Optimalizace

8. Lokální extrémy vázané rovnostmi
(úlohy s omezeními typu rovnosti)

Tomáš Werner, Tomáš Kroupa

FEL ČVUT

Úlohy na extrémy vázané rovnostmi (= úlohy s omezeními typu rovnosti)

Hledáme extrémy funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ na množině

$$X = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) = \mathbf{0} \,\}$$

kde $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$. To také zapisujeme takto:

$$\begin{aligned} \text{min/max} \quad & f(\mathbf{x}) \\ \text{za podmínek} \quad & \mathbf{g}(\mathbf{x}) = \mathbf{0} \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

Příklad (jednoduchý):

$$\begin{aligned} \text{min} \quad & x + y \\ \text{za podmínky} \quad & x^2 + y^2 = 1 \\ & x, y \in \mathbb{R} \end{aligned}$$

- Máme $n = 2$ proměnné a $m = 1$ omezení.
- $f(x, y) = x + y$
- $g(x, y) = x^2 + y^2 - 1$ nebo $g(x, y) = 1 - x^2 - y^2$

Lokální extrémy vázané lineárními rovnostmi

Hledáme extrémy diferencovatelné funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ na afinním podprostoru

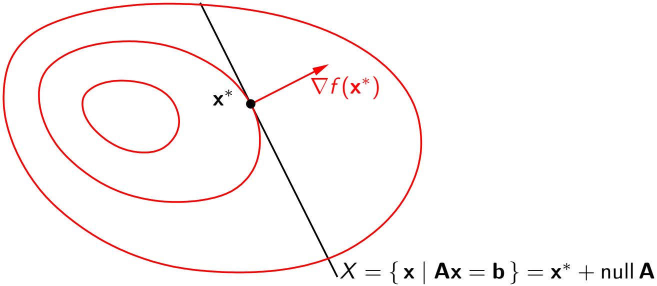
$$X = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b} \,\}$$

kde $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$. Neboli řešíme úlohu

$$\begin{aligned} \text{min/max} \quad & f(\mathbf{x}) \\ \text{za podmínek} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

Tedy $\mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$ je afinní zobrazení.

Podmínka prvního řádu



Věta

Nechť $\mathbf{x}^*$ je lokální extrém funkce f za podmínky $\mathbf{A}\mathbf{x} = \mathbf{b}$. Pak $\nabla f(\mathbf{x}^*) \perp \operatorname{null} \mathbf{A}$.

Důkaz: Protože $\mathbf{A}\mathbf{x}^* = \mathbf{b}$, množinu řešení soustavy $\mathbf{A}\mathbf{x} = \mathbf{b}$ lze psát jako

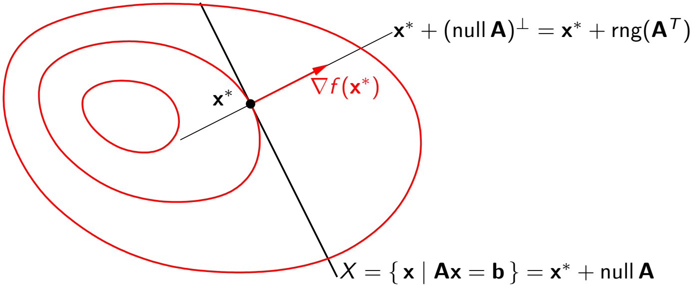
$$\mathbf{x}^* + \operatorname{null} \mathbf{A} = \{\, \mathbf{x}^* + \mathbf{B}\mathbf{y} \mid \mathbf{y} \in \mathbb{R}^k \,\}$$

kde sloupce matice $\mathbf{B} \in \mathbb{R}^{m \times k}$ tvoří bázi podprostoru $\operatorname{null} \mathbf{A}$.

Tedy bod $\mathbf{y} = \mathbf{0}$ musí být volný lokální extrém funkce $\varphi(\mathbf{y}) = f(\mathbf{x}^* + \mathbf{B}\mathbf{y})$. Tedy

$$\varphi'(\mathbf{y}) = f'(\mathbf{x}^* + \mathbf{B}\mathbf{y})\mathbf{B} = f'(\mathbf{x}^*)\mathbf{B} = \mathbf{0}.$$

Ale $f'(\mathbf{x}^*)\mathbf{B} = \mathbf{0}$ je totéž jako $\nabla f(\mathbf{x}^*) \perp \operatorname{null} \mathbf{A}$.



