

# Optimalizace

Metody hledání volných lokálních extrémů

---

Tomáš Kroupa, Mirko Navara

24. 4. 2024

FEL ČVUT

Vnitřek a hranice množiny

Extrémy funkce na množině

Podmínka prvního řádu

Parciální derivace druhého řádu

Taylorův polynom pro funkci  $f: \mathbb{R}^n \rightarrow \mathbb{R}$

Podmínky druhého řádu

Iterační metody **minimalizace funkce**

Gradientní metoda **minimalizace funkce**

Newtonova metoda **řešení soustavy rovnic**

Newtonova metoda **minimalizace funkce**

Jak najít sestupný směr diferencovatelné funkce při **minimalizaci**

Prokletí dimenzionality

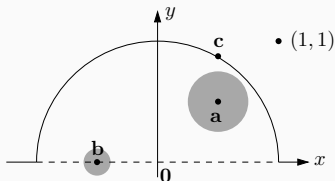
## Vnitřek a hranice množiny

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

- **vnitřní bod**, pokud  $\exists \varepsilon > 0 : B_\varepsilon(\mathbf{x}) \subseteq X$ ,
- **vnější bod**, pokud  $\exists \varepsilon > 0 : B_\varepsilon(\mathbf{x}) \subseteq \mathbb{R}^n \setminus X$ , tj.  $B_\varepsilon(\mathbf{x}) \cap X = \emptyset$ ,
- **hraniční bod**, pokud  $\forall \varepsilon > 0 : B_\varepsilon(\mathbf{x}) \cap X \neq \emptyset, B_\varepsilon(\mathbf{x}) \cap (\mathbb{R}^n \setminus X) \neq \emptyset$ .

**Které body jsou vnitřní a které hraniční?**

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



## Extrémy funkce na množině

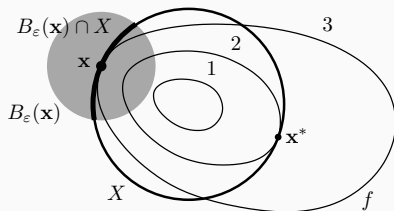
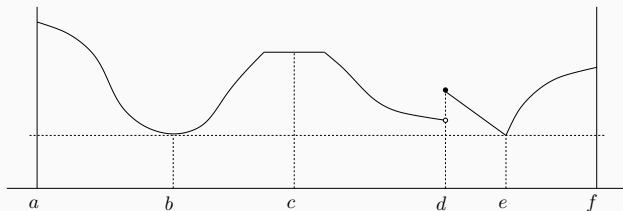
Funkce  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  nabývá na množině  $X \subseteq \mathbb{R}^n$  v bodě  $\mathbf{x} \in X$

- **minima**, pokud  $f(\mathbf{x}) \leq f(\mathbf{y})$  pro všechna  $\mathbf{y} \in X$ ,
- **lokálního minima**, pokud existuje  $\varepsilon > 0$  tak, že  $f(\mathbf{x}) \leq f(\mathbf{y})$  pro všechna  $\mathbf{y} \in X \cap B_\varepsilon(\mathbf{x})$ .

(Lokální) minimum v bodě  $\mathbf{x}$  je

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

# Extrémy funkce na množině – příklady



## Podmínka prvního řádu

**Stacionární bod**  $\mathbf{x} \in \mathbb{R}^n$  funkce  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  splňuje  $f'(\mathbf{x}) = \mathbf{0}$ .

### Věta

Nechť  $X \subseteq \mathbb{R}^n$ ,  $\mathbf{x} \in X$ , je vnitřní bod množiny  $X$  a  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  je diferencovatelná v  $\mathbf{x}$ . Je-li  $\mathbf{x}$  lokální extrém funkce  $f$  na  $X$ , platí

$$f'(\mathbf{x}) = \mathbf{0},$$

tj.  $\mathbf{x}$  je stacionární bod funkce  $f$ .

Stacionární body se celkem dobře hledají, takže mnoho úloh se budeme snažit převést na řešení soustavy rovnic  $f'(\mathbf{x}) = \mathbf{0}$ .

K rozhodnutí, zda stacionární bod je extrém, nám mohou (ale nemusí!) pomoci druhé derivace.

## Parciální derivace druhého řádu

### Věta

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

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = D_j D_i f(\mathbf{x}), \quad \frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_i} = D_i D_j f(\mathbf{x})$$

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

Hessova matice je

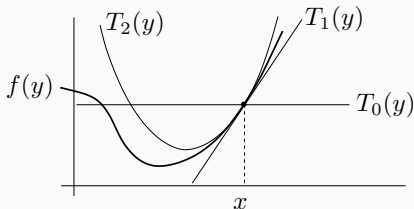
$$f''(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{bmatrix} = \begin{bmatrix} D_1 D_1 f(\mathbf{x}) & \cdots & D_n D_1 f(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ D_1 D_n f(\mathbf{x}) & \cdots & D_n D_n f(\mathbf{x}) \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Obvykle symetrická, je maticí nějaké kvadratické formy.

