Co a k čemu je optimalizace

Tomáš Werner, Tomáš Kroupa

26. září 2020

Co je optimalizace?

Co na to slovník:

- Latinsko-anglický slovník: optimus (adj.) =
  - very good, best
  - excellent
  - most beneficial, most advantageous
- Merriam-Webster dictionary: optimization =
  - An act, process, or methodology of making something (such as a design, system, or decision) as fully perfect, functional, or effective as possible.
  - Specifically, the mathematical procedures (such as finding the maximum of a function) involved in this.
- Business Dictionary: optimization =
  Finding an alternative with the most cost effective or highest achievable performance under the given constraints, by maximizing desired factors and minimizing undesired ones.

Mnoho aplikací

- ekonomie a finance: minimální riziko, maximální zisk, nastavení cen, ...
- logistika: doprava, průmysl, zásobování, válka
- řízení (control engineering): výtahu, robota, vlaku, letadla, aktivní budovy, ...
- rozvrhování a plánování (scheduling): školní rozvrh, výrobní kroky, cesta mobilního robota, sled úkonů robotického manipulátoru, aircrew scheduling
- floor planning: návrh integrovaných obvodů (VLSI design) a plošných spojů
- optimalizace kódu: co nejmenší paměť, co nejrychlejší kód
- routing: IDOS, navigace v autě, návrh počítačové sítě, ...
- pravděpodobnost a statistika: princip maximální věrohodnosti, princip maxima entropie, regrese (modelování funkční závislosti náhodných proměnných), rozhodování za neurčitosti
- počítačové vidění: rekonstrukce scény z obrazů (multiview geometry), segmentace obrazu pomocí řezů v grafu (graph cuts), hledání tváří v obraze (AdaBoost), ...
- inteligentní zpracování jiných signálů (např. audio, EKG, EEG): separace zdrojů, auditory scene analysis, ...
- rozpoznávání a strojové učení: minimální trénovací chyba, nejjednodušší model
- návrh mechanických struktur: most, jeřáb, hák, křídlo letadla
- molekulární modelování: např. protein folding
- teorie her
- přiřazování radiových frekvencí v mobilní síti
- ...

Příroda optimalizuje

"Nothing takes place in the world whose meaning is not that of some maximum or minimum."
– Leonhard Euler [1]

- Fyzikální zákony často ve variačním tvaru:
  - systém v rovnováze je v lokálním minimu potenciální energie
  - princip nejkratšího času v optice (Fermat)
  - princip nejmenšího účinku v klasické mechanice: systém se pohybuje ve stavovém prostoru mezi dvěma časy tak, že celkový účinek má minimální (přesněji: lokálně extrémní) hodnotu
- Occamova břitva, K.Popper: příroda si vždy vybírá nejjednodušší model
- organismus si vybírá nejkratší cestu, minimalizuje námahu na dosažení cíle, ...

[1] Citát ukraden ze slajdů předmětu *Optimalizace a teorie her* (A8B01OGT, M.Bohata) na FEL

Matematická optimalizace (= matematické programování)

Obecná formulace optimalizační úlohy:

- Je dána množina přípustných řešení $X \subseteq X_0$ a účelová funkce $f\colon X_0 \to \mathbb{R}$.
- Úkol: najdi minumum funkce f na množině X.

Základní pojmy:

- Funkce f nabývá minima na množině X v prvku $x^* \in X$, když $f(x^*) \leq f(x)$ pro každé $x \in X$. x^* se nazývá argument minima f na X.
- $f(x^*) = \min_{x \in X} f(x)$ se nazývá minimální hodnota f na X.
- $\operatorname{argmin}_{x \in X} f(x)$ je množina všech argumentů minima f na X.
- Analogicky pro maxima. Maxima a minima se dohromady nazývají extrémy.

Příklad:

$X_0 = \{A, B, C, D, E\}$, $X = \{A, B, C, E\}$, $f(A) = 0$, $f(B) = -1$, $f(C) = 5$, $f(D) = -2$, $f(E) = -1$.