$$\begin{aligned}
\nabla f(\mathbf{x}^*) \perp \operatorname{null}\mathbf{A} &\iff \nabla f(\mathbf{x}^*) \in (\operatorname{null}\mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T) \\
&\iff \nabla f(\mathbf{x}^*) = f'(\mathbf{x}^*)^T \text{ je lin. kombinací sloupců } \mathbf{A}^T \\
&\iff f'(\mathbf{x}^*) = \boldsymbol{\lambda}^T\mathbf{A} \text{ pro nějaké } \boldsymbol{\lambda} \in \mathbb{R}^m
\end{aligned}$$

Podmínka prvního řádu (soustava $m+n$ rovnic s $m+n$ neznámými):

$$f'(\mathbf{x}) = \boldsymbol{\lambda}^T\mathbf{A}$$
$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

Příklad: Řešení lineární soustavy s nejmenší normou

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|\mathbf{x}\|^2 \\ \text{za podmínek} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

Podmínky prvního řádu:

$$\mathbf{x} = \mathbf{A}^T \boldsymbol{\lambda}$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

Má-li $\mathbf{A}$ lineárně nezávislé řádky, pak optimálním řešením je $\mathbf{x} = \underbrace{\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}}_{\mathbf{A}^+}\mathbf{b}$.

Příklad: Úloha nejmenších čtverců s lineárními omezeními

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \\ \text{za podmínek} \quad & \mathbf{C}\mathbf{x} = \mathbf{d} \end{aligned}$$

Podmínky prvního řádu:

$$\mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = \mathbf{C}^T\boldsymbol{\lambda}$$
$$\mathbf{C}\mathbf{x} = \mathbf{d}$$

neboli

$$\begin{bmatrix} \mathbf{A}^T\mathbf{A} & -\mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^T\mathbf{b} \\ \mathbf{d} \end{bmatrix}$$

Tvrzení: Matice $\begin{bmatrix} \mathbf{A}^T\mathbf{A} & -\mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix}$ je regulární, právě když $\begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix}$ má l.n. sloupce a $\mathbf{C}$ má l.n. řádky.

Lokální extrémy vázané nelineárními rovnostmi

Jsou dána spojitě diferencovatelná $f\colon \mathbb{R}^n \to \mathbb{R}$ a $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$.

Hledáme extrémy funkce f na množině

$$X = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) = \mathbf{0} \,\}$$

Neboli řešíme úlohu

$$\begin{aligned} \text{min/max} \quad & f(\mathbf{x}) \\ \text{za podmínek} \quad & \mathbf{g}(\mathbf{x}) = \mathbf{0} \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

Náš postup odvození podmínek na lokální extrémy:

1. Množinu X linearizujeme v okolí lokálního extrému.
2. Tím úlohu převedeme na případ lineárních omezení, který už známe.

Tečný prostor k množině přípustných řešení

Jak 'linearizovat' množinu $X = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) = \mathbf{0} \,\}$ v okolí daného bodu $\mathbf{x}^* \in X$?

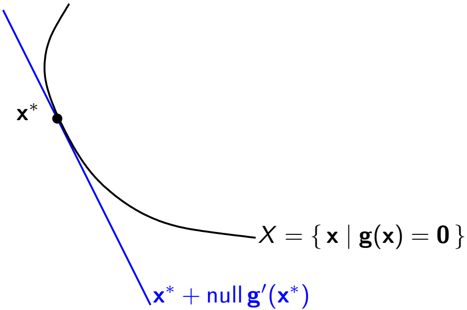
Hledáme 'tečný' prostor k množině X v bodě $\mathbf{x}^*$.

Aproximujme $\mathbf{g}$ v okolí bodu $\mathbf{x}^*$ Taylorovým polynomem 1. stupně:

$$\mathbf{g}(\mathbf{x}) \approx \mathbf{T}^1_{\mathbf{x}^*}(\mathbf{x}) = \underbrace{\mathbf{g}(\mathbf{x}^*)}_{\mathbf{0}} + \mathbf{g}'(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*)$$

Množina X se tím změní na (afinní) podprostor

$$\tilde{X} = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}'(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) = \mathbf{0} \,\} \; = \; \mathbf{x}^* + \operatorname{null} \mathbf{g}'(\mathbf{x}^*)$$



Podmínka regularity

Potíž: Ne vždy se množina X v okolí bodu $\mathbf{x}^*$ podobá afinnímu podprostoru $\tilde{X}$ (přestože se zobrazení $\mathbf{g}$ v okolí bodu $\mathbf{x}^*$ může podobat afinnímu zobrazení $\mathbf{T}^1_{\mathbf{x}^*}$)!

