Optimalizace

Lokální extrémy vázané rovnostmi

Tomáš Kroupa

FEL ČVUT

Úloha s omezeními ve tvaru rovností

Úloha

$$\begin{aligned} \min \quad & f(x_1, \dots, x_n) \\ \text{za podmínek} \quad & g_i(x_1, \dots, x_n) = 0, \quad i = 1, \dots, m \end{aligned}$$

- Budeme hledat podmínky optimality pro lokální extrémy funkce f vázané rovnostmi $\mathbf{g}(\mathbf{x}) = \mathbf{0}$
- Předpokládáme, že f i $\mathbf{g}$ jsou spojitě diferencovatelné

Lineární omezení

Pokud $\mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$, kde $\mathbf{A} \in \mathbb{R}^{m \times n}$, pak řešíme úlohu

$$\min\{f(\mathbf{x}) \mid \mathbf{A}\mathbf{x} = \mathbf{b}\}.$$

Tvrzení

Pro každý lokální extrém $\mathbf{x}$ této úlohy existuje $\boldsymbol{\lambda} \in \mathbb{R}^m$ tak, že

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

Úloha na nejmenší normu řešení nehomogenní soustavy

Hledáme řešení soustavy $\mathbf{A}\mathbf{x} = \mathbf{b}$ s nejmenší normou. Řešíme úlohu

$$\min \left\{ \tfrac{1}{2} \|\mathbf{x}\|^2 \mid \mathbf{A}\mathbf{x} = \mathbf{b} \right\}.$$

Podmínky optimality

$$\mathbf{x} = \mathbf{A}^\top \boldsymbol{\lambda} \qquad \mathbf{A}\mathbf{x} = \mathbf{b}$$

Má-li $\mathbf{A}$ lineárně nezávislé řádky, pak optimálním řešením je

$$\mathbf{x} = \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{b}.$$

Úloha nejmenších čtverců s lineárními omezeními

Hledáme řešení úlohy nejmenších čtverců pro soustavu $\mathbf{Ax} = \mathbf{b}$ s lineárními omezeními $\mathbf{Cx} = \mathbf{d}$. Tedy řešíme úlohu

$$\min \left\{ \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|^2 \mid \mathbf{Cx} = \mathbf{d} \right\}.$$

Podmínky optimality

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

Tečný vektor k množině

Přípustná řešení úlohy s omezeními ve tvaru rovností značíme

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

Definice

Vektor $\mathbf{v} \in \mathbb{R}^n$ je tečný k množině X v bodě $\mathbf{x} \in X$, pokud je v tom bodě tečným vektorem nějaké hladké křivky ležící v X.

Popis tečných vektorů

Tvrzení

Je-li vektor $\mathbf{v} \in \mathbb{R}^n$ tečný k množině X v bodě $\mathbf{x}$, pak $\mathbf{g}'(\mathbf{x})\mathbf{v} = \mathbf{0}$.

Věta

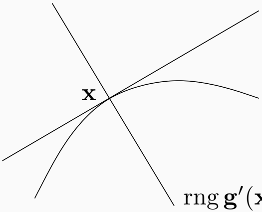
Pokud platí

1. $\operatorname{rank} \mathbf{g}'(\mathbf{x}) = m$
2. $\mathbf{g}'(\mathbf{x})\mathbf{v} = \mathbf{0}$

potom je vektor $\mathbf{v} \in \mathbb{R}^n$ tečný k množině X v bodě $\mathbf{x}$.

V regulárním bodě $\mathbf{x} \in X$ definujeme:

- Tečný prostor jako $\operatorname{null} \mathbf{g}'(\mathbf{x})$
- Ortogonální prostor jako $(\operatorname{null} \mathbf{g}'(\mathbf{x}))^\perp$



$$\operatorname{null} \mathbf{g}'(\mathbf{x}) = (\operatorname{span}\{\nabla g_1(\mathbf{x}), \dots, \nabla g_m(\mathbf{x})\})^\perp$$

$$X = \{\, \mathbf{x} \mid \mathbf{g}(\mathbf{x}) = \mathbf{0} \,\} = \{\, \mathbf{x} \mid g_1(\mathbf{x}) = \cdots = g_m(\mathbf{x}) = 0 \,\}$$

$$\operatorname{rng} \mathbf{g}'(\mathbf{x})^T = \operatorname{span}\{\nabla g_1(\mathbf{x}), \dots, \nabla g_m(\mathbf{x})\}$$

Podmínky prvního řádu

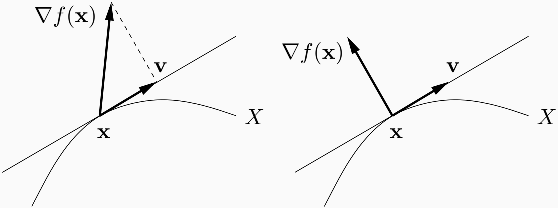
Věta

Pokud

1. $\mathbf{x} \in X$ je lokální extrém funkce f na množině X
2. $\operatorname{rank} \mathbf{g}'(\mathbf{x}) = m$

potom

$$\nabla f(\mathbf{x}) \in \operatorname{span}\{\nabla g_1(\mathbf{x}), \ldots, \nabla g_m(\mathbf{x})\}.$$



Metoda Lagrangeových multiplikátorů

Lagrangeovy multiplikátory $\boldsymbol{\lambda} \coloneqq (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$

Lagrangeova funkce $L\colon \mathbb{R}^{n+m} \to \mathbb{R}$

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

Podmínky optimality

$$\frac{\partial L(\mathbf{x}, \boldsymbol{\lambda})}{\partial \mathbf{x}} = f'(\mathbf{x}) + \boldsymbol{\lambda}^\mathsf{T} \mathbf{g}'(\mathbf{x}) = \mathbf{0}, \qquad \frac{\partial L(\mathbf{x}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \mathbf{g}(\mathbf{x})^\mathsf{T} = \mathbf{0}$$

Za příslušných předpokladů existuje $\boldsymbol{\lambda} \in \mathbb{R}^m$ splňující $L'(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$, tedy $(\mathbf{x}, \boldsymbol{\lambda})$ je stacionárním bodem funkce L.

Úskalí použití Lagrangeových multiplikátorů

- Řešit vzniklou soustavu nelineárních rovnic může být obtížné
- Podmínky druhého řádu pro funkci f na X nelze formulovat pomocí definitnosti Hessiánu Lagrangeovy funkce
- Nelze naivně aplikovat numerické metody na hledání stacionárních bodů Lagrangeovy funkce

Reference

- T. Werner. *Optimalizace* (kapitola 9). Elektronická skripta. FEL ČVUT, 2020.
- S. Boyd, L. Vandenberghe. Introduction to applied linear algebra: vectors, matrices, and least squares. Cambridge University Press, 2018.