# Optimalizace

8. Lokální extrémy vázané rovnostmi
(úlohy s omezeními typu rovnosti)

Tom Werner, Tom Kroupa

FEL ČVUT

# Lokální extrémy vázané lineárními rovnostmi

Hledáme extrémy diferencovatelné funkce $f : \mathbb{R}^n \to \mathbb{R}$ na afinním podprostoru

$$X = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b} \}$$

kde $\mathbf{A} \in \mathbb{R}^{m \times n}$. Neboli řešíme úlohu

$$\begin{aligned} \text{min/max} \quad & f(\mathbf{x}) \\ \text{za podmínek} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

# Podmínka prvního řádu

**Věta**

Jestliže $\mathbf{x}$ je lokální extrém funkce $f$ za podmínky $\mathbf{A}\mathbf{x} = \mathbf{b}$, pak $\nabla f(\mathbf{x}) \perp \operatorname{null} \mathbf{A}$.

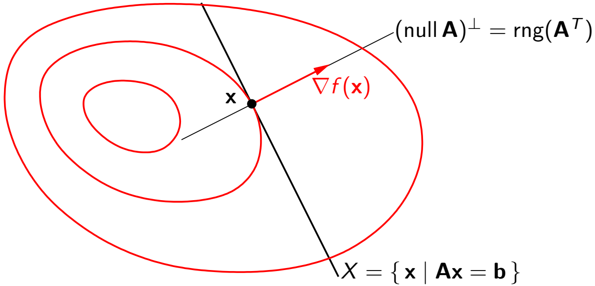
**Důkaz:** Nechť $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$ a $\mathbf{B} \in \mathbb{R}^{m \times k}$ splňuje $\operatorname{rng} \mathbf{B} = \operatorname{null} \mathbf{A}$. Pak

$$\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b} \,\} = \mathbf{x}_0 + \operatorname{null} \mathbf{A} = \{\, \mathbf{x}_0 + \mathbf{B}\mathbf{y} \mid \mathbf{y} \in \mathbb{R}^k \,\}.$$

Teď minimalizujeme funkci $\varphi(\mathbf{y}) = f(\mathbf{x}_0 + \mathbf{B}\mathbf{y})$ bez omezení. Z řetízkového pravidla je

$$\varphi'(\mathbf{y}) = f'(\mathbf{x}_0 + \mathbf{B}\mathbf{y})\mathbf{B} = f'(\mathbf{x})\mathbf{B} = \mathbf{0}.$$

Ale $f'(\mathbf{x})\mathbf{B} = \mathbf{0}$ znamená $\nabla f(\mathbf{x}) \perp \operatorname{null} \mathbf{A}$.




$$\begin{aligned} \nabla f(\mathbf{x}) \perp \operatorname{null}\mathbf{A} &\iff \nabla f(\mathbf{x}) \in (\operatorname{null}\mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T) \\ &\iff f'(\mathbf{x}) = -\boldsymbol{\lambda}^T\mathbf{A} \text{ pro nějaké } \boldsymbol{\lambda} \in \mathbb{R}^m \end{aligned}$$

**Důsledek:** Jestliže $\mathbf{x}$ je lokální extrém, pak existuje $\boldsymbol{\lambda}$ tak, že

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

To je soustava $m+n$ rovnic s $m+n$ neznámými.

## Příklad: Řešení lineární soustavy s nejmenší normou

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|\mathbf{x}\|^2 \\ \text{za podmínek} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

Podmínky prvního řádu:

$$\begin{aligned} \mathbf{x} - \mathbf{A}^T\boldsymbol{\lambda} &= \mathbf{0} \\ \mathbf{A}\mathbf{x} &= \mathbf{b} \end{aligned}$$

Má-li **A** lineárně nezávislé řádky, pak optimálním řešením je

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

## Příklad: Úloha nejmenších čtverců s lineárními omezeními

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \\ \text{za podmínek} \quad & \mathbf{C}\mathbf{x} = \mathbf{d} \end{aligned}$$

Podmínky prvního řádu:

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

**Tvrzení:** Matice $\begin{bmatrix} \mathbf{A}^T\mathbf{A} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix}$ je regulární, právě když $\begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix}$ má lin. nezávislé sloupce a $\mathbf{C}$ má lin. nezávislé řádky.

