1. Přednáška: úvod

Lidé i příroda optimalizuje:

- továrny minimalizují náklady, maximalizují zisk,
- fyzikální systémy mají tendenci zůstávat ve stavu minimální energie.

Matematická optimalizace (progamování):

- **Cílová funkce** (objective function): kvantitativní míra kvality systému, který studujeme (čas, potenciální energie, zisk, kombinace veličin).
- **Proměnné, neznámé** (variables, unknowns): charakteristiky systému, na kterých cílová funkce závisí.
- **Omezení** (constrains): proměnné nemohou nabývat libovolných hodnot.
- Cílem je nalézt konkrétní hodnoty proměnných, ve kterých nabývá cílová funkce optimální hodnoty (min. nebo max.) $\rightarrow$ **optimalizační proces**
  - **Konstrukce modelu** (rozpoznání cílové funkce, proměnných, omezení) - jednoduchý $\times$ komplexní.
  - Použití **optimalizačních algoritmů** - nejsou univerzální, spíše konstruované pro určité typy úloh, chceme, aby byl rychlý a efektivní.
  - Rozpoznat to, zda algoritmus uspěl (zda jsou splněny **podmínky optimality**), pokud ne, může se řešení použít jako odhad a vylepšit, opravit model (interpretace řešení).

Matematická formulace:
Dáno:

- $x \in \mathbb{R}^n \ldots$ vektor proměnných,
- $f : \mathbb{R}^n \rightarrow \mathbb{R} \ldots$ cílová funkce (obvykle 2x spojitě diferencovatelná),
- $h_1(x), \ldots, h_m(x)$, kde $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \ldots, m$
  $q_1(x), \ldots, q_p(x)$, kde $q_j : \mathbb{R}^n \rightarrow \mathbb{R}$, $j = 1, \ldots, p \ldots$ funkce definující omezení nějakou rovností nebo nerovností (obvykle spojitě diferencovatelné).

Úloha:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{za podmínek} \begin{array}{rcl} h(x) & = & 0 \ldots \text{omezení typu rovnost}, \\ q(x) & \leq & 0 \ldots \text{omezení typu nerovnost}, \end{array}$$

kde

$$h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix}, \quad q(x) = \begin{bmatrix} q_1(x) \\ \vdots \\ q_p(x) \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Příklad: Priklad1.m

$$\min_{x\in\mathbb{R}^n}(x_1-2)^2+(x_2-1)^2 \quad \text{za podmínek} \quad \begin{array}{rcl} x_1+x_2 & \leq & 2, \\ x_1^2-x_2 & \leq & 0. \end{array}$$

Klasifikace úloh:

- podle matematické povahy cílové funkce a omezení (lineární $\to$ lineární programování, kvadratická, konvexní, atd.),
- podle počtu proměnných (velké $\times$ malé úlohy),
- podle hladkosti funkcí (diferencovatelnost).

Klasifikace optimalizace:

- **diskrétní** $\times$ **spojitá** (obvykle jednodušší) - zda proměnné uvažuji diskrétní (např. binární) nebo v reálných číslech,
- **podmíněná** (s omezeními, vazbami) $\times$ **nepodmíněná** (bez omezení, vznikne i z podmíněné optimalizace penalizací zohledňující omezení),
- **globální** (u lineárního nebo konvexního programování lokální = globální, obvykle velmi obtížné) $\times$ **lokální** (obvykle u nelinárních funkcí),
- **stochastická** (závislost na neznámých, např. budoucích hodnotách, ekonomie) $\times$ **deterministická** (model je znám).

Algoritmy matematické optimalizace:

- **Iterační**: začínají počátečním odhadem proměnných x a generují posloupnost dalších (lepších) odhadů (iterací), které zastaví při splnění zastavovací podmínky (pokud možno co nejblíže skutečnému řešení, ideálně v řešení).
- Strategie, jak se posunout - **line search**, trust region.
- Cíle: **Robustní** (lze použít na širokém spektru úloh) x **Efektivní** (nízké výpočetní a paměťové nároky, rychle konvergují) x **Přesný** (nalézt dostatečně přesné řešení bez přílišné citlivosti na chyby v datech, zaokrouhlování v průběhu výpočtu, atd.).

Kalkulus funkce více proměnných

- Nechť $f : \mathbb{R}^n \to \mathbb{R}$ je diferencovatelná, tj. $\frac{\partial f}{\partial x_i}$, $i = 1, \ldots, n$ existují, potom vektor

$$\nabla f = \mathrm{grad} f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

  nazýváme gradient.

- Nechť $f : \mathbb{R}^n \to \mathbb{R}$ je 2x diferencovatelná, tj. $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right)$ existují pro $\forall i, j$, potom matici

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

  nazýváme Hessova matice a její determinant Hessián.

- Pokud je f 2x spojitě diferencovatelná, jsou 2. parciální derivace záměnné (nezávisí na pořadí derivování), tj. $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$.

