Optimalizace a teorie her

Podmínky optimality

Martin Bohata

Katedra matematiky
FEL ČVUT v Praze
bohata@math.feld.cvut.cz

Přípustné směry a směry poklesu

- Chceme lokálně aproximovat přípustnou množinu nějakou jednodušší množinou.
- Jakým směrem se lze pohnout o malý krok z daného bodu, aniž opustíme přípustnou množinu?

Definice

Nechť $M \subseteq \mathbb{R}^n$ a $x \in M$.

- Vektor $d \in \mathbb{R}^n$ se nazve přípustný směr množiny M v bodě x, jestliže existuje $\delta > 0$ tak, že pro každé $\alpha \in (0, \delta]$ je $x + \alpha d \in M$.
- Množina $\mathcal{F}(M; x)$ všech přípustných směrů množiny M v bodě x se nazývá kužel přípustných směrů množiny M v bodě x.

- $\mathcal{F}(M; x) \neq \emptyset$.
- Je-li $x \in \operatorname{int}(M)$, pak $\mathcal{F}(M; x) = \mathbb{R}^n$.
- Je-li M konečná (neprázdná), pak $\mathcal{F}(M; x) = \{0\}$ pro každé $x \in M$.

Přípustné směry a směry poklesu

Příklad

1. Je-li $M = S(0;1)$, pak $\mathcal{F}(M;x) = \{0\}$ pro každé $x \in M$.
2. Je-li $C = B(0;1)$ a $\hat{x} = (1,0)^T$, pak
$$\mathcal{F}(C;\hat{x}) = \left\{ \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \in \mathbb{R}^2 \,\middle|\, d_1 < 0 \right\} \cup \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}.$$

- Jakým směrem se pohnout z daného bodu, aby cílová funkce klesala?

Definice

Nechť $D \subseteq \mathbb{R}^n$, $x \in D$ a $f : D \to \mathbb{R}$.

- Vektor $d \in \mathbb{R}^n$ se nazve směr poklesu funkce f v bodě x, jestliže existuje $\delta > 0$ tak, že pro každé $\alpha \in (0,\delta]$ je $f(x + \alpha d) < f(x)$.
- Množina $\mathcal{D}(f;x)$ všech směrů poklesu funkce f v bodě x se nazývá kužel směrů poklesu funkce f v bodě x.

- Definice implicitně obsahuje podmínku $[x, x + \delta d] \subseteq D$.

Přípustné směry a směry poklesu

Tvrzení (Nutná geometrická podmínka lokálního extrému)

Jestliže x je bod lokálního minima funkce $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ na $M \subseteq D$, pak $\mathcal{F}(M;x) \cap \mathcal{D}(f;x) = \emptyset$.

Důkaz: Viz přednáška. ■

- Implikaci nelze obrátit. Jestliže $f(x_1, x_2) = x_2$,

$$M = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \,\middle|\, x_1^2 + x_2^2 = 1 \right\} \quad \text{a} \quad \hat{x} = \frac{1}{5}\begin{pmatrix} 3 \\ 4 \end{pmatrix},$$

pak

$$\mathcal{F}(M;\hat{x}) \cap \mathcal{D}(f;\hat{x}) = \emptyset.$$

Přesto $\hat{x}$ není bodem lokálního minima. Tím je bod $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$.

Přípustné směry a směry poklesu

- Kužel $\mathcal{D}(f;x)$ je obtížné najít. Pokusíme se proto tento kužel „linearizovat“.

Definice

Nechť $\Omega \subseteq \mathbb{R}^n$ je otevřená množina, $x \in \Omega$ a $f \in C^1(\Omega)$.

- Vektor $d \in \mathbb{R}^n$ se nazve silný směr poklesu funkce f v bodě x, jestliže $\langle \nabla f(x), d \rangle < 0$.
- Množina $\mathcal{D}_0(f;x)$ všech silných směrů poklesu funkce f v bodě x se nazývá kužel silných směrů poklesu funkce f v bodě x.