# Lokální extrémy
# vázané nelineárními rovnostmi

Jsou dána spojitě diferencovatelná $f\colon \mathbb{R}^n \to \mathbb{R}$ a $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$.
Hledáme extrémy funkce $f$ na množině

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

Neboli řešíme úlohu

$$\begin{aligned} \text{min/max} \quad & f(\mathbf{x}) \\ \text{za podmínek} \quad & \mathbf{g}(\mathbf{x}) = \mathbf{0} \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

Lineární omezení jsou speciální případ pro $\mathbf{g}(\mathbf{x}) = \mathbf{b} - \mathbf{A}\mathbf{x}$.

**Intuice pro nalezení podmínek:**
Množinu $X$ linearizujeme v okolí extrému, čímž úlohu převedeme na případ lineárních omezení.

# Tečný prostor k množině přípustných řešení

Jak ‘linearizovat’ množinu $X = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) = \mathbf{0} \,\}$ v okolí daného bodu $\mathbf{x} \in X$?

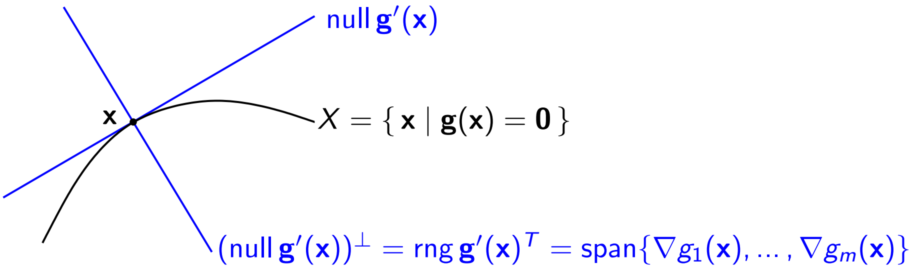
**Intuice:** Hledáme ‘tečný’ prostor k množině $X$ v bodě $\mathbf{x}$.

Aproximujme $\mathbf{g}$ v okolí bodu $\mathbf{x}$ Taylorovým polynomem 1. stupně:

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

Množina $X$ se tím změní na (afinní) podprostor

$$\tilde{X} = \{\, \mathbf{y} \in \mathbb{R}^n \mid \mathbf{g}'(\mathbf{x})(\mathbf{y}-\mathbf{x}) = \mathbf{0} \,\} \;=\; \mathbf{x} + \operatorname{null} \mathbf{g}'(\mathbf{x})$$



## Podmínka regularity

**Potíž:** Ne vždy se podprostor $\tilde{X}$ v okolí bodu $\mathbf{x}$ podobá množině $X$ (přestože zobrazení $\mathbf{T}^1_{\mathbf{x}}$ se v okolí bodu $\mathbf{x}$ podobá zobrazení $\mathbf{g}$)!

**Příklady:** $g(x,y) = xy$ nebo $g(x,y) = x^2$ v bodě $(x,y) = (0,0)$

### Definice

Bod $\mathbf{x} \in \mathbb{R}^n$ je **regulární bod** zobrazení $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$, jestliže řádky Jacobiho matice $\mathbf{g}'(\mathbf{x})$ (tj. gradienty $\nabla g_1(\mathbf{x}), \ldots, \nabla g_m(\mathbf{x})$) jsou lineárně nezávislé.

