# Nepodmíněná optimalizace

- Uvažujeme (striktně) konvexní funkce $f$
- Obor hodnot $\mathcal{H} \subset \mathcal{R}$ – jednokriteriální optimalizace
- Definiční obor $\mathcal{O} = \mathcal{R}^N$ – žádné omezující podmínky
- Jedná se o tzv. *nepodmíněnou optimalizaci* – základní úloha matematického programování
- Pro jednoduchost uvažujeme dostatečně hladké funkce
- Základní myšlenka – aproximace kvadratickými funkcemi

# Nepodmíněná optimalizace v jednom rozměru

- Taylorův rozvoj funkce $f$ v bodě $x_0$

$$f(x) \approx f(x_0) + \frac{\mathrm{d}f}{\mathrm{d}x}(x_0)(x - x_0) + \frac{1}{2}\frac{d^2 f}{d\,x^2}(x_0)(x - x_0)^2$$

- V obecném případě by přibyly ještě členy $(\xi - x_0)^3$, kde

$$x_0 \leq \xi \leq x,$$

ty ale můžeme pro $x$ blízko $x_0$ zanedbat.

- Zvolíme optimální řešení $x_{opt}$ jako $x_0$:

$$\begin{aligned} f(x) &\approx f(x_{opt}) + \frac{\mathrm{d}f}{\mathrm{d}x}(x_{opt})(x - x_{opt}) + \frac{1}{2}\frac{d^2 f}{d\,x^2}(x_{opt})(x - x_{opt})^2 \\ &> f(x_{opt}) \end{aligned}$$

- Pro hladkou konvexní funkci je

$$\frac{d^2 f}{d\,x^2}(x_{opt}) > 0$$

- Podmínka optimality

$$\frac{\mathrm{d}f}{\mathrm{d}x}(x_{opt}) = 0$$

- Pokud pro $x_0$ není předchozí podmínka splněna, provedeme opravu

$$\begin{aligned} 0 \quad &= \quad \frac{\mathrm{d}f}{\mathrm{d}x}(x_{opt}) \approx \frac{\mathrm{d}f}{\mathrm{d}x}(x_0) + \frac{d^2 f}{d\,x^2}(x_0)(x_{opt} - x_0) \\ \Rightarrow \quad & \quad x_{opt} \approx x_0 - (\frac{d^2 f}{d\,x^2}(x_0))^{-1}\frac{\mathrm{d}f}{\mathrm{d}x}(x_0) \end{aligned}$$

- Newtonova metoda pro podmínky optimality $\Rightarrow$ velmi rychle konverguje k optimálnímu řešení

- Jak zkonstruovat algoritmus založený na podmínkách optimality?
  1. Nastav $k = 0$ a zvol počáteční bod $x_0$
  2. Urči
  $$g_k = \frac{\mathrm{d}f}{\mathrm{d}x}(x_k)$$
  3. Pokud $|g_k| < $ tolerance $x_k \rightarrow x_{opt}$ a ukonči algoritmus
  4. Urči
  $$h_k = \frac{d^2 f}{d\,x^2}(x_k)$$
  5. Proveď aktualizaci
  $$x_{k+1} = x_k - h_k^{-1} g_k$$
  6. $k = k + 1$, jdi na krok 2

# Nepodmíněná optimalizace ve více rozměrech

- Taylorův rozvoj funkce $f$ v bodě $\mathbf{x}_0$

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + \nabla \mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}\left(\mathbf{x} - \mathbf{x}_0\right)^{\mathsf{T}} \nabla^2 \mathsf{f}(\mathbf{x}_0)\left(\mathbf{x} - \mathbf{x}_0\right)$$

- $\nabla \mathbf{f}$ je *gradient* funkce $f$ [směr největšího spádu]

$$\nabla \mathbf{f}(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{bmatrix} (\mathbf{x})$$