Determinant Hessovy matice,  $\det f''(\mathbf{x})$ , se nazývá **hessián**.

## Taylorův polynom pro funkci $f: \mathbb{R}^n \rightarrow \mathbb{R}$

Hledáme polynom  $k$ -tého stupně  $T_k: \mathbb{R}^n \rightarrow \mathbb{R}$ , který má v bodě  $\mathbf{x}$  všechny parciální derivace až do řádu  $k$  stejné jako funkce  $f$ .



### Taylorovy polomy v bodě $\mathbf{x}$ do stupně 2

$$T_0(\mathbf{y}) = f(\mathbf{x})$$

$$T_1(\mathbf{y}) = f(\mathbf{x}) + f'(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

$$T_2(\mathbf{y}) = f(\mathbf{x}) + f'(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T f''(\mathbf{x})(\mathbf{y} - \mathbf{x})$$



### Věta

Nechť  $X \subseteq \mathbb{R}^n$ ,  $\mathbf{x} \in X$  je vnitřní bod množiny  $X$  a  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  je dvakrát diferencovatelná v  $\mathbf{x}$ . Je-li  $\mathbf{x}$  stacionární bod, 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, ale **sedlový bod**.

## Iterační metody **minimalizace funkce**

Hledáme lokální minimum funkce  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  pomocí konstrukce posloupnosti bodů  $\mathbf{x}_0, \mathbf{x}_1, \dots \in \mathbb{R}^n$ . Volíme počáteční bod  $\mathbf{x}_0 \in \mathbb{R}^n$ , **směr hledání**  $\mathbf{v}_k \in \mathbb{R}^n$  a parametr  $\alpha_k > 0$ , určující **délku kroku**  $\alpha_k \|\mathbf{v}\|$ .

### Iterace

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{v}_k.$$

- **Sestupná metoda** splňuje  $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$ .
- **Sestupný směr**  $\mathbf{v}_k$  splňuje  $f'(\mathbf{x}_k) \mathbf{v}_k < 0$  (tj.  $\nabla f(\mathbf{x}_k)^T \mathbf{v}_k < 0$ ).

## Iterace gradientní metody

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k) = \mathbf{x}_k - \alpha_k f'(\mathbf{x}_k)^T.$$

- Směr  $-\nabla f(\mathbf{x}_k)$  je sestupný.
- Optimální parametr  $\alpha_k$  lze nalézt minimalizací funkce

$$\varphi_k(\alpha) := f(\mathbf{x}_k + \alpha \mathbf{v}_k),$$

ale často to nestojí za velké úsilí.

- Můžeme zvolit  $\alpha$  velké a exponenciálně zmenšovat, dokud nedocílíme pokles kritéria (pokud každý pokus stojí stejně!).
- Robustní metoda.
- Může konvergovat velmi pomalu.
- Alternativa: **Metoda sdružených gradientů** – zapamatuje si směr předchozích kroků a využije pro další postup.

## Newtonova metoda řešení soustavy rovnic

Hledáme řešení soustavy rovnic  $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ , kde  $\mathbf{g}: \mathbb{R}^n \rightarrow \mathbb{R}^n$  je diferencovatelné zobrazení.

$$T_{1,k}(\mathbf{x}_{k+1}) = \mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = \mathbf{0}.$$

### Iterace

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1} \mathbf{g}(\mathbf{x}_k).$$

- Jacobiho matice  $\mathbf{g}'(\mathbf{x}_k)$  musí být regulární (raději dobře podmíněná).
- Rychlá konvergence, ale nezaručená.
- Nutno začít s dobrou počáteční aproximací řešení,  $\mathbf{x}_0$ .

# Newtonova metoda **minimalizace funkce**

Hledáme možné lokální minimum dvakrát diferencovatelné funkce  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  jako řešení rovnice  $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ , kde  $\mathbf{g}(\mathbf{x}) := \nabla f(\mathbf{x}) = f'(\mathbf{x})^\top$ .

## Iterace Newtonovy metody

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1} \mathbf{g}(\mathbf{x}_k) = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^\top,$$

obecněji

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1} \mathbf{g}(\mathbf{x}_k) = \mathbf{x}_k - \alpha_k f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^\top.$$

Pro srovnání:

## Iterace gradientní metody

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{1} f'(\mathbf{x}_k)^\top.$$

## Newtonova metoda **minimalizace funkce**

Hledáme možné lokální minimum dvakrát diferencovatelné funkce  $f$  jako řešení rovnice  $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ , kde  $\mathbf{g}(\mathbf{x}) := \nabla f(\mathbf{x}) = f'(\mathbf{x})^T$ .

