

Optimalizace a teorie her

Konvexní funkce

Martin Bohata

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

Definice konvexní funkce

Definice

Nechť $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ a $C \subseteq D$ je neprázdna konvexní množina.

Řekneme, že f je

- ① **konvexní** na C , jestliže pro každé $x, y \in C$ a každé $\lambda \in [0, 1]$ je

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

- ② **ryze konvexní** na C , jestliže pro každé dva různé body $x, y \in C$ a každé $\lambda \in (0, 1)$ je

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y).$$

- ③ **konkávní** (resp. **ryze konkávní**) na C , jestliže $-f$ je konvexní (resp. ryze konvexní) na C .

Úmluva: U pojmů z předchozí definice budeme vynechávat „na C “, jestliže $C = D$.

Základní příklady

Příklad

Z úvodního kurzu matematické analýzy víme, že

- 1 $f(x) = x^2$ je ryze konvexní;
- 2 $f(x) = x^3$ není konvexní ani konkávní. Avšak je ryze konvexní na $[0, \infty)$ a ryze konkávní na $(-\infty, 0]$;
- 3 $f(x) = e^x$ je ryze konvexní;
- 4 $f(x) = \ln x$ je ryze konkávní;
- 5 $f(x) = \sqrt{x}$ je ryze konkávní.

Příklad

- 1 Afinní zobrazení $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (tj. $f(x) = \langle x, a \rangle + b$, kde $a \in \mathbb{R}^n$ a $b \in \mathbb{R}$) je konvexní a také konkávní.
- 2 Funkce $f(x) = \|x\|$ je konvexní.

Dolní úrovněová množina

Definice

Dolní úrovněová množina funkce $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ hladiny $\alpha \in \mathbb{R}$ je množina

$$\text{lev}_{\leq}(f; \alpha) := \{x \in D \mid f(x) \leq \alpha\}.$$

Tvrzení

Je-li f konvexní na $C \subseteq \mathbb{R}^n$, potom $\text{lev}_{\leq}(f|_C; \alpha)$ je konvexní pro každé $\alpha \in \mathbb{R}$.

Důkaz: Viz přednáška. ■

- Opačné tvrzení neplatí (viz např. $f(x) = x^3$).

Příklad

Množina $M = \left\{ x \in \mathbb{R}^2 \mid \|x\| \leq 1, \left\langle x, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\rangle \leq 1 \right\}$ je konvexní.

Tvrzení

Jestliže f, g jsou konvexní funkce na $C \subseteq \mathbb{R}^n$ a $\alpha \geq 0$, potom $f + g$ a αf jsou konvexní na C .

Důkaz: Viz přednáška. ■

Příklad

Funkce

$$f(x) = e^x - 3 \ln x + 2x$$

je konvexní.

- Součin konvexních funkcí již není obecně konvexní funkce ($f(x) = x$ a $g(x) = -x$ jsou konvexní, ale $(fg)(x) = -x^2$ není konvexní).

Operace zachovávající konvexitu

- Skládání konvexních funkcí není obecně konvexní funkce. Funkce $f(x) = x^2$ a $g(x) = x^2 - 1$ jsou konvexní, ale

$$(f \circ g)(x) = f(g(x)) = (x^2 - 1)^2$$

není konvexní.

Tvrzení

Nechť f je konvexní na $K \subseteq \mathbb{R}^m$, $C \subseteq \mathbb{R}^n$ je neprázdňá konvexní a $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ je afinní. Jestliže $g(C) \subseteq K$, pak $f \circ g$ je konvexní na C .

Důkaz: Viz přednáška. ■

Příklad

- 1 Funkce $f(x) = \|Ax + b\|$, kde $A \in \mathbb{M}_{m,n}(\mathbb{R})$ a $b \in \mathbb{R}^m$, je konvexní.
- 2 Funkce $f(x) = e^{x_1 - 2x_2 + 3x_3} + e^{2x_1} - x_3$ je konvexní.

Tvrzení

Jestliže f je konvexní a neklesající na intervalu I , g je konvexní na $C \subseteq \mathbb{R}^n$ a $g(C) \subseteq I$, pak $f \circ g$ je konvexní na C .

Důkaz: Viz přednáška. ■

Příklad

- ❶ Funkce $f(x) = \|x\|^2$ je konvexní.
- ❷ Funkce $f(x) = e^{\|x\|}$ je konvexní.

Definice (lokální extrémy)

Nechť $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, $M \subseteq D$.