- Nechť $h : \mathbb{R}^n \to \mathbb{R}^m$ je diferencovatelná, tj. $\frac{\partial h_i}{\partial x_j}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ existují, potom matici

$$J = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix}$$

  nazýváme Jacobiho maticí a její determinant Jacobián.

- Složené funkce:

  - Nechť $f : \mathbb{R}^n \to \mathbb{R}$, x závisí na skalární proměnné $t \in \mathbb{R}$, tj. $x = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}$, potom

$$\frac{d}{dt}\left[f(x(t))\right] = \frac{d}{dt} f\left(x_1(t), \ldots, x_n(t)\right) = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt} = [\frac{dx_1}{dt}, \ldots, \frac{dx_n}{dt}] \cdot \nabla f.$$

  - Nechť $f : \mathbb{R}^n \to \mathbb{R}$, x závisí na vektorové proměnné $t = \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}$, označme $h(t) = f(x(t))$, $h : \mathbb{R}^m \to \mathbb{R}$, potom $\frac{\partial h}{\partial t_j} = [\frac{\partial x_1}{\partial t_j}, \ldots, \frac{\partial x_n}{\partial t_j}] \cdot \nabla f$, $j = 1, \ldots, m$, tj.

vektorově

$$\begin{bmatrix} \frac{\partial h}{\partial t_1} \\ \frac{\partial h}{\partial t_2} \\ \vdots \\ \frac{\partial h}{\partial t_m} \end{bmatrix} = \begin{bmatrix} \frac{\partial x_1}{\partial t_1} & \cdots & \frac{\partial x_n}{\partial t_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial t_m} & \cdots & \frac{\partial x_n}{\partial t_m} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix},$$

$$\nabla_t h = J^T \nabla_x f,$$

kde J je Jacobiho matice zobrazení $x : \mathbb{R}^m \to \mathbb{R}^n$.

- Nechť $f : \mathbb{R}^n \to \mathbb{R}$ je diferencovatelná v bodě x, d je jednotkový směr, potom

$$\frac{\partial f(x)}{\partial d} = D(f,d) = \lim_{\varepsilon \to 0} \frac{f(x+\varepsilon d) - f(x)}{\varepsilon}$$

je derivace funkce f ve směru vektoru d a platí $\frac{\partial f(x)}{\partial d} = \nabla f(x)^T d$.

- **Taylorova věta:** Nechť $f : \mathbb{R}^n \to \mathbb{R}$ je spojitě diferencovatelná v konvexní otevřené množině $\Omega \subset \mathbb{R}^n$. Nechť $x \in \Omega$ a $d \in \mathbb{R}^n$ je takový, že $x + d \in \Omega$, potom existuje $z \in (x, x+d)$ tak, že

$$f(x+d) = f(x) + \nabla f^T(z) d,$$

tj. $f(x+d) = f(x) + \nabla f^T(x+td)d$, pro nějaké $t \in (0,1)$.
Pokud je f 2x spojitě diferencovatelná

$$f(x+d) = f(x) + \int_0^1 \nabla^2 f(x+td) d\, dt$$

a $f(x+d) = f(x) + \nabla f^T(x)d + \frac{1}{2} d^T \nabla^2 f(x+td) d$, pro nějaké $t \in (0,1)$.

Příklady:

Priklady1.nb

Základní vlastnosti řešení

Úloha:

$$\min_{x\in\Omega} f(x),$$

kde $x \in \mathbb{R}^n$, $n \geq 1$ a $f : \mathbb{R}^n \to \mathbb{R}$ je dvakrát spojitě diferencovatelná, Ω je přípustná oblast, pro nepodmíněnou optimalizaci $\Omega = \mathbb{R}^n$.

Definice: *x^* je bod lokálního minima (lokální minimizér) funkce f na Ω, pokud existuje $\varepsilon > 0$ tak, že $f(x) \geq f(x^*)$, $\forall x \in \Omega : |x - x^*| < \varepsilon$. Nastane-li $\forall x \neq x^*$, $|x - x^*| < \varepsilon$ ostrá nerovnost, tj. $f(x) > f(x^*)$, potom jde o bod ostrého lokálního minima.*

x^* je bod globálního minima (globální minimizér) funkce f na Ω, pokud platí $f(x) \geq f(x^*)$, $\forall x \in \Omega$. Pokud platí $f(x) > f(x^*)$, $\forall x \in \Omega, x \neq x^*$, pak x^* je bodem ostrého globálního minima.

- Slabé lokální minimum, izolované lokální minimum.
- Lokální minimum je globálním za určitých předpokladů, kladených na fci f (konvexita) a oblast Ω.
- Úloha $\max_{x\in\Omega} f(x) = \min_{x\in\Omega} -f(x)$.

Jak poznat lokální minimum, aniž bychom zkoušeli všechny funkční hodnoty na nějakém okolí?