- Kužel $\mathcal{D}_0(f;x)$ je množina všech řešení lineární nerovnice

$$\langle \nabla f(x), d \rangle < 0.$$

- $\mathcal{D}_0(f;x)$ je konvexní kužel.

Přípustné směry a směry poklesu

Tvrzení

Nechť $\Omega \subseteq \mathbb{R}^n$ *je otevřená množina,* $x \in \Omega$ *a* $f \in C^1(\Omega)$*. Potom platí:*

1. *Je-li* $d \in \mathcal{D}(f;x)$*, potom* $\langle \nabla f(x), d \rangle \leq 0$*.*
2. $\mathcal{D}_0(f;x) \subseteq \mathcal{D}(f;x)$ *(tj. jestliže* $\langle \nabla f(x), d \rangle < 0$*, pak* $d \in \mathcal{D}(f;x)$*).*

Důkaz: Viz přednáška. ■

- Implikace v 1 nelze obrátit. Uvažme například funkci $f(x) = 1$. Pak $f'(x)d = 0$ pro každé $d \in \mathbb{R}$, ale $\mathcal{D}(f;x) = \emptyset$.
- Inkluze v 2 nelze obrátit. Uvažme například funkci $f(x) = x^3$. Ta má v 0 směr poklesu $d = -1$, ale $d \notin \mathcal{D}_0(f;0)$.

Základní podmínky optimality

Věta (Fermatova věta)

Nechť $\Omega \subseteq \mathbb{R}^n$ je otevřená množina, $M \subseteq \Omega$ a $\hat{x} \in M$ je bodem lokálního minima funkce $f \in C^1(\Omega)$ na M. Potom platí:

1. $\mathcal{F}(M;\hat{x}) \cap \mathcal{D}_0(f;\hat{x}) = \emptyset$ *(tj.* $\langle \nabla f(\hat{x}), d \rangle \geq 0$ *pro všechny* $d \in \mathcal{F}(M;\hat{x})$*).*
2. *Jestliže* $\hat{x} \in \operatorname{int}(M)$, *pak* $\nabla f(\hat{x}) = 0$.

Důkaz: Viz přednáška. ■

- Podmínka 1 (a tedy také nulovost gradientu v bodě 2) je nutná, nikoli postačující.
- Jaké dodatečné požadavky zajistí, aby podmínka 1 byla postačující pro existenci lokálního minima v bodě $\hat{x}$?

Základní podmínky optimality

Věta (O nutných a postačujících podmínkách pro konvexní úlohu)

Nechť $\Omega \subseteq \mathbb{R}^n$ je otevřená množina, $f \in C^1(\Omega)$ je konvexní na $C \subseteq \Omega$ a $\hat{x} \in C$. Potom platí:

1. $\hat{x} \in \operatorname{argmin}_{x \in C} f(x)$ *právě tehdy, když* $\mathcal{F}(C; \hat{x}) \cap \mathcal{D}_0(f; \hat{x}) = \emptyset$.
2. *Předpokládejme, že* $\hat{x} \in \operatorname{int}(C)$. *Pak* $\hat{x} \in \operatorname{argmin}_{x \in C} f(x)$ *právě tehdy, když* $\nabla f(\hat{x}) = 0$.

Důkaz: Viz přednáška. ■

Příklad

Je dána funkce

$$f(x_1, x_2) = x_1^2 + 3x_2^2 - 2x_1x_2 + x_1 - 2x_2.$$

Funkce f je ryze konvexní. Jediným bodem minima funkce f je $\frac{1}{4}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Základní podmínky optimality

Věta (O podmínkách optimality 2. řádu)

Nechť $\Omega \subseteq \mathbb{R}^n$ *je otevřená množina,* $M \subseteq \Omega$, $\hat{x} \in \operatorname{int}(M)$ *a* $f \in C^2(\Omega)$. *Potom platí:*