- Řekneme, že f **nabývá v $\hat{x} \in M$ lokálního minima** (resp. **ostrého lokálního minima**) **na M** , jestliže existuje $\delta > 0$ tak, že $f(\hat{x}) \leq f(x)$ pro všechna $x \in M \cap U(\hat{x}, \delta)$ (resp. $f(\hat{x}) < f(x)$ pro všechna $x \in M \cap P(\hat{x}, \delta)$).
 - Nabývá-li f v $\hat{x}$ lokálního minima (resp. ostrého lokálního minima), pak $f(\hat{x})$ se nazývá **lokální minimum** (resp. **ostré lokální minimum**) funkce f na M a $\hat{x}$ se nazývá **bod lokálního minima** (resp. **bod ostrého lokálního minima**) funkce f na M .
-
- Úmluva: U pojmů z předchozí definice budeme vynechávat „na M “, jestliže $M = D$.
 - Analogicky definujeme pojmy v případě lokálního maxima.

Body minima konvexních funkcí

Věta (O extrémech a konvexitě)

Nechť f je konvexní na $C \subseteq \mathbb{R}^n$. Potom platí:

- 1 *Každý bod lokálního minima f na C je bodem minima f na C .*
- 2 *Množina $\operatorname{argmin}_{x \in C} f(x)$ je konvexní. Je-li navíc f ryze konvexní na C , pak existuje nejvýše jeden bod minima funkce f na C .*

Důkaz: Viz přednáška. ■

Příklad

Nechť $f(x) = \|Ax\|^2$, kde $A \in \mathbb{M}_n(\mathbb{R})$. V souladu s předchozí větou je

$$\operatorname{argmin}_{x \in \mathbb{R}^n} f(x) = \{x \in \mathbb{R}^n \mid Ax = 0\}$$

konvexní množina.

Opakování

- Gradient:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

- Hessova matice:

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}.$$

Příklad

1. Ať $f(x) = \langle x, c \rangle$, kde $c \in \mathbb{R}^n$. Pak $\nabla f(x) = c$ a $H_f(x) = 0$.
2. Ať $f(x) = \langle Ax, x \rangle$, kde $A \in \mathbb{M}_n(\mathbb{R})$. Pak $\nabla f(x) = Ax + A^T x$ a $H_f(x) = A + A^T$.

Věta (O konvexitě a první derivaci)

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

- ① *f je konvexní na C právě tehdy, když pro každé $x, y \in C$ je*

$$f(x) + \langle \nabla f(x), y - x \rangle \leq f(y).$$

- ② *f je ryze konvexní na C právě tehdy, když pro každé dva různé body $x, y \in C$ je*

$$f(x) + \langle \nabla f(x), y - x \rangle < f(y).$$

Důkaz: Důkaz ① viz přednáška. Důkaz ② vynecháváme. ■

- Geometrická interpretace: graf leží nad tečnou nadrovinou.

Opakování

- Řekneme, že $A \in \mathbb{M}_n(\mathbb{R})$ je **symetrická**, jestliže $A^T = A$.

Příklad

Ať $A \in \mathbb{M}_n(\mathbb{R})$.

- 1 Existují jednoznačně určené matice $B, C \in \mathbb{M}_n(\mathbb{R})$ takové, že $B^T = B$, $C^T = -C$ a $A = B + C$.
 - 2 Existuje symetrická matice $B \in \mathbb{M}_n(\mathbb{R})$ taková, že $\langle Ax, x \rangle = \langle Bx, x \rangle$ pro každé $x \in \mathbb{R}^n$. Tedy v případě funkce $f(x) = \langle Ax, x \rangle$ můžeme vždy bez újmy na obecnosti předpokládat, že A je symetrická.
- Symetrická matice A je **pozitivně semidefinitní**, jestliže $\langle Ax, x \rangle \geq 0$ pro každé $x \in \mathbb{R}^n$.
 - Symetrická matice A je **pozitivně definitní**, jestliže $\langle Ax, x \rangle > 0$ pro každé nenulové $x \in \mathbb{R}^n$.

