Optimalizace

6. Nelineární funkce a zobrazení

Tom Werner, Tom Kroupa

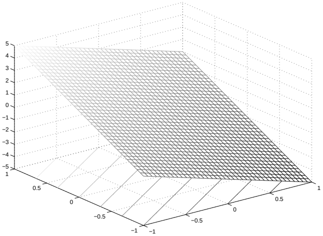
FEL ČVUT

Příklady funkcí $f : \mathbb{R}^n \to \mathbb{R}$

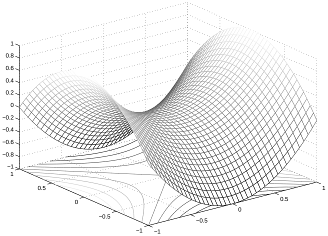
- $f(x_1, x_2) = 3x_1 - x_1^3 - 3x_1x_2^2$
- $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$
- $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$
- skalární pole $\mathbb{R}^3 \to \mathbb{R}$

Zopakujme:

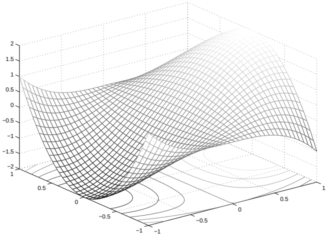
- **Graf** funkce f je množina $\{\, (\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid \mathbf{x} \in \mathbb{R}^n,\ f(\mathbf{x}) = y \,\}$.
- **Vrstevnice** funkce f výšky y je množina $\{\, \mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = y \,\}$.



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



$f(x_1, x_2) = x_1^2 - x_2^2$



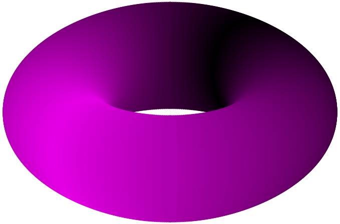
$f(x_1, x_2) = 3x_1 - x_1^3 - 3x_1x_2^2$

Příklady zobrazení $\mathbf{f}\colon \mathbb{R} \to \mathbb{R}^m$

- $\mathbf{f}(t) = \mathbf{x} + t\mathbf{v}$
- $\mathbf{f}(t) = (\cos t,\ \sin t)$
- $\mathbf{f}(t) = (\cos t,\ \sin t,\ t)$
- poloha částice jako funkce času

Příklady zobrazení $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$

- $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$
- vektorové pole $\mathbb{R}^3 \to \mathbb{R}^3$
- $\mathbf{f}(u, v) = \big(\, (R + r\cos v)\cos u,\ (R + r\cos v)\sin u,\ r\sin v \,\big)$

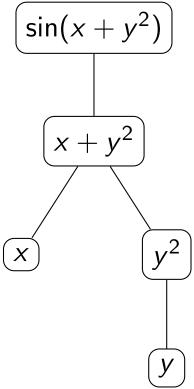


Spojitost

Zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ je **spojité** v bodě $\mathbf{x} \in \mathbb{R}^n$, jestliže

$$\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{f}(\mathbf{x}).$$

Postačující podmínka: Spojitost se zachovává sčítáním/násobením/skládáním funkcí.



Derivace

Funkce $f\colon \mathbb{R} \to \mathbb{R}$ je **diferencovatelná** v bodě x, jestliže existuje $a \in \mathbb{R}$ tak, že

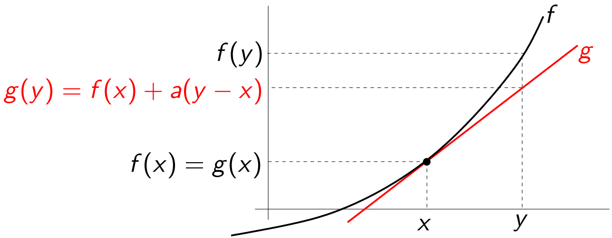
$$\lim_{y \to x} \frac{f(y) - f(x) - a(y - x)}{|y - x|} = 0.$$

Pak se a nazývá **derivace** funkce f v bodě x a píšeme $f'(x) = \dfrac{\mathrm{d}f(x)}{\mathrm{d}x} = a$.

Tvrzení: Tato definice je ekvivalentní obvyklé definici derivace.