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1} \mathbf{g}(\mathbf{x}_k) = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T.$$

- Používá affinní aproximaci zobrazení  $\mathbf{g} = f'$ , a tedy kvadratickou aproximaci funkce  $f$ .
- Zatímco při řešení obecné soustavy  $\mathbf{g}(\mathbf{x}) = \mathbf{0}$  může být Jacobiho matice  $\mathbf{g}'(\mathbf{x}_k)$  zobrazení  $\mathbf{g}$  libovolná, zde je *symetrická*, neboť je to Hessova matice  $f''(\mathbf{x}_k)$  funkce  $f$ .
- Přesto může být výpočet této matice a její inverze pracný.
- Je-li  $f'(\mathbf{x}_k) = \mathbf{0}$ , pak jsme našli stacionární bod (řešení?).

## Kdy je Newtonův směr sestupný (mimo stacionární body)

Pro  $f'(\mathbf{x}_k) \neq \mathbf{0}$  je **Newtonův směr**  $-f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T$  sestupný, pokud je matice  $f''(\mathbf{x}_k)$  pozitivně definitní.

Obecněji:

### Tvrzení

Je-li  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{x} \neq \mathbf{0}$ , a  $\mathbf{A} \in \mathbb{R}^{n \times n}$  symetrická pozitivně definitní matice, pak úhel  $\angle(\mathbf{x}, \mathbf{A}\mathbf{x})$  je ostrý.

### Důkaz

Výraz

$$\mathbf{x}^T \mathbf{A} \mathbf{x}$$

je současně skalárním součinem vektorů  $\mathbf{x}$ ,  $\mathbf{A}\mathbf{x}$  i pozitivně definitní kvadratickou formou, takže je kladný.

Totéž platí i pro inverzní matici  $\mathbf{A}^{-1}$ .

# Jak snadno najít sestupný směr diferencovatelné funkce

**Pro náhodně zvolený vektor  $\mathbf{v}_k \in \mathbb{R}^n$  vypočteme směrovou derivaci**

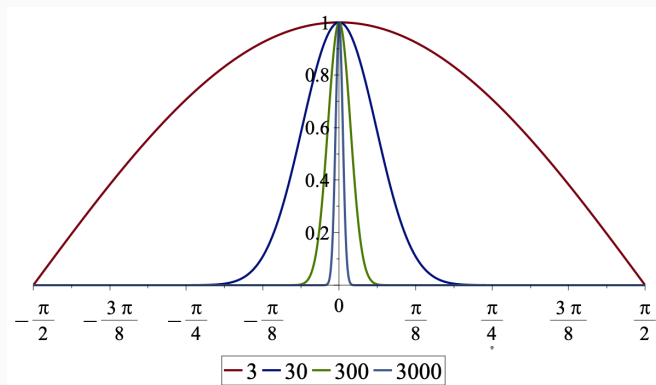
- je-li nulová, možná už jsme ve stacionárním bodě;
- je-li záporná,  $\mathbf{v}_k$  je sestupný směr;
- je-li kladná,  $-\mathbf{v}_k$  je sestupný směr.

Mohli bychom tedy směr volit náhodně, ale snažili jsme se postupovat lépe a rychleji.



# Prokletí dimenzionality

Ve velké dimenzi jsou skoro všechny směry skoro kolmé:



Tvar (neznormované!) hustoty rozdělení zeměpisných šířek z rovnoměrného rozdělení na sféře v  $n$  dimenzích

Alternativa: Soustředíme se na menší počet dimenzí, v tom umíme vybrat směr a nedá to moc práce.






Opakujeme pro další dimenze

⇒ *BCD = block coordinate descent.*

Někdy se dá úloha rozložit na více menších, které se neovlivňují, nebo se aspoň ovlivňují málo.

Vhodné dimenze umí vybrat *PCA = principal component analysis.*

## Reference

-  T. Werner. *Optimalizace* (kapitola 8). Elektronická skripta. FEL ČVUT, 17. 10. 2022. [https://cw.fel.cvut.cz/wiki/\\_media/courses/b0b33opt/opt.pdf](https://cw.fel.cvut.cz/wiki/_media/courses/b0b33opt/opt.pdf)
-  J. Hamhalter a J. Tišer. *Diferenciální počet funkcí více proměnných*. FEL ČVUT, 2005. <https://math.fel.cvut.cz/en/people/tiser/dipocet.pdf>
-  G. Goh. *Why Momentum Really Works*. <https://distill.pub/2017/momentum/>
-  *Matlab Activities for Multivariable Calculus*. <https://mse.redwoods.edu/darnold/math50c/matlab/>
-  M. Navara. *Matematika spojitého světa*. Elektronická skripta. FEL ČVUT, 2011. [https://cmp.felk.cvut.cz/~navara/MSS/MSS\\_print.pdf](https://cmp.felk.cvut.cz/~navara/MSS/MSS_print.pdf)