1. *Jestliže* $\hat{x}$ *je bod lokálního minima funkce* f *na* M*, pak* $\nabla^2 f(\hat{x})$ *je pozitivně semidefinitní.*
2. *Jestliže* $\nabla f(\hat{x}) = 0$ *a* $\nabla^2 f(\hat{x})$ *je pozitivně definitní, pak* $\hat{x}$ *je bod ostrého lokálního minima.*

Důkaz: Vynecháváme. ∎

- Pokud $\hat{x} \in \operatorname{int}(M)$ je bodem ostrého lokálního minima, pak nutně neplatí, že $\nabla^2 f(\hat{x})$ je pozitivně definitní (viz $f(x) = x^4$ a $\hat{x} = 0$).
- Analogie druhého tvrzení z předchozí věty pro semidefinitní matice neplatí (viz $f(x) = x^3$ a $\hat{x} = 0$).

Základní podmínky optimality

Příklad

Je dána funkce

$$f(x_1, x_2) = \frac{1}{3}x_1^3 + \frac{1}{2}x_2^2 + x_1x_2 + 2x_2.$$

Podezřelé body z lokálního extrému jsou

$$u = \begin{pmatrix} 2 \\ -4 \end{pmatrix} \quad \text{a} \quad v = \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

- Bod u je bodem ostrého lokálního minima.
- Bod v není bodem lokálního minima ani maxima, neboť v tomto bodě není splněna nutná podmínka druhého řádu pro funkci f ani pro funkci $-f$.

Omezení ve tvaru nerovnosti

- Jak popsat informaci obsaženou v $\mathcal{F}(M;\hat{x})$ pomocí jednodušeji konstruovatelného kuželu?

Definice

Ať $g_1, \ldots, g_k$ jsou reálné funkce definované na množině $\Omega \subseteq \mathbb{R}^n$, $M = \{x \in \Omega \,|\, g_1(x) \leq 0, \ldots, g_k(x) \leq 0\}$ a $x \in M$.

- Množina $\mathcal{I}\left((g_i)_{i=1}^k; x\right) := \{i \in \{1, \ldots, k\} \,|\, g_i(x) = 0\}$ se nazývá indexová množina aktivních omezení v bodě x.
- Jestliže $i \in \mathcal{I}\left((g_i)_{i=1}^k; x\right)$, pak $g_i(x) \leq 0$ se nazve aktivní omezení (ve tvaru nerovnosti) v bodě x.
- Jestliže $i \notin \mathcal{I}\left((g_i)_{i=1}^k; x\right)$, pak $g_i(x) \leq 0$ se nazve neaktivní omezení (ve tvaru nerovnosti) v bodě x.

Omezení ve tvaru nerovnosti

Definice

Nechť $\Omega \subseteq \mathbb{R}^n$ je otevřená množina, $g_1, \ldots, g_k \in C^1(\Omega)$, $x \in M$ a $M = \{x \in \Omega \,|\, g_1(x) \leq 0, \ldots, g_k(x) \leq 0\}$. Definujeme množinu

$$\begin{aligned} \mathcal{G}\left((g_i)_{i=1}^k; x\right) &:= \left\{d \in \mathbb{R}^n \,\middle|\, \langle \nabla g_i(x), d \rangle \leq 0 \ \ \forall i \in \mathcal{I}\left((g_i)_{i=1}^k; x\right)\right\} \\ &= \bigcap_{i \in \mathcal{I}\left((g_i)_{i=1}^k; x\right)} \{d \in \mathbb{R}^n \,|\, \langle \nabla g_i(x), d \rangle \leq 0\}. \end{aligned}$$

- $\mathcal{G}\left((g_i)_{i=1}^k; x\right)$ je uzavřený konvexní kužel obsahující 0.
- $\mathcal{G}\left((g_i)_{i=1}^k; x\right)$ (a také $\mathcal{I}\left((g_i)_{i=1}^k; x\right)$) závisí na popisu množiny M. Naproti tomu $\mathcal{F}(M; x)$ je geometrický koncept (tj. nezávisí na konkrétním popisu množiny M).

