

# Optimalizace

Metody hledání volných lokálních extrémů – pokračování

---

Tomáš Kroupa, Mirko Navara

31. 3. 2023

FEL ČVUT

## 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  $\mathbf{A}^{-1}$ .

Nechť  $\mathbf{g}: \mathbb{R}^n \rightarrow \mathbb{R}^m$  je diferencovatelné zobrazení. Hledáme **přibližné řešení** přeuročené soustavy  $\mathbf{g}(\mathbf{x}) = \mathbf{0}$  ve smyslu nejmenších čtverců.

**Úloha:** Minimalizujte funkci

$$f(\mathbf{x}) := \|\mathbf{g}(\mathbf{x})\|^2 = \sum_{i=1}^m g_i(\mathbf{x})^2, \quad \mathbf{x} \in \mathbb{R}^n .$$

## Gaussova-Newtonova metoda

Speciální případ: přeuročená nehomogenní soustava lineárních rovnic

$$\mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$$

vede na optimální řešení

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

Zobrazení  $\mathbf{g}$  aproximujeme v okolí bodu  $\mathbf{x}_k$  afinním zobrazením  $\mathbf{T}_1$ ,

$$\mathbf{g}(\mathbf{y}) \approx \mathbf{T}_1(\mathbf{y}) = \underbrace{\mathbf{g}(\mathbf{x}_k)}_{-\mathbf{b}} + \underbrace{\mathbf{g}'(\mathbf{x}_k)}_{\mathbf{A}} \underbrace{(\mathbf{y} - \mathbf{x}_k)}_{\mathbf{x}}$$

a místo hledání minima  $\|\mathbf{g}(\mathbf{y})\|^2$  volíme

$$\begin{aligned} \mathbf{x}_{k+1} - \mathbf{x}_k &= \underset{\mathbf{y}}{\operatorname{argmin}} \|\mathbf{T}_1(\mathbf{y})\|^2 = \mathbf{A}^+ \mathbf{b} = \\ &= -\mathbf{g}'(\mathbf{x}_k)^+ \mathbf{g}(\mathbf{x}_k) = -(\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}'(\mathbf{x}_k))^{-1} \mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}(\mathbf{x}_k). \end{aligned}$$

# Gaussova-Newtonova metoda

Iterace Gaussovy-Newtonovy metody (přibližného řešení soustavy  $m$  rovnic o  $n$  neznámých)

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

Jacobiho matice  $\mathbf{g}'(\mathbf{x}_k)$  musí mít LN sloupce, pak je  $\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}'(\mathbf{x}_k)$  pozitivně definitní.

Speciální případ:

Iterace Newtonovy metody (řešení soustavy  $n$  rovnic o  $n$  neznámých)

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

Minimalizujeme  $f(\mathbf{x}) = \|\mathbf{g}(\mathbf{x})\|^2 = \mathbf{g}(\mathbf{x})^\top \mathbf{g}(\mathbf{x})$ , kde

$$f'(\mathbf{x}) = 2 \mathbf{g}(\mathbf{x})^\top \mathbf{g}'(\mathbf{x}),$$

$$f''(\mathbf{x}) = 2 \mathbf{g}'(\mathbf{x})^\top \mathbf{g}'(\mathbf{x}) + 2 \sum_i g_i(\mathbf{x}) g_i''(\mathbf{x}).$$

## Iterace Gaussovy-Newtonovy metody

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k - (\mathbf{g}'(\mathbf{x}_k)^\top \mathbf{g}'(\mathbf{x}_k))^{-1} \mathbf{g}'(\mathbf{x}_k)^\top \mathbf{g}(\mathbf{x}_k) \\ &= \mathbf{x}_k - (2 \mathbf{g}'(\mathbf{x}_k)^\top \mathbf{g}'(\mathbf{x}_k))^{-1} f'(\mathbf{x}_k)^\top. \end{aligned}$$

## Iterace Newtonovy metody (optimalizační)

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

## Iterace Gaussovy-Newtonovy metody

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (2\mathbf{g}'(\mathbf{x}_k)^\top \mathbf{g}'(\mathbf{x}_k))^{-1} f'(\mathbf{x}_k)^\top .$$

Má-li  $\mathbf{g}'(\mathbf{x}_k)$  LN sloupce, je  $\mathbf{g}'(\mathbf{x}_k)^\top \mathbf{g}'(\mathbf{x}_k)$  pozitivně definitní.

## Iterace gradientní metody

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



# Levenbergova-Marquardtova metoda

## Iterace Gaussovy-Newtonovy metody

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

## Iterace Levenbergovy-Marquardtovy metody

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

Regularizační parametr  $\mu_k > 0$  umožňuje plynule kombinovat mezi

- Gaussovou-Newtonovou metodou (pro  $\mu_k$  malé),
- gradientní metodou (pro  $\mu_k$  velké).

Zvolíme konstantu  $q > 1$ , např. 2, začneme velkou hodnotou  $\mu_0$ ,

- když se účelová funkce sníží, krok přijmeme a  $\mu_{k+1} := \mu_k / q$ ,
- když se účelová funkce zvýší, krok odmítneme a  $\mu_{k+1} := \mu_k q$ .

# Jak snadno najít sestupný směr diferencovatelné funkce

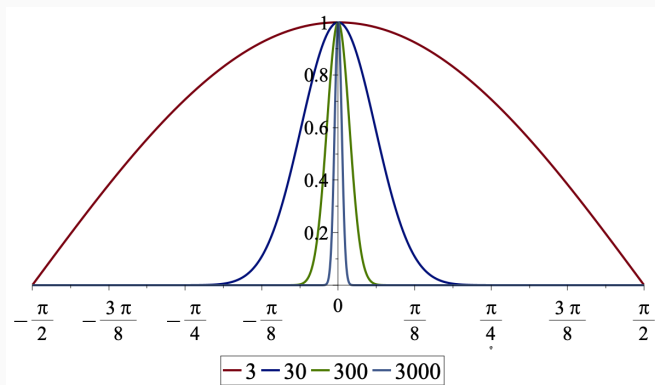
**Pro náhodně zvolený vektor  $\mathbf{v}_k \in \mathbb{R}^n$  vypočteme směrovou derivaci**

- 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)