Příklady: $g(x, y) = xy$ nebo $g(x, y) = x^2$ v bodě $(x^*, y^*) = (0, 0)$

Definice

Bod $\mathbf{x} \in \mathbb{R}^n$ je **regulární bod** zobrazení $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$, jestliže řádky Jacobiho matice $\mathbf{g}'(\mathbf{x})$ (tj. gradienty $\nabla g_1(\mathbf{x}), \dots, \nabla g_m(\mathbf{x})$) jsou lineárně nezávislé.

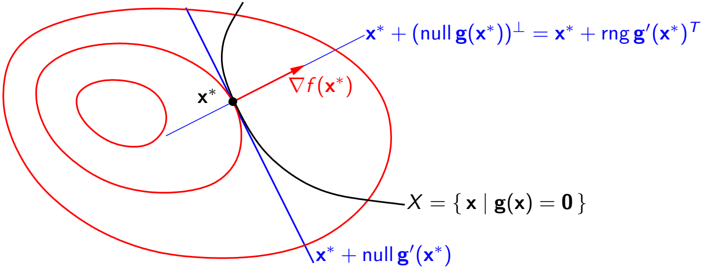
Speciálně: Bod $\mathbf{x}$ je regulární bod funkce $g\colon \mathbb{R}^n \to \mathbb{R}$, jestliže $\nabla g(\mathbf{x}) \neq \mathbf{0}$.

Tvrzení (neformální)

Nechť $\mathbf{x}^* \in X$ je regulární bod zobrazení $\mathbf{g}$.
Pak množina X je v okolí bodu $\mathbf{x}^*$ 'podobná' afinnímu podprostoru $\tilde{X}$.
Ten má dimenzi $n - m$ a je 'tečný' k množině X v bodě $\mathbf{x}^*$.

Věta

Nechť $\mathbf{x}^*$ je regulární bod zobrazení $\mathbf{g}$ a lokální extrém funkce f za podmínky $\mathbf{g}(\mathbf{x}) = \mathbf{0}$. Pak $\nabla f(\mathbf{x}^*) \perp \operatorname{null} \mathbf{g}'(\mathbf{x}^*)$.



$$\begin{aligned}
\nabla f(\mathbf{x}^*) \perp \operatorname{null} \mathbf{g}'(\mathbf{x}^*) \iff \nabla f(\mathbf{x}^*) &\in (\operatorname{null} \mathbf{g}'(\mathbf{x}^*))^\perp = \operatorname{rng}(\mathbf{g}'(\mathbf{x}^*)^T) \\
\iff f'(\mathbf{x}^*) &= -\boldsymbol{\lambda}^T \mathbf{g}'(\mathbf{x}^*) \text{ pro nějaké } \boldsymbol{\lambda} \in \mathbb{R}^m
\end{aligned}$$

Podmínka prvního řádu (soustava $m + n$ rovnic s $m + n$ neznámými):

$$\begin{aligned}
f'(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}'(\mathbf{x}) &= \mathbf{0} \\
\mathbf{g}(\mathbf{x}) &= \mathbf{0}
\end{aligned}$$

Zavedeme-li **Lagrangeovu funkci** $L\colon \mathbb{R}^{m+n} \to \mathbb{R}$

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}).$$

pak soustava výše jde napsat jako

$$\begin{aligned} L_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}) &= f'(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}'(\mathbf{x}) &= \mathbf{0} \\ L_{\boldsymbol{\lambda}}(\mathbf{x}, \boldsymbol{\lambda}) &= \mathbf{g}(\mathbf{x})^T &= \mathbf{0} \end{aligned}$$

neboli $L'(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$, tj. $(\mathbf{x}, \boldsymbol{\lambda})$ je stacionární bod funkce L.

Čísla $(\lambda_1, \ldots, \lambda_m) = \boldsymbol{\lambda}$ se nazývají **Lagrangeovy multiplikátory**.

Příklad

$$\begin{aligned} \text{min} \quad & x + y \\ \text{za podmínky} \quad & x^2 + y^2 = 1 \end{aligned}$$

Lagrangeova funkce

$$L(x, y, \lambda) = x + y + \lambda(1 - x^2 - y^2)$$

Stacionární podmínky pro L:

$$\begin{aligned} L_x(x, y, \lambda) &= 1 - 2\lambda x &&= 0 \\ L_y(x, y, \lambda) &= 1 - 2\lambda y &&= 0 \\ L_\lambda(x, y, \lambda) &= 1 - x^2 - y^2 &&= 0 \end{aligned}$$

Řešení: $(x, y, \lambda) = \pm(1, 1, 1)/\sqrt{2}$.
Našli jsme dva kandidáty na lok. extrémy: $(x, y) = \pm(1, 1)/\sqrt{2}$.