Omezení ve tvaru nerovnosti

Příklad

Je dána množina

$$M = \left\{ (x_1, x_2)^T \in \mathbb{R}^2 \,\middle|\, x_2 - x_1^3 \leq 0, -x_2 \leq 0 \right\}$$

a bod $\hat{x} = (0, 0)^T$. Označme $g_1(x_1, x_2) = x_2 - x_1^3$ a $g_2(x_1, x_2) = -x_2$. Pak

$$\begin{aligned} \mathcal{F}(M; \hat{x}) &= \left\{ \begin{pmatrix} d_1 \\ 0 \end{pmatrix} \,\middle|\, d_1 \geq 0 \right\}, \\ \mathcal{G}\left((g_1, g_2), \hat{x}\right) &= \left\{ \begin{pmatrix} d_1 \\ 0 \end{pmatrix} \,\middle|\, d_1 \in \mathbb{R} \right\}. \end{aligned}$$

Vidíme, že $\mathcal{F}(M; \hat{x}) \subsetneq \mathcal{G}\left((g_1, g_2), \hat{x}\right)$. Položme $g_3(x_1, x_2) = -x_1 - x_2$. Pak

$$\begin{aligned} M &= \left\{ (x_1, x_2)^T \in \mathbb{R}^2 \,\middle|\, g_i(x_1, x_2) \leq 0 \ \ \forall i \in \{1, 2, 3\} \right\}, \\ \mathcal{G}\left((g_1, g_2, g_3), \hat{x}\right) &= \mathcal{F}(M; \hat{x}). \end{aligned}$$

Omezení ve tvaru nerovnosti

- Úmluva: Nemůže-li dojít k nedorozumění, budeme psát jen $\mathcal{I}(x)$ a $\mathcal{G}(x)$ místo $\mathcal{I}\left((g_i)_{i=1}^k; x\right)$ a $\mathcal{G}\left((g_i)_{i=1}^k; x\right)$.
- Lze ukázat, že $\mathcal{F}(M; x) \subseteq \mathcal{G}(x)$. Jak jsme viděli v předchozím příkladě, rovnost obecně nenastává.
- Lze podmínku optimality $\mathcal{F}(M; \hat{x}) \cap \mathcal{D}_0(f; \hat{x}) = \emptyset$ nahradit podmínkou $\mathcal{G}(\hat{x}) \cap \mathcal{D}_0(f; \hat{x}) = \emptyset$?

Příklad

$$\begin{aligned} \text{minimalizujte } & x_1 + x_2 \\ \text{za podmínek } & x_2 - x_1^3 \leq 0, \\ & -x_2 \leq 0. \end{aligned}$$

Bod $\hat{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ je (jediné) řešení, avšak $\mathcal{G}(\hat{x}) \cap \mathcal{D}_0(f; \hat{x}) \neq \emptyset$.

Omezení ve tvaru nerovnosti

- Kužel $\mathcal{G}(x)$ je obecně příliš veliký. Nutné klást dodatečnou podmínku.

Věta (O nutných KKT podmínkách)

Nechť $\Omega \subseteq \mathbb{R}^n$ *je otevřená množina,* $f, g_1, \ldots, g_k \in C^1(\Omega)$*,*

$$M = \{x \in \Omega \,|\, g_1(x) \leq 0, \ldots, g_k(x) \leq 0\}$$

a $\hat{x} \in M$*. Jestliže* $\overline{\mathcal{F}(M;\hat{x})} = \mathcal{G}(\hat{x})$ *a* $\hat{x}$ *je bod lokálního minima* f *na* M*, pak existuje* $(\mu_1, \ldots, \mu_k)^T \in \mathbb{R}^k$ *tak, že*

$$\begin{aligned} \nabla f(\hat{x}) + \sum_{i=1}^{k} \mu_i \nabla g_i(\hat{x}) &= 0, \\ \mu_i g_i(\hat{x}) &= 0 \quad \textit{pro všechna } i \in \{1, \ldots, k\}, \\ \mu_i &\geq 0 \quad \textit{pro všechna } i \in \{1, \ldots, k\}. \end{aligned}$$