- f nabývá na X minima v prvcích B a E
- $\min_{x \in X} f(x) = -1$
- $\operatorname{argmin}_{x \in X} f(x) = \{B, E\}$

Typy optimalizačních úloh

- kombinatorická optimalizace: X konečná (ale obrovská) ($X \subseteq X_0 = \{0,1\}^n$ nebo X obsahuje textové řetězce, grafy, konfigurace Rubikovy kostky, ...).
- spojitá optimalizace: $X \subseteq X_0 = \mathbb{R}^n$ nespočetná ("spojité proměnné") (to znáte z analýzy)
- variační počet: X je množina diferencovatelných funkcí $\varphi\colon D \to \mathbb{R}$, kde $D \subseteq \mathbb{R}^n$ je pevné. Účelová funkce $f\colon X \to \mathbb{R}$ se pak nazývá funkcionál.

V kursu se zaměříme hlavně na spojitou optimalizaci.

Příklad kombinatorické optimalizace: Problém obchodního cestujícího

Máme n měst. Mezi každou dvojicí měst i, j je silnice o známé délce $d(i,j) \geq 0$. Najdi nejkratší možnou trasu, která navštíví každé město právě jednou a vrátí se nazpět do výchozího města.

- X je množina všech permutací n prvků, tj. n-tic $(i_1, \ldots, i_n)$ kde $i_1, \ldots, i_n \in \{1, \ldots, n\}$ jsou různé
- Pro permutaci $(i_1, \ldots, i_n)$ je

$$f(i_1, \ldots, i_n) = \sum_{k=1}^{n-1} d(i_k, i_{k+1}) + d(i_n, i_1)$$

- (Když nějaká dvojice měst i, j není spojena silnicí, nastavíme $d(i,j)$ jako velmi velké číslo.)
- NP-těžká úloha.
  Věří se, že nikdy nebude znám algoritmus, který by řešil NP-těžké úlohy v přijatelném čase (= v čase, který je polynomiální funkcí velikosti úlohy, zde $m + n$).

Příklad spojité optimalizace: Bod na hyperbole nejbližší danému bodu

Najdi bod na hyperbole s rovnicí $xy = 1$, který je nejblíže bodu (x_0, y_0).

- $X = \{ (x, y) \in \mathbb{R}^2 \mid xy = 1 \}$
- $f(x, y) = \sqrt{(x - x_0)^2 + (y - x_0)^2}$

Příklad úlohy variačního počtu: Nejkratší křivka spojující dva body

Najděte nejkratší křivku spojující dva dané body $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ v rovině.
(BÚNO předpokládáme $x_1 \neq x_2$.)

- Křivka spojující dané body je graf diferencovatelné funkce
$$\varphi\colon [x_1, x_2] \to \mathbb{R}$$
takové, že $\varphi(x_1) = y_1$ a $\varphi(x_2) = y_2$.
- Minimalizuj délku křivky
$$f(\varphi) = \int_{x_1}^{x_2} \sqrt{1 + \varphi'(x)^2}\, dx$$
za podmínek $\varphi(x_1) = y_1$ a $\varphi(x_2) = y_2$.
- Řešením je afinní funkce. Grafem této funkce je úsečka procházející danými dvěma body.

Další klasické úlohy z variačního počtu:

- isoperimetrický problém (najdi uzavřenou rovinou křivku dané délky obepínající největší plochu)
- brachistochrona
- řetězovka
- optimální trajektorie chapadla robota za daných okrajových podmínek

Obecná úloha spojité optimalizace

Jsou dány čísla $n, m, \ell \in \mathbb{N}$ a funkce $f, g_1, \ldots, g_m, h_1, \ldots, h_\ell \colon \mathbb{R}^n \to \mathbb{R}$.

- $X \subseteq \mathbb{R}^n$ je množina všech řešení $\mathbf{x} = (x_1, \ldots, x_n)$ soustavy

$$\begin{aligned} g_i(x_1, \ldots, x_n) &\leq 0, \quad i = 1, \ldots, m \\ h_i(x_1, \ldots, x_n) &= 0, \quad i = 1, \ldots, \ell \end{aligned}$$

- Hledáme minimum funkce f na množině X.

Tato úloha se zapisuje také jako

$$\begin{array}{rl} \text{min} & f(x_1, \ldots, x_n) \\ \text{za podmínek} & g_i(x_1, \ldots, x_n) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x_1, \ldots, x_n) = 0, \quad i = 1, \ldots, \ell \\ & x_1, \ldots, x_n \in \mathbb{R} \end{array}$$

nebo kratčeji

$$\min\{\, f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{g}(\mathbf{x}) \leq \mathbf{0},\ \mathbf{h}(\mathbf{x}) = \mathbf{0} \,\}$$

kde $\mathbf{g} \colon \mathbb{R}^n \to \mathbb{R}^m$ a $\mathbf{h} \colon \mathbb{R}^n \to \mathbb{R}^\ell$.

