Optimalizace

Lokální extrémy vázané rovnostmi

Tomáš Kroupa, Mirko Navara

19. 4. 2023

FEL ČVUT

Úloha s omezeními ve tvaru rovností

Úloha

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

- 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.

Obecně: Metoda Lagrangeových multiplikátorů

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

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

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

Podmínky optimality

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

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.

Sekvenční kvadratické programování (SQP)

Newtonova metoda minimalizace funkce (opakování)

Řešíme rovnici $f'(\mathbf{x}) = \mathbf{0}$ iterací $\mathbf{x}_{k+1} = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1}\, f'(\mathbf{x}_k)^\mathsf{T}$.

Použijeme pro $f(\mathbf{x}) \coloneqq L(\mathbf{x}, \boldsymbol{\lambda})$:

$$
\begin{aligned}
L(\mathbf{x}, \boldsymbol{\lambda}) &= f(\mathbf{x}) + \boldsymbol{\lambda}^\mathsf{T}\, \mathbf{g}(\mathbf{x}) \,, \\
L'(\mathbf{x}, \boldsymbol{\lambda}) &= \big(f'(\mathbf{x}) + \underbrace{\boldsymbol{\lambda}^\mathsf{T}\, \mathbf{g}'(\mathbf{x})}_{\sum_{i=1}^m \lambda_i\, g_i'(\mathbf{x})} \,,\ \mathbf{g}(\mathbf{x})^\mathsf{T} \big) \,, \\
L''(\mathbf{x}, \boldsymbol{\lambda}) &= \begin{bmatrix} f''(\mathbf{x}) + \sum_{i=1}^m \lambda_i\, g_i''(\mathbf{x}) & \mathbf{g}'(\mathbf{x})^\mathsf{T} \\ \mathbf{g}'(\mathbf{x}) & \mathbf{0} \end{bmatrix} \,,
\end{aligned}
$$

$$
\begin{bmatrix} \mathbf{x}_{k+1} - \mathbf{x}_k \\ \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k \end{bmatrix} = -L''(\mathbf{x}_k, \boldsymbol{\lambda}_k)^{-1}\, L'(\mathbf{x}_k, \boldsymbol{\lambda}_k)^\mathsf{T} =
$$

$$
= - \begin{bmatrix} f''(\mathbf{x}_k) + \sum_{i=1}^m \lambda_i\, g_i''(\mathbf{x}_k) & \mathbf{g}'(\mathbf{x}_k)^\mathsf{T} \\ \mathbf{g}'(\mathbf{x}_k) & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} f'(\mathbf{x}_k)^\mathsf{T} + \mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\, \boldsymbol{\lambda}_k \\ \mathbf{g}(\mathbf{x}_k) \end{bmatrix} .
$$

Tečný vektor k množině

Množina všech přípustných řešení úlohy s omezeními ve tvaru rovností:

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

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

O bodu $\mathbf{x} \in \mathbf{X}$ chceme rozhodnout, zda může být extrémem *nějaké* spojitě diferencovatelné funkce f.

Tečný prostor $T :=$ je lineární prostor vektorů tečných k X (v bodě $\mathbf{x}$).
Obvykle $\dim T = n - m$.

Ortogonální prostor $T^\perp$.
Nutná podmínka pro vázaný extrém: $\nabla f(\mathbf{x}) \in T^\perp$.
Obvykle $\dim T^\perp = m$.

$G := \operatorname{span}\{\nabla g_1(\mathbf{x}), \ldots, \nabla g_m(\mathbf{x})\} \subseteq T^\perp$.

Lagrangeovy multiplikátory dovolují najít *možný* extrém, právě když $\nabla f(\mathbf{x}) \in G$, přičemž $\dim G \leq m$.

Co Lagrangeovy multiplikátory nedovolují

Tvrzení

Je-li vektor $\mathbf{v} \in \mathbb{R}^n$ tečný k množině X v bodě $\mathbf{x}$, pak $\mathbf{g}'(\mathbf{x})\,\mathbf{v} = \mathbf{0}$, neboli $\mathbf{v}^\mathsf{T}\,\nabla\mathbf{g}(\mathbf{x}) = \mathbf{0}$.

Tj. $\mathbf{v} \in T \implies \mathbf{v} \in G^\perp$,
ale $\mathbf{v} \in G^\perp \not\Longrightarrow \mathbf{v} \in T$.
$T + G$ nemusí být celé $\mathbb{R}^n$; doplníme na ortogonální rozklad:
$\mathbb{R}^n = T + G + D$, kde $D := (T + G)^\perp = (T \cup G)^\perp$.