Důkaz: Viz přednáška. ■

Omezení ve tvaru nerovnosti

Terminologie:

- Podmínky

  (P1) $\nabla f(\hat{x}) + \sum_{i=1}^{k} \mu_i \nabla g_i(\hat{x}) = 0$,
  (P2) $\mu_i g_i(\hat{x}) = 0$,
  (P3) $\mu_i \geq 0$,

  se souhrně nazývají KKT podmínky.
- Podmínka (P1) se nazývá podmínka stacionarity.
- Podmínka (P2) se nazývá podmínka komplementarity.
- Koeficienty $\mu_1, \ldots, \mu_k \in \mathbb{R}$ splňující KKT podmínky se nazývají KKT multiplikátory (Lagrangeovy multiplikátory) v bodě $\hat{x}$.
- Bod $\hat{x}$ se nazve KKT bod, existuje-li vektor $(\mu_1, \ldots, \mu_k)^T$ KKT multiplikátorů v bodě $\hat{x}$.

Omezení ve tvaru nerovnosti

- Geometrická interpretace KKT podmínek: Gradienty funkcí z aktivních omezení (funkcí g_i, $i \in \mathcal{I}(\hat{x})$) v bodě $\hat{x}$ generují kužel, ve kterém leží $-\nabla f(\hat{x})$.

Příklad

$$\begin{aligned} \text{minimalizujte} \quad & x_1 + x_2 \\ \text{za podmínek} \quad & x_1, x_2 \geq 0. \end{aligned}$$

Lze ukázat, že

$$\overline{\mathcal{F}(\mathbb{R}^2_+; x)} = \mathcal{G}(x)$$

pro každé $x \in \mathbb{R}^2_+$. Bod $\hat{x} = 0$ je jediný KKT bod a jedná se o bod minima.

- Předpoklad $\overline{\mathcal{F}(M; \hat{x})} = \mathcal{G}(\hat{x})$ je jednou z tzv. podmínek regularity. Nahradíme-li $\overline{\mathcal{F}(M; \hat{x})} = \mathcal{G}(\hat{x})$ ve větě O nutných KKT podmínkách jakoukoli jinou podmínkou regularity, zůstane věta stále v platnosti.

Omezení ve tvaru nerovnosti

Definice

Nechť $\Omega \subseteq \mathbb{R}^n$ je otevřená množina, $g_1, \dots, g_k \in C^1(\Omega)$ a

$$M = \{x \in \Omega \,|\, g_1(x) \leq 0, \dots, g_k(x) \leq 0\}\,.$$

Řekneme, že $(g_i)_{i=1}^k$ splňuje

- (APR) Afinní podmínku regularity, jestliže $g_1, \dots, g_k$ jsou afinní;
- (SPR) Slaterovu podmínku regularity, jestliže $g_1, \dots, g_k$ jsou konvexní na Ω a existuje $x \in \Omega$ tak, že pro každé $i \in \{1, \dots, k\}$ je $g_i(x) < 0$;
- (PLN) Podmínku lineární nezávislosti v $\hat{x} \in M$, jestliže $\{\nabla g_i(\hat{x}) \,|\, i \in \mathcal{I}(\hat{x})\}$ je lineárně nezávislá množina vektorů;
- (ZPR) Zangwillovu podmínku regularity v $\hat{x} \in M$, jestliže $\overline{\mathcal{F}(M;\hat{x})} = \mathcal{G}(\hat{x})$.

Omezení ve tvaru nerovnosti

Platí:

- Každá z podmínek (APR) a (SPR) implikuje (ZPR) pro každé $\hat{x} \in M$.
- Podmínka (PLN) implikuje (ZPR).

