \sin(\alpha_0) = 1 \Rightarrow \alpha_0 = 90^\circ$$



(b) czas do opadnięcia do ustalonego pułapu y_f wynosi

$$t_f = \frac{v_0 \sin \alpha_0 + \sqrt{v_0^2 \sin^2 \alpha_0 - 2gy_f}}{g} = t_1 + \frac{\sqrt{v_0^2 \sin^2 \alpha_0 - 2gy_f}}{g}$$

Zasięg

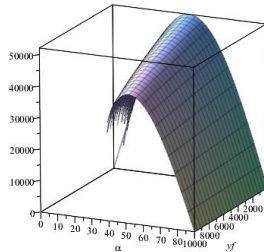
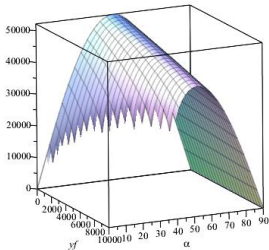
$$\begin{aligned} x(t; \alpha_0 | y(t) = y_f) &= x(t_f; \alpha_0) = v_0 t_f \cos \alpha_0 \\ &= \frac{v_0^2 \sin 2\alpha_0}{2g} + \frac{\sqrt{(v_0^2 \sin 2\alpha_0)^2 - 8v_0 g y_f \cos \alpha_0}}{2g} \end{aligned}$$

Gdy $y_f = 0$, to

$$x(t_f) = x(2t_1) = \frac{v_0^2}{g} \sin(2\alpha_0) \Rightarrow \alpha_0 = 45^\circ$$

Zatem

$$\begin{cases} y_f < 0 & \alpha_0 < 45^\circ \\ y_f = 0, & \alpha_0 = 45^\circ \\ y_f > 0 & \alpha_0 > 45^\circ \end{cases}$$

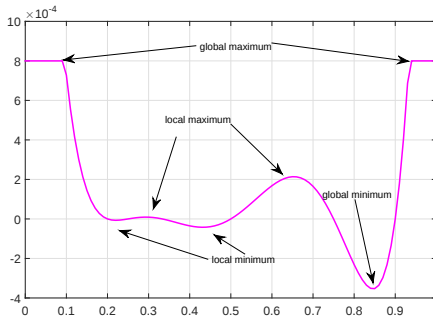


Przykładowo, dla $y_f = 1000$ m, to $\alpha_0 \approx 45.7^\circ$,
a dla $y_f = 8848$ m $\alpha_0 \approx 50.9^\circ$!

Zadanie do samodzielnego rozważenia

Jak zmieni się rozwiązanie gdy założymy współczynnik oporu powietrza $k = 9 \cdot 10^{-6}$ kg/m? Czy taka mała wartość współczynnika oporu ruchu uzasadnia zaniedbanie tego składnika?

Minima lokalne vs globalne



- ▶ Funkcja $f(x)$ przyjmuje w danym punkcie x^* minimum (maksimum) lokalne, jeśli w pewnym otwartym otoczeniu tego punktu funkcja nigdzie nie przyjmuje wartości mniejszych (większych).
- ▶ Jeśli dodatkowo w pewnym otwartym otoczeniu tego punktu funkcja nie przyjmuje wartości równych $f(x^*)$, to nazywamy takie ekstremum lokalne właściwym.
- ▶ Największa i najmniejsza wartość funkcji f w całej swojej dziedzinie nazywane są odpowiednio maksimum i minimum globalnym.

Wartości i wektory własne

Niech macierz kwadratowa A reprezentuje przekształcenie liniowe przestrzeni wektorowej na siebie samą. Jeżeli dla pewnego wektora v i pewnego skalaru λ spełniony jest warunek

$$Av = \lambda v$$

to v nazywamy wektorem własnym tego przekształcenia, a λ odpowiadającą mu wartością własną.