Intuice: Funkce f musí být v okolí bodu x podobná nějaké afinní funkci

$$g(y) = f(x) + a(y - x)$$



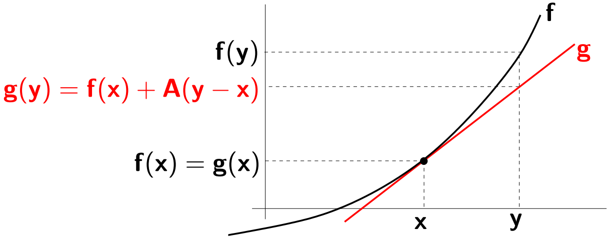
Zobrazení $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$ je **diferencovatelné** v bodě $\mathbf{x}$, jestliže existuje matice $\mathbf{A} \in \mathbb{R}^{m \times n}$ tak, že

$$\lim_{\mathbf{y} \to \mathbf{x}} \frac{\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x}) - \mathbf{A}(\mathbf{y} - \mathbf{x})}{\|\mathbf{y} - \mathbf{x}\|} = \mathbf{0}.$$

Pak se $\mathbf{A}$ nazývá (totální) **derivace** (nebo **Jacobiho matice**) zobrazení $\mathbf{f}$ v bodě $\mathbf{x}$ a píšeme $\mathbf{f}'(\mathbf{x}) = \dfrac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}} = \mathbf{A}$.

Intuice: Zobrazení $\mathbf{f}$ musí být v okolí bodu $\mathbf{x}$ podobné afinnímu zobrazení

$$\mathbf{g}(\mathbf{y}) = \mathbf{f}(\mathbf{x}) + \mathbf{A}(\mathbf{y} - \mathbf{x})$$



Tvrzení

Je-li $\mathbf{f}$ v bodě $\mathbf{x}$ diferencovatelné, pak existují všechny parciální derivace $\frac{\partial f_i(\mathbf{x})}{\partial x_j}$ a

$$\mathbf{f}'(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m\times n}.$$

Speciální případy:

- Pro $f\colon \mathbb{R} \to \mathbb{R}$ je $f'(x)$ skalár
- Pro $\mathbf{f}\colon \mathbb{R} \to \mathbb{R}^m$ je $\mathbf{f}'(x) = \begin{bmatrix} f_1'(x) \\ \vdots \\ f_m'(x) \end{bmatrix}$ sloupcový vektor
- Pro $f\colon \mathbb{R}^n \to \mathbb{R}$ je $f'(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$ řádkový vektor

Věta (postačující podmínka pro diferencovatelnost)

Existují-li v bodě $\mathbf{x}$ všechny parciální derivace $\frac{\partial f_i(\mathbf{x})}{\partial x_j}$ a funkce $\frac{\partial f_i}{\partial x_j}$ jsou v bodě $\mathbf{x}$ spojité, pak je $\mathbf{f}$ v bodě $\mathbf{x}$ diferencovatelné.

Věta (řetízkové pravidlo)

Pro diferencovatelná zobrazení

$$\mathbb{R}^n \xrightarrow{\mathbf{f}} \mathbb{R}^m \xrightarrow{\mathbf{g}} \mathbb{R}^\ell$$

platí

$$(\mathbf{g} \circ \mathbf{f})'(\mathbf{x}) = \frac{\mathrm{d}\mathbf{g}(\mathbf{f}(\mathbf{x}))}{\mathrm{d}\mathbf{x}} = (\mathbf{g}' \circ \mathbf{f})(\mathbf{x})\, \mathbf{f}'(\mathbf{x}) = \mathbf{g}'(\mathbf{f}(\mathbf{x}))\, \mathbf{f}'(\mathbf{x}).$$

V Leibnizově značení: Je-li $\mathbf{y} = \mathbf{g}(\mathbf{u})$ a $\mathbf{u} = \mathbf{f}(\mathbf{x})$, pak

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\mathbf{u}} \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}\mathbf{x}}$$

Přirozeně se zobecní pro složení více zobrazení.

Pravidla pro derivování vektorových/maticových výrazů.