Nutné podmínky prvního řádu:

Definice: *Nechť $\Omega \subseteq \mathbb{R}^n$. Vektor d nazveme přípustným směrem v x, pokud $\exists \bar{\alpha} > 0$, takové, že $x + \alpha d \in \Omega$, $\forall \alpha : 0 \leq \alpha \leq \bar{\alpha}$.*

Věta: *Nechť $\Omega \subseteq \mathbb{R}^n$ a nechť $f \in C^1(\Omega)$. Je-li x^* lokální minimum f na Ω, potom pro všechny přípustné směry $d \in \mathbb{R}^n$ v x^* platí*

$$\nabla f(x^*)^T d \geq 0.$$

Důsledek: *Nutná podmínka pro lokální minimum ve vnitřním bodě Ω, nebo pokud $\Omega = \mathbb{R}^n$:*

$$\nabla f(x^*) = 0.$$

Body, ve kterých tato podmínka platí nazýváme stacionární body.

Podmínky druhého řádu:

Věta: *Nechť $\Omega \subseteq \mathbb{R}^n$ a nechť $f \in C^2(\Omega)$. Je-li x^* lokální minimum f na Ω, potom pro všechny přípustné směry $d \in \mathbb{R}^n$ v x^* platí*

- $\nabla f(x^*)^T d \geq 0$,
- *pokud* $\nabla f(x^*)^T d = 0$, *potom* $d^T \nabla^2 f(x^*) d \geq 0$.

Důsledek: *Pro lokální minimum ve vnitřním bodě* Ω*, nebo pokud* $\Omega = \mathbb{R}^n$*:*

- $\nabla f(x^*) = 0$,
- $d^T \nabla^2 f(x^*) d \geq 0$, $\forall d \in \mathbb{R}^n$ *(pozitivně semidefinitní Hessova matice).*

Věta: *(Postačující podmínky) Nechť* $\Omega \subseteq \mathbb{R}^n$*,* $f \in C^2(\Omega)$ *a* x^* *je vnitřním bodem* Ω*. Předpokládejme, že jsou splněny podmínky*

- $\nabla f(x^*) = 0$,
- $H(x^*) = \nabla^2 f(x^*)$ *je pozitivně definitní Hessova matice,*

potom je x^* *bodem ostrého lokálního minima funkce* f *na* Ω.

Poznámky, příklady:

- Sylvestrovo rozhodovací kritérium: pokud Δ_i jsou levé horní subdeterminanty, potom
  - pokud $\Delta_i > 0$ pro $\forall i$ jde o minimum,
  - pokud $\Delta_1 < 0$, $\Delta_2 > 0$, $\Delta_3 < 0$, $\ldots$, $(-1)^n \Delta_n > 0$ jde o maximum.
- Sedlový bod: $\nabla f(x) = 0$, ale funkce v tomto bodě nenabývá ani svého lokálního maxima ani lokálního minima (typ minmax - koňské sedlo, jiné druhy).
- Priklady2.m

Konvexita:

Definice: *Množina* $\Omega \subset \mathbb{R}^n$ *je konvexní, pokud pro* $\forall x, y \in \Omega$ *platí* $\alpha x + (1-\alpha) y \in \Omega$, $\forall \alpha \in \langle 0, 1 \rangle$.

Definice: *Funkce* f *definovaná na konvexní množině* Ω *je konvexní, pokud pro* $\forall x, y \in \Omega$ *platí,* $f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$*, pro* $\forall \alpha \in \langle 0, 1 \rangle$.

Definice: *Funkce* f *definovaná na konvexní množině* Ω *je ostře konvexní, pokud pro* $\forall x, y \in \Omega$*,* $x \neq y$ *platí,* $f(\alpha x + (1-\alpha) y) < \alpha f(x) + (1-\alpha) f(y)$*, pro* $\forall \alpha \in (0, 1)$.

Poznámky:

- f definovaná na konvexní množině Ω je konkávní, pokud $-f$ je konvexní,
- f definovaná na konvexní množině Ω je ostře konkávní, pokud $-f$ je ostře konvexní.

Lemma: *Jsou-li* f_1 *a* f_2 *konvexní na konvexní množině* Ω, *potom i* $f_1 + f_2$ *je konvexní na* Ω.

Lemma: *Je-li* f *konvexní na konvexní množině* Ω, *potom i* $a \cdot f$ *je konvexní na* Ω *pro* $\forall a \geq 0$.

Důsledek: *Nezáporná lineární kombinace konvexních funkcí na konvexní množině je konvexní funkcí.*

Lemma: *Nechť* f *je konvexní funkce na konvexní množině* Ω. *Potom je množina* $\Gamma_c \equiv \{x : x \in \Omega, f(x) \leq c\}$ *konvexní pro* $\forall c \in \mathbb{R}$.

Omezení jsou ve tvaru $q_1(x) \leq c_1, \ldots, q_m(x) \leq c_m$, pokud jsou $q_i(x)$, $\forall i$ konvexní, potom je i přípustná množina konvexní (průnik konvexních množin je konvexní množina).

Věta: *Nechť* f *je konvexní funkce na konvexní množině* Ω. *Potom je množina* Γ, *kde* f *nabývá svého minima, konvexní, a každé lokální minimum funkce* f *je globálním minimem. Pokud je navíc* f *difencovatelná, potom každý stacionární bod je globálním minimem funkce* f.

Priklady2.nb