1. Wartości własne są pierwiastkami wielomianu charakterystycznego macierzy A :
 $p_A(\lambda) = \det(A - \lambda I) = 0$, gdzie I jest macierzą jednostkową,
2. Jeżeli A jest symetryczna to wszystkie jej wartości własne są liczbami rzeczywistymi.
3. Jeżeli znamy wartość własną λ_i , to odpowiadający jej wektor własny v^i jest rozwiązaniem układu równań:
 $(A - \lambda I) \cdot v_i = 0$.

Interpretacja: Jeżeli A definiuje przekształcenie liniowe przestrzeni (np. translację, obrót, skalowanie, rzut perspektywiczny, etc.) to wektory własne pokazują kierunki które są niezmiennicze względem przekształcenia, a odpowiadające im wartości własne pokazują jak przestrzeń jest skalowana (rozciągana/ściskana) w tym kierunku.

Przykład: Rozważmy przekształcenie $y = Ax$ gdzie:

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$$

Równanie charakterystyczne

$$\det \left(\begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} - \lambda \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) = 0 \Rightarrow$$

$$\Rightarrow (1-\lambda) \cdot (2-\lambda) - 2 \cdot 0 = 0 \quad \text{Wartości}$$

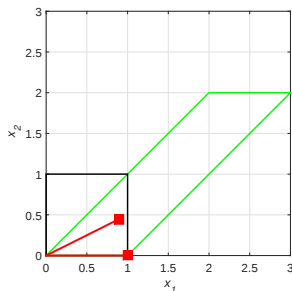
własne:

$$\lambda_1 = 1, \quad \lambda_2 = 2$$

Wektory własne:

$$v^1 = [1.0, 0.0]^T,$$

$$v^2 = [0.8944, 0.4472]^T$$



Określoność macierzy

Następujące stwierdzenia są równoważne:

1. Symetryczna macierz kwadratowa A jest dodatnio (ujemnie) określona, co zapisujemy $A \succ 0$ ($A \prec 0$).
 2. Dla każdego niezerowego wektora $x \in \mathbb{R}^n$, zachodzi $x^T A x > 0$ (< 0).
 3. Wszystkie wartości własne są dodatnie (ujemne), tzn. $\forall i, \lambda_i > 0$ (< 0).
-
- ▶ Analogicznie definiuje się nieujemną (niedodatnią) określoność macierzy, np. $A \succcurlyeq 0$ ($A \preccurlyeq 0$) $\Leftrightarrow \forall i, \lambda_i \leq 0$ (≥ 0).
 - ▶ Numerycznie często łatwiej sprawdzać np. elementy piwotalne macierzy lub wyznaczniki minorów głównych.

Gradient

Gradientem funkcji $f(x)$ nazywamy wektor:

$$\text{grad}(f(x)) = \nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1} \quad \frac{\partial f(x)}{\partial x_2} \quad \cdots \quad \frac{\partial f(x)}{\partial x_n} \right]^T$$

Interpretacja: Gradient (antygradient) funkcji w punkcie x_0 pokazuje kierunek największego przyrostu (spadku) funkcji widziany z tego punktu. Długość tego wektora jest ilościową miarą nachylenia funkcji w tym kierunku.

- ▶ Gradient prowadzi do uogólnienie pochodnej dla funkcji wielu zmiennych,
- ▶ Stosowany do opisu przestrzennego rozkładu zmian skalarnej wielkości fizycznej: temperatury, stężenia, ciśnienia, potencjału elektrycznego etc.

Przykład:

Rozważmy funkcję $f(x_1, x_2) = x_2 \sin(\pi x_1) + x_1 \cos(\pi x_2)$ w obszarze $(x_1, x_2) \in [-1, 1]^2$

Gradient:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} \pi x_2 \cos(\pi x_1) + \cos(\pi x_2) \\ \sin(\pi x_1) - \pi x_1 \sin(\pi x_2) \end{bmatrix}$$

Np. w punkcie $(0, 0)$:

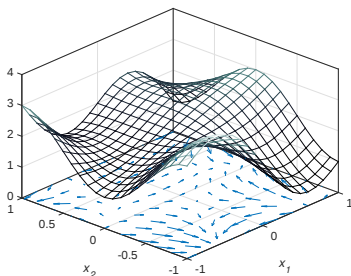
$\nabla f(0, 0) = [1, 0]^T$ - kierunek osi x_1

$$\|\nabla f(0, 0)\| = 1 \Rightarrow \operatorname{tg} \alpha = 1 \Rightarrow \alpha = 45^\circ$$

Rzeczywiście, dla $x_2 = 0$ mamy

$$f(x_1, 0) = x_1$$

czyli w przekroju wzdłuż osi x_1 funkcja jest liniowa ze współczynnikiem kierunkowym równym 1!



Stacjonarność

Punktem stacjonarnym funkcji $f : \mathbb{R}^n \rightarrow \mathbb{R}$ należącej do klasy C^1 , nazywamy punkt x^s , dla którego spełniony jest warunek:

$$\nabla f(x^s) = 0$$

Interpretacja: Zerowy maksymalny przyrost (równoważny dla odpowiednio gładkich funkcji zerowemu minimalnemu spadkowi) w punkcie oznacza, że funkcja jest stała ('wypłaszcza się') w pewnym infinitezymalnym otoczeniu tego punktu.

Twierdzenie 1: Warunek konieczny istnienia ekstremum

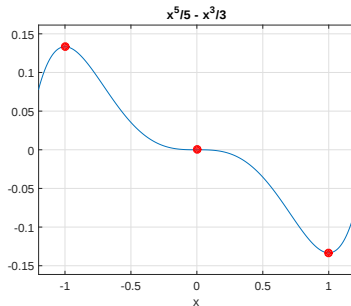
Jeżeli punkt $x^s \in \mathbb{R}^n$ jest minimum (maksimum) lokalnym funkcji $f \in C^1$ to jest punktem stacjonarnym tej funkcji, tzn. $\nabla f(x^s) = 0$

Twierdzenie odwrotne do Tw. 1 **nie jest prawdziwe!**

Przykład:

Rozważmy funkcję $f(x) = \frac{1}{5}x^5 - \frac{1}{3}x^3$. Znaleźć punkty stacjonarne.

Badamy gradient: $\nabla f(x) = f'(x) = x^2(x-1)(x+1) = 0$



- ▶ dobre narzędzie selekcji kandydatów, ale nie rozróżnia ekstremów,
- ▶ niezbędna jest analiza wypukłości funkcji.

Hesjan

Hesjanem funkcji $f(x) \in C^2$ nazywamy macierz:

$$H(x) = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n x_1} & \frac{\partial^2 f(x)}{\partial x_n x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

- **Interpretacja:** Hesjan jest uogólnieniem drugiej pochodnej dla funkcji wielu zmiennych. Dla $u, v \in \mathbb{R}^n$ forma biliniowa $u^T H(x)v$ jest drugą pochodną kierunkową funkcji f .
- Hesjan jest macierzą symetryczną, można ją zastosować do sprawdzenia wypukłości (wklęsłości) funkcji.

Twierdzenie 2: Warunek wystarczający istnienia ekstremum

Punkt $x^* \in \mathbb{R}^n$ jest minimum (maksimum) lokalnym funkcji $f \in C^2$ wtedy i tylko wtedy, gdy jest jej punktem stacjonarnym oraz hesjan w tym punkcie jest dodatnio (ujemnie) określony, tzn.

$$\nabla f(x^*) = 0, \quad \text{oraz} \quad \nabla^2 f(x^*) \succ 0 \quad (\prec 0)$$

- ▶ Jeżeli dodatkowo funkcja f jest wypukła (wklęsła) w całej dziedzinie, to powyższy warunek wystarcza, aby x^* był minimum (maksimum) globalnym
- ▶ Twierdzenie pozwala identyfikować różne typy punktów stacjonarnych na podstawie wartości własnych hesjanu.