Názvosloví

$$
\begin{array}{rll}
\min & f(\mathbf{x}) & \\
\text{za podmínek} & g_i(\mathbf{x}) \le 0, & i = 1, \ldots, m \\
& h_i(\mathbf{x}) = 0, & i = 1, \ldots, \ell \\
& \mathbf{x} \in \mathbb{R}^n &
\end{array}
$$

- $\mathbf{x} = (x_1, \ldots, x_n)$ jsou proměnné úlohy
- f je účelová funkce (to už známe)
- $g_i(\mathbf{x}) \le 0$ a $h_i(\mathbf{x}) = 0$ jsou omezující podmínky (omezení)
- omezení (nerovnost) $g_i(\mathbf{x}) \le 0$ je aktivní v bodě $\mathbf{x}$, jestliže $g_i(\mathbf{x}) = 0$
- prvky množiny X jsou přípustná řešení úlohy
- prvky množiny $\operatorname{argmin}_{\mathbf{x} \in X} f(\mathbf{x})$ jsou optimální řešení úlohy

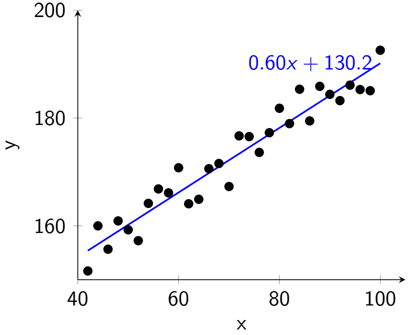
Základní vlastnosti úlohy

Vidíme-li úlohu spojité optimalizace, ihned si položíme otázky:

- Jaký je počet n proměnných úlohy? Jaké jsou tyto proměnné?
- Jaký je počet m, ℓ omezení?
  - $m = \ell = 0$: úloha bez omezení, minimalizujeme f na $\mathbb{R}^n$
- Jaké jsou funkce f, g_i, h_i?
  - f chybí ($=$ je konstantní): úloha na přípustnost
  - Jsou spojité?
  - Jsou všude diferencovatelné? V kterých bodech nejsou?
  - Jsou afinní? (pak jde o lineární optimalizaci/programování)
  - Jsou to polynomy? (pak jde o polynomiální optimalizaci)
  - Jsou konvexní/konkávní? (Jsou-li f, g_i konvexní a h_i afinní, jde o konvexní optimalizační úlohu.)
- Jaká je množina X?
  - Je omezená?
  - Je uzavřená/otevřená?
  - Je konvexní?
- Je úloha přípustná, tj. $X \neq \emptyset$?
- Má úloha optimální řešení?
  - Nemusí mít, např. funkce $f(x) = 1/x$ na intervalu $X = [0, +\infty)$ nemá minimum
  - Postačující podmínka (Weierstrass): Spojitá funkce na uzavřené omezené množině nabývá minima.
- Jak je úloha obtížná? Máme podezření, že je NP-těžká?

Prokládáme body přímkou

Odhadujeme funkční vztah váhy x [kg] a výšky y [cm] z m měření $(x_i, y_i) \in \mathbb{R}^2$:



Hledáme přímku, která 'co nejtěsněji' proloží černé body.

- Vztah má být lineární (přesněji afinní) funkce, tj. $y = \theta_1 + \theta_2 x$ item Formulace ve smyslu nejmenších čtverců: na množině $\mathbb{R}^2$ minimalizuj (bez omezení) funkci

$$f(\boldsymbol{\theta}) = f(\theta_1, \theta_2) = \sum_{i=1}^{m} (\theta_1 + \theta_2 x_i - y_i)^2 = \|\mathbf{A}\boldsymbol{\theta} - \mathbf{y}\|^2$$

- f je konvexní kvadratická funkce dvou proměnných
- Řešení splňují normální rovnice $\mathbf{A}^T\mathbf{A}\boldsymbol{\theta} = \mathbf{A}^T\mathbf{y}$ (důkaz: derivacemi nebo lin. algebrou)

Optimální směs zeleniny

Pro 3 druhy zeleniny jsou dány měrné obsahy živin a minimální požadavky pro jednu přílohu oběda:

	Mrkev	*Bílé zelí*	*Okurka*	**Požadavek**
Vitamín A [mg/kg]	35	0.5	0.28	0.5 mg
Vitamín C [mg/kg]	60	300	80	15 mg
Vláknina [g/kg]	30	20	10	4 g
Cena [Kč/kg]	26	22	60	

Najdi hmotnosti zelenin, které splní výživové požadavky při minimální celkové ceně.