Příklad: Najděme derivaci kvadratické formy (**A** nemusí být symetrická)

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}x_i x_j.$$

Parciální derivace jsou

$$\frac{\partial f(\mathbf{x})}{\partial x_k} = \sum_{j=1}^{n} a_{kj}x_j + \sum_{i=1}^{n} a_{ik}x_i.$$

Tedy

$$f'(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} = \mathbf{x}^T\mathbf{A} + \mathbf{x}^T\mathbf{A}^T = \mathbf{x}^T(\mathbf{A} + \mathbf{A}^T).$$

Derivace některých zobrazení $\mathbb{R}^n \to \mathbb{R}^m$

$\mathbf{f}(\mathbf{x})$	$\mathbf{f}'(\mathbf{x})$
$\mathbf{x}$	$\mathbf{I}$
$\mathbf{A}\mathbf{x}+\mathbf{b}$	$\mathbf{A}$
$\mathbf{A}\mathbf{g}(\mathbf{x})$	$\mathbf{A}\mathbf{g}'(\mathbf{x})$
$\mathbf{g}(\mathbf{A}\mathbf{x}+\mathbf{b})$	$\mathbf{g}'(\mathbf{A}\mathbf{x}+\mathbf{b})\mathbf{A}$

Derivace některých funkcí $f\colon \mathbb{R}^n \to \mathbb{R}$

$f(\mathbf{x})$	$f'(\mathbf{x})$
$\mathbf{a}^T\mathbf{x} + b$	$\mathbf{a}^T$
$\mathbf{x}^T\mathbf{x} = \\|\mathbf{x}\\|^2$	$2\mathbf{x}^T$
$\mathbf{x}^T\mathbf{A}\mathbf{x}$	$\mathbf{x}^T(\mathbf{A} + \mathbf{A}^T)$
$\\|\mathbf{x}\\|$	$\mathbf{x}^T/\\|\mathbf{x}\\|$
$\mathbf{g}(\mathbf{x})^T\mathbf{g}(\mathbf{x}) = \\|\mathbf{g}(\mathbf{x})\\|^2$	$2\mathbf{g}(\mathbf{x})^T\mathbf{g}'(\mathbf{x})$
$\mathbf{g}(\mathbf{x})^T\mathbf{h}(\mathbf{x})$	$\mathbf{g}(\mathbf{x})^T\mathbf{h}'(\mathbf{x}) + \mathbf{h}(\mathbf{x})^T\mathbf{g}'(\mathbf{x})$

Derivace funkce $\mathbb{R}^{m\times n} \to \mathbb{R}$

Derivaci funkce $f\colon \mathbb{R}^{m\times n} \to \mathbb{R}$ v bodě $\mathbf{X} \in \mathbb{R}^{m\times n}$ je přirozené definovat jako matici

$$f'(\mathbf{X}) = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{1n}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{bmatrix} \in \mathbb{R}^{n\times m}$$

protože pro $n = 1$ je to Jacobiho matice zobrazení $f\colon \mathbb{R}^m \to \mathbb{R}$ (tj. řádkový vektor).

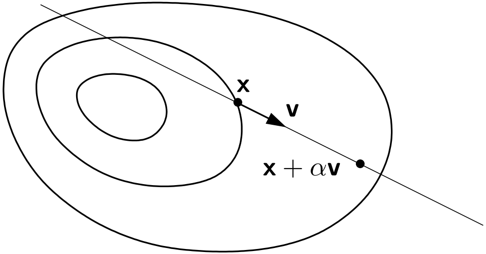
$f(\mathbf{X})$	$f'(\mathbf{X})$	poznámka
$\operatorname{tr}\mathbf{X} = \langle \mathbf{I}, \mathbf{X}\rangle$	$\mathbf{I}$	$\mathbf{X}$ čtvercová
$\operatorname{tr}(\mathbf{A}\mathbf{X}) = \langle \mathbf{A}^T, \mathbf{X}\rangle$	$\mathbf{A}$	
$\lVert\mathbf{X}\rVert^2 = \langle \mathbf{X}, \mathbf{X}\rangle = \operatorname{tr}(\mathbf{X}^T\mathbf{X})$	$2\mathbf{X}^T$	
$\operatorname{tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) = \langle \mathbf{X}, \mathbf{A}\mathbf{X}\rangle$	$\mathbf{X}^T(\mathbf{A} + \mathbf{A}^T)$	
$\mathbf{a}^T\mathbf{X}\mathbf{b}$	$\mathbf{b}\mathbf{a}^T$	
$\operatorname{tr}(\mathbf{A}\mathbf{X}\mathbf{B})$	$\mathbf{B}\mathbf{A}$	
$\operatorname{tr}(\mathbf{X}^k)$	$k\mathbf{X}^{k-1}$	k celé nenulové, pro $k < 0$ je $\mathbf{X}$ regulární
$\det\mathbf{X}$	$(\det\mathbf{X})\mathbf{X}^{-1}$	$\mathbf{X}$ regulární
$\ln\det\mathbf{X}$	$\mathbf{X}^{-1}$	$\mathbf{X}$ positivně definitní