Příklad

$$\begin{aligned} \text{minimalizujte} \quad & (x_1+1)^2 + (x_2+1)^2 \\ \text{za podmínek} \quad & x_1 + x_2 \leq 0, \\ & -x_1 \leq 0, \\ & -x_2 \leq 0. \end{aligned}$$

Je splněna afinní podmínka regularity. V každém bodě přípustné množiny jsou tak KKT podmínky nutnými podmínkami optimality.

Omezení ve tvaru nerovnosti

Příklad

$$\begin{aligned} \text{minimalizujte} \quad & x_1 \\ \text{za podmínek} \quad & x_1^2 + (x_2 - 1)^2 \le 1, \\ & x_1^2 + (x_2 + 1)^2 \le 1. \end{aligned}$$

Bod $(0,0)^T$ je jediný přípustný bod, ale není to KKT bod. Splnění některé z podmínek regularity je tedy nutné i v případě konvexních úloh!

Věta (O postačujících KKT podmínkách)

Nechť $\Omega \subseteq \mathbb{R}^n$ je otevřená množina, $f, g_1, \ldots, g_k \in C^1(\Omega)$ jsou konvexní funkce na $C = \{x \in \Omega \mid g_1(x) \le 0, \ldots, g_k(x) \le 0\}$. Jestliže $\hat{x} \in C$ je KKT bod, pak $\hat{x}$ je bod minima funkce f na C.

Důkaz: Viz přednáška. ■

Omezení ve tvaru nerovnosti

Příklad

$$\begin{aligned} \text{minimalizujte } & 2x_1^2 + x_2^2 \\ \text{za podmínek } & x_1^2 + x_2^2 \leq 1, \\ & -x_1 \leq 0. \end{aligned}$$

Je splněna Slaterova podmínka regularity a všechny funkce vystupující v optimalizační úloze jsou spojitě diferencovatelné a konvexní (na $\mathbb{R}^2$). Tedy KKT podmínky jsou nutné a postačující podmínky optimality.

Lze ukázat, že $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ je jediný bod minima.

Omezení ve tvaru rovnosti a nerovnosti

Předpokládejme, že $\Omega \subseteq \mathbb{R}^n$ je neprázdná otevřená množina a $f, g_1, \dots, g_k, h_1, \dots, h_l \in C^1(\Omega)$. Je dána úloha

$$\begin{aligned} \text{minimalizujte } & f(x) \\ \text{za podmínek } & g_i(x) \le 0 \quad \text{pro každé } i \in \{1, \dots, k\}, \\ & h_j(x) = 0 \quad \text{pro každé } j \in \{1, \dots, l\}, \\ & x \in \Omega. \end{aligned}$$

- Rovnice $h_1(x) = 0, \dots, h_l(x) = 0$ lze zapsat pomocí nerovnic $h_1(x) \le 0, -h_1(x) \le 0, \dots, h_l(x) \le 0, -h_l(x) \le 0$, a tím problém převedeme na optimalizační úlohu obsahující jen omezení ve tvaru nerovnosti.
- V případě nelineárních funkcí $h_1, \dots, h_l$ ale není kužel $\mathcal{F}(M; x)$ příliš použitelný (běžně $\mathcal{F}(M; x) = \{0\}$). V takovém případě je nutné vybudovat obecnější teorii.

Omezení ve tvaru rovnosti a nerovnosti

Časté jsou ale úlohy, kdy f je navíc konvexní na přípustné množině a $g_1, \ldots, g_k, h_1, \ldots, h_l$ jsou afinní. Pak

$$\begin{aligned} \text{minimalizujte} \quad & f(x) \\ \text{za podmínek} \quad & g_i(x) \leq 0 \quad \text{pro každé } i \in \{1, \ldots, k\}, \\ & h_j(x) \leq 0 \quad \text{pro každé } j \in \{1, \ldots, l\}, \\ & -h_j(x) \leq 0 \quad \text{pro každé } j \in \{1, \ldots, l\}, \\ & x \in \Omega. \end{aligned}$$

je konvexní úloha a KKT podmínky jsou nutné a postačující v každém bodě přípustné množiny (je splněna afinní podmínka regularity).