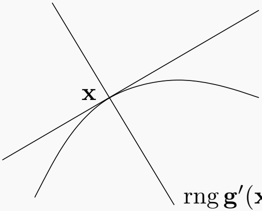
Ve směrech s nenulovou složkou z D nepomohou Lagrangeovy multiplikátory testovat nutnou podmínku $\nabla f(\mathbf{x}) \in T^\perp$.

Pokud $\dim G = m$, tj. $\nabla g_1(\mathbf{x}), \dots, \nabla g_m(\mathbf{x})$ jsou LN, pak $\mathbb{R}^n = T + G$, $\dim D = 0$ a Lagrangeovy multiplikátory dovolí testovat $\nabla f(\mathbf{x}) \in T^\perp = G$. V takovém případě říkáme, že $\mathbf{x}$ je regulární bod zobrazení $\mathbf{g}$ (nikoli regulární bod množiny X).

Tečný a ortogonální prostor v regulárních bodech

V regulárním bodě $\mathbf{x} \in X$ zobrazení $\mathbf{g}$ je:

- tečný prostor $T = \operatorname{null} \mathbf{g}'(\mathbf{x})$
- ortogonální prostor $T^\perp = (\operatorname{null} \mathbf{g}'(\mathbf{x}))^\perp$



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

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

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

Popis tečných vektorů v regulárních bodech

Věta

Pokud platí

1. $\operatorname{rank} \mathbf{g}'(\mathbf{x}) = m$, neboli $\mathbf{g}'(\mathbf{x})$ má plnou hodnost (pro $m \leq n$),
2. $\mathbf{g}'(\mathbf{x})\,\mathbf{v} = \mathbf{0}$, neboli $\mathbf{v}^T \nabla \mathbf{g}(\mathbf{x}) = \mathbf{0}$,

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

Podmínky prvního řádu v regulárních bodech

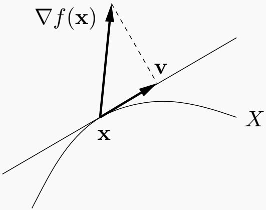
Věta

Pokud

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

potom

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



Co s body, které nejsou regulární?

- Pokud je jich málo, vyzkoušíme je všechny.
- Pokud je některý gradient $\nabla g_i(\mathbf{x}) = \mathbf{0}$, zkusme ho rozložit na součin a vyšetřit každý činitel zvlášť.
- Chybějící tečné vektory můžeme najít sami a zpracovat stejně jako gradienty omezujících funkcí.
- Místo toho můžeme odvodit jiné omezující funkce, určující stejný obor X.

Co s body, které nejsou regulární?

Speciální případ: Pokud jsou omezující podmínky lineární (opak.):

$$\mathbf{g}(\mathbf{x}) = \mathbf{C}\,\mathbf{x} - \mathbf{d}\,, \qquad \mathbf{g}'(\mathbf{x}) = \mathbf{C}\,, \qquad \nabla\mathbf{g}(\mathbf{x}) = \mathbf{C}^\mathsf{T}\,.$$

1. Jsou-li řádky matice $\mathbf{C}$ (neboli $\nabla g_1(\mathbf{x}), \ldots, \nabla g_m(\mathbf{x})$) lineárně nezávislé, pak jsou všechny body regulární.
2. Jsou-li řádky matice $\mathbf{C}$ (neboli $\nabla g_1(\mathbf{x}), \ldots, \nabla g_m(\mathbf{x})$) lineárně závislé, pak není *žádný* bod regulární a záleží na $\mathbf{d}$:
   - $X = \emptyset$: není co řešit.
   - $X \neq \emptyset$: můžeme z $[\mathbf{C} \;\; \mathbf{d}]$ vynechat LZ řádky a pokračovat dle bodu 1, nebo na to nedbat a řešit původní soustavu

$$f'(\mathbf{x}) = \boldsymbol{\lambda}^\mathsf{T}\,\mathbf{C}\,, \text{ neboli } \nabla f(\mathbf{x}) = \mathbf{C}^\mathsf{T}\,\boldsymbol{\lambda}\,,$$
$$\mathbf{C}\,\mathbf{x} = \mathbf{d}\,.$$

(To pro nelineární podmínky nejde.)

Prokletí dimenzionality

Pro velký počet omezujících funkcí je těžké zajistit nebo ověřit regularitu.