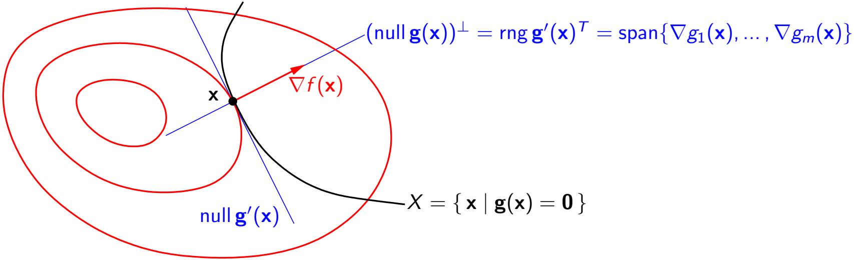
### Tvrzení (neformální)

Je-li $\mathbf{x} \in X$ regulární bod zobrazení $\mathbf{g}$,
pak množina $X$ je v okolí bodu $\mathbf{x}$ 'podobná' afinnímu podprostoru $\tilde{X}$.
Ten má dimenzi $n - m$ a je 'tečný' k množině $X$ v bodě $\mathbf{x}$.

### Věta

Nechť $\mathbf{x}$ je regulární bod zobrazení $\mathbf{g}$.

Jestliže $\mathbf{x}$ je lokální extrém funkce $f$ za podmínky $\mathbf{g}(\mathbf{x}) = \mathbf{0}$, pak $\nabla f(\mathbf{x}) \perp \operatorname{null} \mathbf{g}'(\mathbf{x})$.




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

$$\iff f'(\mathbf{x}) = -\boldsymbol{\lambda}^T \mathbf{g}'(\mathbf{x}) \text{ pro nějaké } \boldsymbol{\lambda} \in \mathbb{R}^m$$ (mínus je konvence)

**Důsledek:** Jestliže $\mathbf{x}$ je lokální extrém, pak existuje $\boldsymbol{\lambda}$ tak, že

$$f'(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}'(\mathbf{x}) = \mathbf{0}$$
$$\mathbf{g}(\mathbf{x}) = \mathbf{0}$$

To je soustava $m + n$ rovnic s $m + n$ neznámými.

# Lagrangeova funkce

Zavedeme-li **Lagrangeovu funkci** $L\colon \mathbb{R}^{m+n} \to \mathbb{R}$

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

pak soustava výše jde napsat jako

$$\begin{aligned} L_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}) &= f'(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}'(\mathbf{x}) &&= \mathbf{0} \\ L_{\boldsymbol{\lambda}}(\mathbf{x}, \boldsymbol{\lambda}) &= \mathbf{g}(\mathbf{x})^T &&= \mathbf{0} \end{aligned}$$

neboli $L'(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$, tj. $(\mathbf{x}, \boldsymbol{\lambda})$ je stacionární bod funkce $L$.

($L_{\mathbf{x}}$ resp. $L_{\boldsymbol{\lambda}}$ značí parciální derivace $L$ podle $\mathbf{x}$ resp. $\boldsymbol{\lambda}$.)

Čísla $(\lambda_1, \ldots, \lambda_m) = \boldsymbol{\lambda}$ se nazývají **Lagrangeovy multiplikátory**.

# Příklad

$$\begin{aligned} \min \quad & x + y \\ \text{za podmínky} \quad & x^2 + y^2 = 1 \end{aligned}$$

Lagrangeova funkce

$$L(x, y, \lambda) = x + y + \lambda(1 - x^2 - y^2)$$

Stacionární podmínky pro $L$:

$$\begin{aligned} L_x(x, y, \lambda) &= 1 - 2\lambda x &&= 0 \\ L_y(x, y, \lambda) &= 1 - 2\lambda y &&= 0 \\ L_\lambda(x, y, \lambda) &= 1 - x^2 - y^2 &&= 0 \end{aligned}$$

Řešení: $(x, y, \lambda) = \pm(1, 1, 1)/\sqrt{2}$.
Našli jsme dva kandidáty na lok. extrémy: $(x, y) = \pm(1, 1)/\sqrt{2}$.

## Příklad: Nesplněné podmínky regularity

$$\begin{aligned} \min \quad & x \\ \text{za podmínky} \quad & x^2 = 0 \end{aligned}$$

Lokální extrém $x = 0$ není regulární bod, protože $g'(x) = 2x = 0$ (tedy $\operatorname{rank} g'(x) < 1$).