- Formulace problému (lineární program):

$$
\begin{array}{rrcrcrcr}
\text{min} & 26x_1 & + & 22x_2 & + & 60x_3 & & \\
\text{za podmínek} & 35x_1 & + & 0.5x_2 & + & 0.28x_3 & \geq & 0.5 \\
& 60x_1 & + & 300x_2 & + & 80x_3 & \geq & 15 \\
& 30x_1 & + & 20x_2 & + & 10x_3 & \geq & 4 \\
& & & & & x_1, x_2, x_3 & \geq & 0
\end{array}
$$

- Optimální řešení je $x_1 \doteq 0.12$, $x_2 \doteq 0.03$, $x_3 = 0$ za cenu 3.59 Kč.
- Při požadavku $x_3 \geq 0.1$ (okurka!) je řešení $x_1 \doteq 0.097$, $x_2 \doteq 0.004$, $x_3 = 0.1$ za 8.62 Kč.

Aritmetický průměr, těžiště

Jsou dána čísla $a_1, \dots, a_m \in \mathbb{R}$. Najdi minimum (na $\mathbb{R}$) funkce

$$f(x) = \sum_{i=1}^{m} (x - a_i)^2$$

- argument minima je aritmetický průměr $x^* = \frac{1}{m} \sum_{i=1}^{m} a_i$

Jsou dány body $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^n$. Najdi minimum (na $\mathbb{R}^n$) funkce

$$f(\mathbf{x}) = f(x_1, \dots, x_n) = \sum_{i=1}^{m} \|\mathbf{x} - \mathbf{a}_i\|^2$$

- f je kvadratická funkce n proměnných, je konvexní a diferencovatelná
- Je
$$f(x_1, \dots, x_n) = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_j - a_{ij})^2 = \sum_{j=1}^{n} \sum_{i=1}^{m} (x_j - a_{ij})^2 = \sum_{j=1}^{n} f_j(x_j)$$
Tedy minimum funkce f najdeme tak, že najdeme minimum každé funkce f_j zvlášť.
- argument minima je těžiště $\mathbf{x}^* = \frac{1}{m} \sum_{i=1}^{m} \mathbf{a}_i$.

Medián, geometrický medián

Jsou dána čísla $a_1, \ldots, a_m \in \mathbb{R}$. Najdi minimum (na $\mathbb{R}$) funkce

$$f(x) = \sum_{i=1}^{m} |x - a_i|$$

- f je konvexní po částech afinní funkce, není diferencovatelná v bodech a_i
- argument minima je medián čísel $a_1, \ldots, a_m$

Fermat-Weberův problém:
Jsou dány body $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$. Najdi minimum (na $\mathbb{R}^n$) funkce

$$f(\mathbf{x}) = f(x_1, \ldots, x_n) = \sum_{i=1}^{m} \|\mathbf{x} - \mathbf{a}_i\|$$

- f je konvexní funkce n proměnných, není diferencovatelná v bodech $\mathbf{a}_i$
- argument minima se nazývá je geometrický medián

Nejmenší koule obsahující dané body. Optimální umístění

Najdi nejmenší (n-rozměrnou) kouli obsahující dané body $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$.

- Koule v $\mathbb{R}^n$ se středem $\mathbf{x}$ a poloměrem $\sqrt{y}$ je množina $\{\, \mathbf{a} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{a}\|^2 \le y \,\}$.
- Úlohu lze napsat jako

$$\begin{aligned} \min \quad & y \\ \text{za podmínek} \quad & \|\mathbf{x} - \mathbf{a}_i\|^2 \le y, \quad i = 1, \ldots, m \\ & \mathbf{x} \in \mathbb{R}^n \\ & y \in \mathbb{R} \end{aligned}$$

- To je konvexní úloha kvadratického programování (QCQP).