- $\nabla^2\mathsf{f}$ označuje Hessián funkce $f$:

$$\nabla^2\mathsf{f}(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1} & \frac{\partial^2 f}{\partial x_N \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix} (\mathbf{x})$$

- Pro striktně konvexní funkci platí $\mathbf{a}^\mathsf{T}\nabla^2\mathsf{f}(\mathbf{x})\mathbf{a} > 0$ pro $\mathbf{a} \neq \mathbf{0}$, Hessián je tedy *pozitivně definitní* a tudíž invertovatelný

- Opět zvolíme optimální řešení $\mathbf{x}_{opt}$ jako $\mathbf{x}_0$:

$$\begin{aligned} f(\mathbf{x}_{opt}) \quad &\approx \quad f(\mathbf{x}_{opt}) + \nabla \mathbf{f}^\mathsf{T}(\mathbf{x}_{opt})(\mathbf{x} - \mathbf{x}_{opt}) \\ &+ \quad \frac{1}{2}\left(\mathbf{x} - \mathbf{x}_{opt}\right)^\mathsf{T} \nabla^2\mathsf{f}(\mathbf{x}_{\mathsf{opt}})\left(\mathbf{x} - \mathbf{x}_{\mathsf{opt}}\right) > \mathsf{f}(\mathbf{x}_{\mathsf{opt}}) \end{aligned}$$

- Podmínka optimality
$$\nabla \mathbf{f}(\mathbf{x}_{opt}) = \mathbf{0}$$
- Pokud pro $\mathbf{x}_0$ není předchozí podmínka splněna, provedeme lineární opravu
$$\mathbf{x}_{opt} \approx \mathbf{x}_0 - (\nabla^2 \mathsf{f}(\mathbf{x}_0))^{-1} \nabla \mathbf{f}(\mathbf{x}_0)$$
- Algoritmus
  1. Nastav $k = 0$ a zvol počáteční bod $\mathbf{x}_0$
  2. Urči
  $$\mathbf{g}_k = \nabla \mathbf{f}(x_k)$$
  3. Pokud $||g_k|| <$ tolerance $\mathbf{x}_k \rightarrow \mathbf{x}_{opt}$ a ukonči algoritmus
  4. Urči
  $$\mathsf{H}_\mathsf{k} = \nabla^2 \mathsf{f}(\mathbf{x}_\mathsf{k})$$
  5. Proveď aktualizaci
  $$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathsf{H}_\mathsf{k}^{-1} \mathbf{g}_\mathsf{k}$$
  6. $k = k + 1$, jdi na krok 2

# Výhody a nevýhody předchozího algoritmu

- Metoda konverguje velmi rychle
- V každém kroku se řeší *plná* soustava lineárních rovnic $\rightarrow$ může být časově náročné $\rightarrow$ Navrženy iterační metody které "rekonstruují" Hessián [2] nebo ho vůbec nepotřebují – např. metoda sdružených gradientů [3]
- Silně závisí na požadavcích konvexity cílové funkce, pro jiné úlohy může být nestabilní
- *Globální* varianty těchto algoritmů $\rightarrow$ zajištěna konvergence v konečném množství kroků do *lokálního* minima [1, 2]

# Reference

[1] J. F. Bonnans, J. C. Gilbert, J. C. Lemarechal, and C. A. Sagastizabal, *Numerical optimization: Theoretical and practical aspects*, Springer-Verlag, Berlin, Heidelberg, 2004.

[2] L. Lukšan, *Numerické optimalizační metody*, Tech. Report 930, Ústav informatiky AVČR, 2005, `http://www.cs.cas.cz/ics/reports/v930-05.ps`.

[3] J.R. Shewchuk, *An introduction to the conjugate gradient method without the agonizing pain*, Tech. report, Computer Science Division, University of California at Berkeley, 1994, `http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf`.

□

**Prosba.** V případě, že v textu objevíte nějakou chybu nebo budete mít námět na jeho vylepšení, ozvěte se prosím na zemanj@cml.fsv.cvut.cz.

$\boxed{\text{Verze } -003}$