Můžeme doufat, že množina bodů, v nichž je regularita porušena, je „řídká“, takže i když přes ni projdeme, nezůstaneme v ní.

Alternativa: Metoda projektovaných gradientů

Někdy dovedeme k libovolnému bodu z $\mathbb{R}^n$ *snadno* najít „nejbližší“ přípustný z X („projekci“ na X), např. pro

$$X_1 = \{\mathbf{x} \mid \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\}\,,$$
$$X_2 = \{\mathbf{x} \mid \|\mathbf{x}\| \leq r\}\,,$$
$$X_3 = \{\mathbf{x} \mid \|\mathbf{x}\| = r\}\,.$$

Pak můžeme střídat kroky:

1. Postup ve směru $\nabla f(\mathbf{x})$ bez respektování omezujících podmínek.
2. Projekce na X.

Problém: V blízkosti extrému nás 1. krok zavádí dál od cíle.

Podmínky druhého řádu pro volné extrémy (opakování)

Věta

Nechť $X \subseteq \mathbb{R}^n$, $\mathbf{x} \in X$ je vnitřní bod množiny X a $f\colon \mathbb{R}^n \to \mathbb{R}$ je dvakrát diferencovatelná v $\mathbf{x}$. Pro stacionární bod $\mathbf{x}$ funkce f platí:

- Pokud je $\mathbf{x}$ lokální minimum, pak je Hessova matice $f''(\mathbf{x})$ pozitivně semidefinitní.
- Pokud je $f''(\mathbf{x})$ pozitivně definitní, pak je $\mathbf{x}$ ostré lokální minimum.
- Pokud je $f''(\mathbf{x})$ indefinitní, pak $\mathbf{x}$ není lokální extrém.

Matice $\mathbf{C}$ je pozitivně definitní na lineárním prostoru Y, jestliže $\forall \mathbf{x} \in Y \setminus \{\mathbf{0}\} : \mathbf{x}^\mathsf{T} C \mathbf{x} > 0$.

Podmínky druhého řádu pro vázané extrémy

$\frac{\partial^2 L}{\partial \mathbf{x}^2}(\mathbf{x}, \boldsymbol{\lambda}) = f''(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i''(\mathbf{x})\,.$

Věta

Nechť $\mathbf{x} \in X$ a funkce $f\colon \mathbb{R}^n \to \mathbb{R}, \mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$ jsou dvakrát diferencovatelné v $\mathbf{x}$. Nechť $\mathbf{x}$ je regulární bod zobrazení $\mathbf{g}$.

- Je-li $\mathbf{x}$ lokální minimum funkce f za podmínky $\mathbf{g}(\mathbf{x}) = \mathbf{0}$, pak $\exists \boldsymbol{\lambda} : L'(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$, $\frac{\partial^2 L}{\partial \mathbf{x}^2}(\mathbf{x}, \boldsymbol{\lambda})$ je pozitivně semidefinitní na $\operatorname{null} \mathbf{g}'(\mathbf{x})$.
- Jestliže $\exists \boldsymbol{\lambda} : L'(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$, $\frac{\partial^2 L}{\partial \mathbf{x}^2}(\mathbf{x}, \boldsymbol{\lambda})$ je pozitivně definitní na $\operatorname{null} \mathbf{g}'(\mathbf{x})$, pak $\mathbf{x}$ je ostré lokální minimum funkce f za podmínky $\mathbf{g}(\mathbf{x}) = \mathbf{0}$.
- Jestliže $\frac{\partial^2 L}{\partial \mathbf{x}^2}(\mathbf{x}, \boldsymbol{\lambda})$ je indefinitní na $\operatorname{null} \mathbf{g}'(\mathbf{x})$, pak $\mathbf{x}$ není lokální extrém.

Reference

- T. Werner. *Optimalizace* (kapitola 8). Elektronická skripta. FEL ČVUT, 17. 10. 2022. `https://cw.fel.cvut.cz/wiki/_media/courses/b0b33opt/opt.pdf`
- S. Boyd, L. Vandenberghe. Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares. Cambridge University Press, 2018.
- J. Hamhalter a J. Tišer. *Diferenciální počet funkcí více proměnných*. FEL ČVUT, 2005. `https://math.fel.cvut.cz/en/people/tiser/dipocet.pdf`
- M. Navara. *Matematika spojitého světa*. Elektronická skripta. FEL ČVUT, 2011. `https://cmp.felk.cvut.cz/~navara/MSS/MSS_print.pdf`