Příklad: Nesplněné podmínky regularity

$$\begin{aligned} \min \quad & x + y \\ \text{za podmínky} \quad & (1 - x^2 - y^2)^2 = 0 \end{aligned}$$

Množina přípustných řešení X je stejná kružnice jako minule, ale gradient

$$\nabla g(x, y) = (\,-4x(1 - x^2 - y^2),\, -4y(1 - x^2 - y^2)\,)$$

je na kružnici X nulový, tedy žádný bod X není regulární.

Lagrangeova funkce

$$L(x, y, \lambda) = x + y + \lambda(1 - x^2 - y^2)^2$$

Podmínky stacionarity

$$\begin{aligned} L_x(x, y, \lambda) &= 1 - 4\lambda x(1 - x^2 - y^2) &&= 0 \\ L_y(x, y, \lambda) &= 1 - 4\lambda y(1 - x^2 - y^2) &&= 0 \\ L_y(x, y, \lambda) &= (1 - x^2 - y^2)^2 &&= 0 \end{aligned}$$

nemají řešení. Lokální extrémy jsme nenašli, i když existují.

$$\begin{aligned} \min \quad & x \\ \text{za podmínky} \quad & x^2 = 0 \end{aligned}$$

Lokální extrém $x = 0$ není regulární bod, protože $g'(x) = 2x = 0$.

Lagrangeova funkce

$$L(x, \lambda) = x + \lambda x^2$$

Podmínky stacionarity

$$\begin{aligned} L_x(x, \lambda) &= 1 + 2x = 0 \\ L_\lambda(x, \lambda) &= x^2 \qquad = 0 \end{aligned}$$

nemají řešení. Lokální extrém jsme nenašli.

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{za podmínky} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

Lagrangeova funkce

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{b} - \mathbf{A}\mathbf{x})$$

Podmínky stacionarity

$$\begin{aligned} L_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}) = f'(\mathbf{x}) - \boldsymbol{\lambda}^T\mathbf{A} &= \mathbf{0} \\ L_{\boldsymbol{\lambda}}(\mathbf{x}, \boldsymbol{\lambda}) = (\mathbf{b} - \mathbf{A}\mathbf{x})^T \quad &= \mathbf{0} \end{aligned}$$

To jsme už odvodili dříve.

Pro lineární omezení jsme ale nemuseli předpokládat $\operatorname{rank} \mathbf{g}'(\mathbf{x}) = \operatorname{rank} \mathbf{A} = m$.

Příklad z optiky: Zákon odrazu z Fermatova principu

Fermatův princip nejkratšího času v optice (1662)

Paprsek mezi dvěma body letí takovou dráhou, aby doba letu byla nejkratší.

Modernější upřesnění: Doba letu je *stacionární* (vzhledem k variacím dráhy).

Křivé zrcadlo je plocha $X = \{\, \mathbf{x} \in \mathbb{R}^n \mid g(\mathbf{x}) = 0 \,\}$ kde $g\colon \mathbb{R}^n \to \mathbb{R}$.
Délka paprsku mezi dvěma body $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ s odrazem od zrcadla v bodě $\mathbf{x} \in X$ je

$$f(\mathbf{x}) = \|\mathbf{x} - \mathbf{a}\| + \|\mathbf{x} - \mathbf{b}\|.$$

Hledáme stacionární bod funkce f za podmínky $g(\mathbf{x}) = 0$.

Lagrangeova funkce

$$L(\mathbf{x}, \lambda) = \|\mathbf{x} - \mathbf{a}\| + \|\mathbf{x} - \mathbf{b}\| + \lambda g(\mathbf{x})$$

Podmínka $L_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$ je (po transpozici)

$$(\mathbf{x} - \mathbf{a})^0 + (\mathbf{x} - \mathbf{b})^0 + \lambda \nabla g(\mathbf{x}) = \mathbf{0} \qquad \text{kde jsme označili } \mathbf{y}^0 = \frac{\mathbf{y}}{\|\mathbf{y}\|}.$$

To říká, že vektor $\nabla g(\mathbf{x})$ (normála k zrcadlu v bodě $\mathbf{x}$) leží v jedné rovině s jednotkovými vektory $(\mathbf{x} - \mathbf{a})^0$ a $(\mathbf{x} - \mathbf{b})^0$ a půlí úhel mezi nimi.