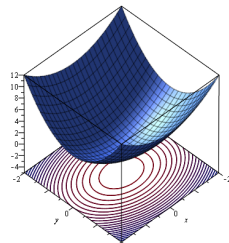
Typy punktów stacjonarnych:

- a) minima, $\forall i \lambda_i > 0$ (wszystkie w. własne są dodatnie),
 - b) maksima, $\forall i \lambda_i < 0$ (wszystkie w. własne są ujemne),
 - c) punkty siodłowe, $\forall i \lambda_i \neq 0 \wedge \exists i, j \lambda_i \lambda_j < 0$ (wszystkie w. własne są niezerowe oraz przyjmują wartości zarówno dodatnie jak i ujemne),
 - d) punkty osobliwe (zdegenerowane), $\exists i \lambda_i = 0$ (istnieją zerowe wartości własne).
-
- ▶ w przypadkach (a)-(c) dla odróżnienia od sytuacji (d) często mówi się o 'zwykajnych' lub 'standardowych' punktach stacjonarnych,
 - ▶ przypadek (d) nie jest rozstrzygający i niezbędna jest dalsza identyfikacja innymi metodami, punkt może być nadal ekstremum lub p. siodłowym (nazywanymi wtedy zdegenerowanymi) lub całkowicie innym punktem krytycznym (np. punktem przegięcia).

(a) $f(x, y) = x^2 + 2y^2$ w punkcie $(x, y) = (0, 0)$

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f(x)}{\partial x} \\ \frac{\partial f(x)}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x \\ 4y \end{bmatrix} \Rightarrow \nabla f(0, 0) = \begin{bmatrix} 2 \cdot 0 \\ 4 \cdot 0 \end{bmatrix} = \mathbf{0}$$

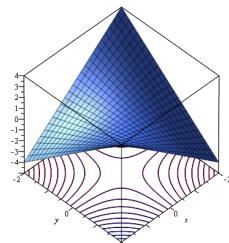
$$\nabla^2 f(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \Rightarrow \lambda_1 = 2, \lambda_2 = 4 \Rightarrow \text{minimum!}$$



(c) $f(x, y) = xy$ w punkcie $(x, y) = (0, 0)$

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f(x)}{\partial x} \\ \frac{\partial f(x)}{\partial y} \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix} \Rightarrow \nabla f(0, 0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{0}$$

$$\nabla^2 f(x, y) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \Rightarrow \lambda_1 = 1, \lambda_2 = -1 \Rightarrow \text{p. siodłowy!}$$



$$(d1) f(x, y) = x^2y + xy^2$$

$$\nabla f(x, y) = \begin{bmatrix} 2xy + y^2 \\ x^2 + 2xy \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow (x, y) = (0, 0)$$

$$\nabla^2 f(x, y) = \begin{bmatrix} 2y & 2x + 2y \\ 2x + 2y & 2x \end{bmatrix} \Rightarrow \nabla^2 f(0, 0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$\Rightarrow \lambda_1 = \lambda_2 = 0 \Rightarrow$ p. osobliwy!

Zauważmy, że w kierunkach $x = 0$, $y = 0$, $y = x$ oraz $y = -x$, funkcja jest stała, nie mamy zatem do czynienia z ekstremum właściwym!

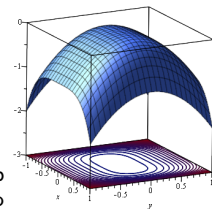
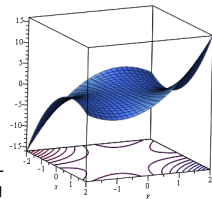
$$(d2) f(x, y) = -x^4 - y^2$$

$$\nabla f(x, y) = \begin{bmatrix} -4x^3 \\ -2y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow (x, y) = (0, 0)$$

$$\nabla^2 f(x, y) = \begin{bmatrix} -12x^2 & 0 \\ 0 & -2 \end{bmatrix} \Rightarrow \nabla^2 f(0, 0) = \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}$$

$\Rightarrow \lambda_1 = 0, \lambda_2 = -2 \Rightarrow$ p. osobliwy!