Opakování

- Nenulový vektor $v \in \mathbb{R}^n$ nazýváme **vlastní vektor matice A** , existuje-li $\lambda \in \mathbb{R}$ tak, že $Av = \lambda v$. Číslo λ se nazývá **vlastní číslo matice A** .
- Je-li A symetrická, pak všechna její vlastní čísla jsou reálná a existuje ortonormální báze v $\mathbb{R}^n$ tvořená vlastními vektory matice A .
- A je pozitivně (resp. negativně) definitní právě tehdy, když má všechna vlastní čísla kladná (resp. záporná).
- A je pozitivně (resp. negativně) semidefinitní právě tehdy, když má všechna vlastní čísla nezáporná (resp. nekladná).
- Ať $I \subseteq \{1, \dots, n\}$ a $A \in \mathbb{M}_n(\mathbb{R})$. Symbolem A_I označíme matici, která vznikne z A vynecháním všech řádků a všech sloupců indexovaných prvky z množiny $\{1, \dots, n\} \setminus I$.

Věta (Sylvesterovo kritérium)

Ať $A \in \mathbb{M}_n(\mathbb{R})$ je symetrická.

- 1 A je pozitivně definitní právě tehdy, když pro každé $k \in \{1, \dots, n\}$ je $\det A_{I_k} > 0$, kde $I_k = \{1, \dots, k\}$.
- 2 A je pozitivně semidefinitní právě tehdy, když $\det A_I \geq 0$ pro každou neprázdnou množinu $I \subseteq \{1, \dots, n\}$.

Důkaz: Vynecháváme. ■

Příklad

Určete typ definitnosti matice A , jestliže

$$\textcircled{1} A = \begin{pmatrix} 15 & 3 & 2 \\ 3 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}; \quad \textcircled{2} A = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 0 \end{pmatrix}; \quad \textcircled{3} A = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -2 & 2 \\ 1 & 2 & -3 \end{pmatrix}.$$

Věta (O Taylorově polynomu 1. a 2. řádu)

Nechť $\Omega \subseteq \mathbb{R}^n$ je otevřená a $x \in \Omega$. Potom platí:

- ❶ *Jestliže $f \in C^1(\Omega)$, pak existuje $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ tak, že $\lim_{h \rightarrow 0} \omega(h) = 0$ a pro každé $h \in \mathbb{R}^n$ splňující $x + h \in \Omega$ je*

$$f(x + h) = f(x) + \langle \nabla f(x), h \rangle + \|h\| \omega(h).$$

- ❷ *Jestliže $f \in C^2(\Omega)$ a $h \in \mathbb{R}^n$ splňuje $[x, x + h] \subseteq \Omega$, pak existuje $\xi \in \{x + \lambda h \mid \lambda \in (0, 1)\}$ tak, že*

$$f(x + h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle H_f(\xi)h, h \rangle.$$

- ❸ *Jestliže $f \in C^2(\Omega)$, pak existuje $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ tak, že $\lim_{h \rightarrow 0} \omega(h) = 0$ a pro každé $h \in \mathbb{R}^n$ splňující $x + h \in \Omega$ je*

$$f(x + h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle H_f(x)h, h \rangle + \|h\|^2 \omega(h).$$

Věta (O konvexitě a druhé derivaci)

Nechť $\Omega \subseteq \mathbb{R}^n$ je otevřená, $C \subseteq \Omega$ neprázdná konvexní a $f \in C^2(\Omega)$.

Potom platí:

- 1 Jestliže pro každé $x \in C$ je $H_f(x)$ pozitivně semidefinitní matice, pak f je konvexní na C .
- 2 Jestliže f je konvexní na C a C je otevřená, potom $H_f(x)$ je pozitivně semidefinitní matice pro každé $x \in C$.
- 3 Jestliže pro každé $x \in C$ je $H_f(x)$ pozitivně definitní matice, pak f je ryze konvexní na C .

Důkaz: Viz přednáška. ■

Příklad

Funkce $f(x, y) = x^2 + xy + y^2$ je ryze konvexní.

Příklad

1 Množina

$$M = \left\{ (x, y)^T \in \mathbb{R}^2 \mid x + 2e^{-x+y^2} \leq 4, -x^2 + 3xy - 3y^2 \geq -1 \right\}$$

je konvexní.

2 Ať $A = \begin{pmatrix} 2 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 2 & \alpha \end{pmatrix}$ a $\alpha \in \mathbb{R}$. Funkce $f(x) = \langle Ax, x \rangle$ je konvexní právě tehdy, když $\alpha \geq 2$.

3 Funkce

$$f(x_1, \dots, x_n) = \sum_{i=1}^n x_i \ln(x_i) - \left(\sum_{i=1}^n x_i \right) \ln \left(\sum_{i=1}^n x_i \right)$$

je konvexní.