Lagrangeova funkce

$$L(x, \lambda) = x + \lambda x^2$$

Podmínky stacionarity

$$\begin{aligned} L_x(x, \lambda) &= 1 + 2x = 0 \\ L_\lambda(x, \lambda) &= x^2 \qquad = 0 \end{aligned}$$

nemají řešení. Lokální extrém jsme nenašli.

## Příklad: Nesplněné podmínky regularity

$$\begin{aligned} \min \quad & x + y \\ \text{za podmínky} \quad & (1 - x^2 - y^2)^2 = 0 \end{aligned}$$

Množina přípustných řešení $X$ je stejná kružnice jako minule, ale gradient funkce $g(x, y) = (1 - x^2 - y^2)^2$ je na kružnici $X$ nulový, tedy žádný bod $X$ není íregulární.

Lagrangeova funkce

$$L(x, y, \lambda) = x + y + \lambda(1 - x^2 - y^2)^2$$

Podmínky stacionarity

$$\begin{aligned} L_x(x, y, \lambda) &= 1 - 4\lambda x(1 - x^2 - y^2) = 0 \\ L_y(x, y, \lambda) &= 1 - 4\lambda y(1 - x^2 - y^2) = 0 \\ L_y(x, y, \lambda) &= (1 - x^2 - y^2)^2 \qquad\quad = 0 \end{aligned}$$

nemají řešení. Lokální extrémy jsme nenašli, i když existují.

## Příklad: Lineární omezení

$$\begin{aligned} \text{min} \quad & f(\mathbf{x}) \\ \text{za podmínky} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

Lagrangeova funkce

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{b} - \mathbf{A}\mathbf{x})$$

Podmínky stacionarity

$$\begin{aligned} L_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}) = f'(\mathbf{x}) - \boldsymbol{\lambda}^T\mathbf{A} &= \mathbf{0} \\ L_{\boldsymbol{\lambda}}(\mathbf{x}, \boldsymbol{\lambda}) = (\mathbf{b} - \mathbf{A}\mathbf{x})^T \quad &= \mathbf{0} \end{aligned}$$

To jsme už odvodili dříve.

Pro lineární omezení jsme ale nemuseli předpokládat rank $\mathbf{g}'(\mathbf{x}) = \text{rank}\,\mathbf{A} = m$.

# Příklad z optiky: Zákon odrazu z Fermatova principu

### Fermatův princip nejkratšího času v optice (1662)

Paprsek mezi dvěma body letí takovou dráhou, aby doba letu byla nejkratší.

Modernější upřesnění: Doba letu je *stacionární* (vzhledem k variacím dráhy).

Křivé zrcadlo je plocha $X = \{\, \mathbf{x} \in \mathbb{R}^n \mid g(\mathbf{x}) = 0 \,\}$ kde $g\colon \mathbb{R}^n \to \mathbb{R}$.
Délka paprsku mezi dvěma body $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ s odrazem od zrcadla v bodě $\mathbf{x} \in X$ je

$$f(\mathbf{x}) = \|\mathbf{x} - \mathbf{a}\| + \|\mathbf{x} - \mathbf{b}\|.$$

Hledáme stacionární bod funkce $f$ za podmínky $g(\mathbf{x}) = 0$.

Lagrangeova funkce

$$L(\mathbf{x}, \lambda) = \|\mathbf{x} - \mathbf{a}\| + \|\mathbf{x} - \mathbf{b}\| + \lambda g(\mathbf{x})$$

Podmínka $L_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$ je (po transpozici)

$$(\mathbf{x} - \mathbf{a})^0 + (\mathbf{x} - \mathbf{b})^0 + \lambda \nabla g(\mathbf{x}) = \mathbf{0} \qquad \text{kde jsme označili } \mathbf{y}^0 = \frac{\mathbf{y}}{\|\mathbf{y}\|}.$$

To říká, že vektor $\nabla g(\mathbf{x})$ (normála k zrcadlu v bodě $\mathbf{x}$) leží v jedné rovině s jednotkovými vektory $(\mathbf{x} - \mathbf{a})^0$ a $(\mathbf{x} - \mathbf{b})^0$ a půlí úhel mezi nimi.