Definice

Směrová derivace zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ v bodě $\mathbf{x} \in \mathbb{R}^n$ ve směru $\mathbf{v} \in \mathbb{R}^n$ je

$$\mathbf{f}_{\mathbf{v}}(\mathbf{x}) = \varphi'_{+}(0) = \lim_{\alpha\to 0^+} \frac{\varphi(\alpha) - \varphi(0)}{\alpha} = \lim_{\alpha\to 0^+} \frac{\mathbf{f}(\mathbf{x} + \alpha\mathbf{v}) - \mathbf{f}(\mathbf{x})}{\alpha}$$

kde $\varphi(\alpha) = \mathbf{f}(\mathbf{x} + \alpha\mathbf{v})$.



Věta

Je-li zobrazení $\mathbf{f}$ v bodě $\mathbf{x}$ (totálně) diferencovatelné, pak $\mathbf{f}_{\mathbf{v}}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\mathbf{v}$.

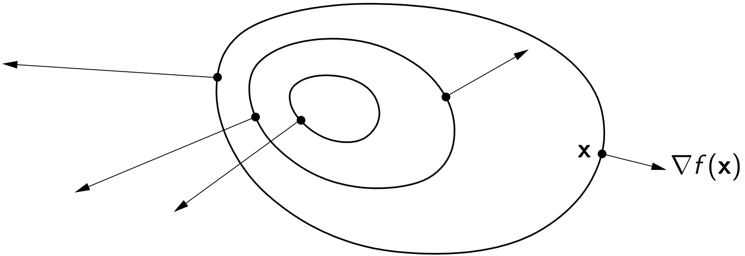
Myšlenka **důkazu:** Zobrazení $\varphi(\alpha) = \mathbf{f}(\mathbf{x} + \alpha\mathbf{v})$ je složením zobrazení $\mathbf{f}$ a zobrazení $g(\alpha) = \mathbf{x} + \alpha\mathbf{v}$.
Věta pak plyne z řetízkového pravidla.

Gradient funkce $f : \mathbb{R}^n \to \mathbb{R}$ je sloupcový vektor

$$\nabla f(\mathbf{x}) = f'(\mathbf{x})^T = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n$$

Směrová derivace $f_{\mathbf{v}}(\mathbf{x}) = f'(\mathbf{x})\mathbf{v} = \nabla f(\mathbf{x})^T \mathbf{v}$ je

- maximální ve směru $\dfrac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$ (za podmínky $\|\mathbf{v}\| = 1$)
- nulová ve směru ortogonálním na $\nabla f(\mathbf{x})$



Parciální derivace druhého řádu

Věta

Nechť $f\colon \mathbb{R}^n \to \mathbb{R}$ a $\mathbf{x} \in \mathbb{R}^n$. Jestliže druhé parciální derivace

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}, \qquad \frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_i}$$

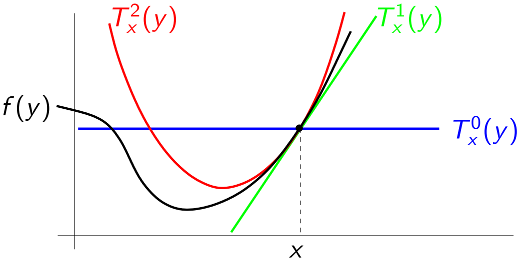
v bodě $\mathbf{x}$ existují a jsou v bodě $\mathbf{x}$ spojité, potom jsou si rovny.

Druhá derivace funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ v bodě $\mathbf{x}$ (**Hessova matice**):

$$f''(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Taylorův polynom funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ stupně k v bodě $\mathbf{x}$ je polynom $T^k_{\mathbf{x}}\colon \mathbb{R}^n \to \mathbb{R}$ stupně k, který má v bodě $\mathbf{x}$ parciální derivace řádu $0, 1, \dots, k$ stejné jako f.