Można zastosować podstawienie $r = x^2$ (dlaczego to zadziała?) lub eksplorując wyższe pochodne składników aż pojawi się pierwsza o wartości niezerowej, tutaj $f_x^{(IV)}(0, 0) = -24$ (czy wystarczy zbadać tylko pochodne czyste?). Stąd mamy do czynienia z maksimum zdegenerowanym!!!



Przykład:

Zidentyfikować minima i maksima funkcji

$$f(p, q, r) = (2p^2 + 3q^2 + 4r^2) \exp(1 - p^2 - q^2 - r^2)$$

Zerujemy gradient:

$$\begin{cases} 2p(2 - (2p^2 + 3q^2 + 4r^2))e^{(1-p^2-q^2-r^2)} = 0 \\ 2q(3 - (2p^2 + 3q^2 + 4r^2))e^{(1-p^2-q^2-r^2)} = 0 \\ 2r(4 - (2p^2 + 3q^2 + 4r^2))e^{(1-p^2-q^2-r^2)} = 0 \end{cases}$$

Stąd, wyznaczamy punkty stacjonarne $S = (p, q, r)$:

$$S^1 = (0, 0, 0), \quad S^{2/3} = (\pm 1, 0, 0), \quad S^{4/5} = (0, \pm 1, 0), \quad S^{6/7} = (0, 0, \pm 1)$$

Pochodne 2-go rzędu:

$$\frac{\partial^2 f}{\partial p^2} = [4 - 20p^2 - 6q^2 - 8r^2 + (4(2p^2 + 3q^2 + 4r^2))p^2]e^{(1-p^2-q^2-r^2)}$$

$$\frac{\partial^2 f}{\partial p \partial q} = [-20qp + (4(2p^2 + 3q^2 + 4r^2))qp]e^{(1-p^2-q^2-r^2)}$$

$$\frac{\partial^2 f}{\partial p \partial r} = [-24rp + (4(2p^2 + 3q^2 + 4r^2))rp]e^{(1-p^2-q^2-r^2)}$$

$$\frac{\partial^2 f}{\partial q^2} = [6 - 30q^2 - 4p^2 - 8r^2 + (4(2p^2 + 3q^2 + 4r^2))q^2]e^{(1-p^2-q^2-r^2)}$$

$$\frac{\partial^2 f}{\partial q \partial r} = [-28rq + (4(2p^2 + 3q^2 + 4r^2))rq]e^{(1-p^2-q^2-r^2)}$$

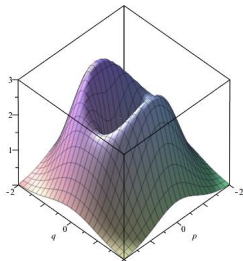
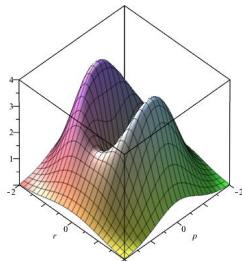
$$\frac{\partial^2 f}{\partial r^2} = [8 - 40r^2 - 4p^2 - 6q^2 + (4(2p^2 + 3q^2 + 4r^2))r^2]e^{(1-p^2-q^2-r^2)}$$

$$H(S^1) = \begin{bmatrix} 4e & 0 & 0 \\ 0 & 6e & 0 \\ 0 & 0 & 8e \end{bmatrix} \Rightarrow \text{minimum}$$

$$H(S^{2/3}) = \begin{bmatrix} -8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix} \Rightarrow \text{p. siodłowe}$$

$$H(S^{4/5}) = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -12 & 0 \\ 0 & 0 & 2 \end{bmatrix} \Rightarrow \text{p. siodłowe}$$

$$H(S^{6/7}) = \begin{bmatrix} -4 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -16 \end{bmatrix} \Rightarrow \text{maksima}$$

rzut $r = 0$ rzut $q = 0$ rzut $p = 0$ 