Příklad pro $n = 1$:



Taylorovy polynomy funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ v bodě $\mathbf{x}$ do stupně dva:

$$T^0_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{x})$$

$$T^1_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{x}) + f'(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

$$T^2_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{x}) + f'(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \tfrac{1}{2}(\mathbf{y} - \mathbf{x})^T f''(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

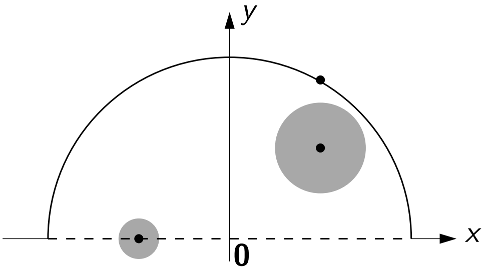
Lokální extrémy

Pro $\mathbf{x} \in \mathbb{R}^n$ a $\epsilon > 0$ označme $B_\epsilon(\mathbf{x}) = \{\, \mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{y}\| < \epsilon \,\}$.

Mějme množinu $X \subseteq \mathbb{R}^n$. Bod $\mathbf{x} \in \mathbb{R}^n$ je její

- **vnitřní bod**, jestliže $B_\epsilon(\mathbf{x}) \subseteq X$ pro nějaké $\epsilon > 0$,
- **hraniční bod**, jestliže $B_\epsilon(\mathbf{x}) \cap X \neq \emptyset$ a $B_\epsilon(\mathbf{x}) \cap (\mathbb{R}^n \setminus X) \neq \emptyset$ pro všechna $\epsilon > 0$.

Příklad: $X = \{\, (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1,\ y > 0 \,\}$



Množina $X \subseteq \mathbb{R}^n$ je

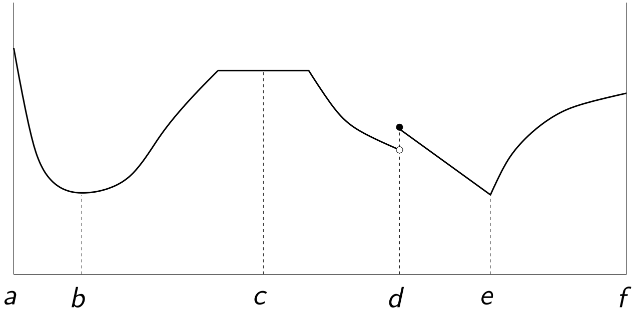
- **otevřená**, jestliže každý její bod je vnitřní,
- **uzavřená**, jestliže obsahuje každý svůj hraniční bod.

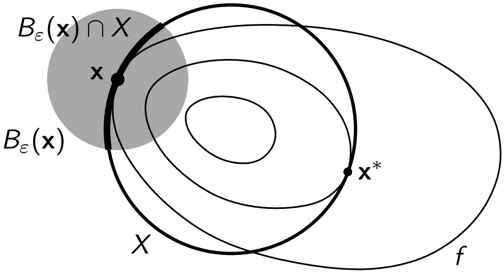
Globální a lokální extrémy funkce na množině

Funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ nabývá na množině $X \subseteq \mathbb{R}^n$ v bodě $\mathbf{x} \in X$

- **(globálního) minima**, jestliže $f(\mathbf{x}) \leq f(\mathbf{x}')$ pro všechna $\mathbf{x}' \in X$,
- **lokálního minima**, jestliže existuje $\epsilon > 0$ takové, že funkce f nabývá v bodě $\mathbf{x}$ (globálního) minima na množině $X \cap B_\epsilon(\mathbf{x})$.

Příklady:





Volné a vázané extrémy

(Globální, lokální) minimum $\mathbf{x}$ funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ na množině $X \subseteq \mathbb{R}^n$ je

- **volné**, pokud $\mathbf{x}$ je vnitřním bodem množiny X,
- **vázané**, pokud $\mathbf{x}$ je hraničním bodem množiny X.

Tvrzení

Nechť $f\colon \mathbb{R}^n \to \mathbb{R}$. Nechť $\mathbf{x}$ je vnitřní bod množiny $X \subseteq \mathbb{R}^n$.
Následující výroky jsou ekvivalentní:

- $\mathbf{x}$ je lokální minimum funkce f na množině X.
- $\mathbf{x}$ je lokální minimum funkce f na množině $\mathbb{R}^n$.

Důkaz implikace $\Rightarrow$ je triviální.
Důkaz $\Leftarrow$: Nechť $\mathbf{x}$ je (globální) minimum f na $X \cap B_\epsilon(\mathbf{x})$.
Protože $\mathbf{x}$ je vnitřní bod X, můžeme ϵ zmenšit tak, aby $B_\epsilon(\mathbf{x}) \subseteq X$.