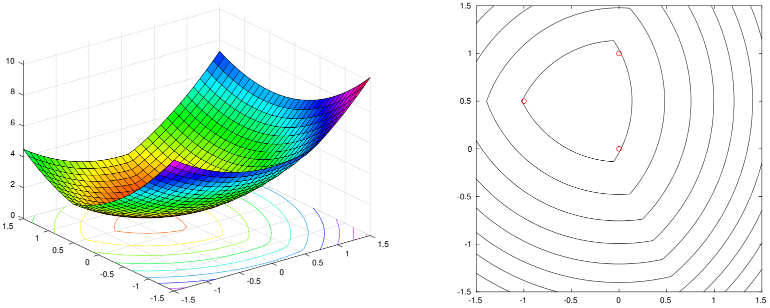
- Tvrzení: Je-li $(\mathbf{x}, y)$ optimální řešení, pak $y = \max_{i=1}^{m} \|\mathbf{x} - \mathbf{a}_i\|^2$.
  Tedy úloha je ekvivalentní minimalizaci funkce

$$f(\mathbf{x}) = \max_{i=1}^{m} \|\mathbf{x} - \mathbf{a}_i\|^2$$

  přes proměnné $\mathbf{x} \in \mathbb{R}^n$ bez omezení.

  Interpretace jako úloha na optimální umístění (pro $n = 2$):
  Najdi místo pro heliport, z něhož dolétne helikoptéra po úsečce do nejvzdálenějšího z daných bodů $\mathbf{a}_1, \ldots, \mathbf{a}_m$ v nejkratším čase.

Graf a vrstevnice funkce f pro $n = 2$, $m = 3$ (body $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ jsou červená kolečka):



- funkce f je konvexní
- funkce f není všude diferencovatelná
- lze ukázat, že f má na $\mathbb{R}^n$ vždy minimum

Shlukování

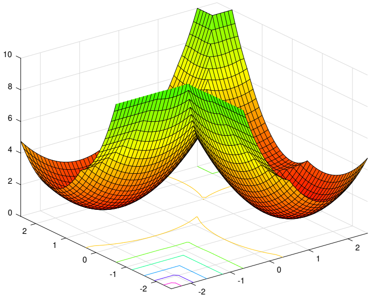
Máme m bodů $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$. Rozmísti dalších k bodů $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ tak, aby průměrná vzdálenost bodu $\mathbf{a}_i$ k nejbližšímu bodu $\mathbf{x}_j$ byla co nejmenší.

- Minimalizujeme (bez omezení) funkci

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \sum_{i=1}^{m} \min_{j=1}^{k} \|\mathbf{a}_i - \mathbf{x}_j\|^2$$

- NP-těžká úloha.

Graf účelové funkce pro $n = 1$, $m = 3$, $k = 2$, $a_1 = -1$, $a_2 = 1$, $a_3 = 2$:

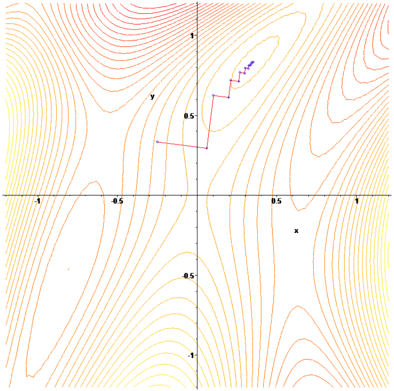


$$f(x_1, x_2) = \sum_{i=1}^{m} \min\{|a_i - x_1|^2, |a_i - x_2|^2\}$$

Lokální minima

Označme kouli v $\mathbb{R}^n$ se středem $\mathbf{x} \in \mathbb{R}^n$ a poloměrem ε jako $B_\varepsilon(\mathbf{x}) = \{\, \mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon \,\}$.

- Minimum funkce na množině se také nazývá globální minimum funkce na množině.
- Funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ na množině $X \subseteq \mathbb{R}^n$ nabývá lokálního minima v bodě $\mathbf{x}^* \in X$, nabývá-li f (globálního) minima na množině $X \cap B_\varepsilon(\mathbf{x}^*)$ pro nějaké $\varepsilon > 0$.



Konvexita

- Množina $X \subseteq \mathbb{R}^n$ je konvexní, jestliže pro každé dva body $\mathbf{x}, \mathbf{y} \in X$ a libovolné $\alpha \in [0, 1]$ platí
$$(1 - \alpha)\mathbf{x} + \alpha\mathbf{y} \in X.$$
- Funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je konvexní na konvexní množině $X \subseteq \mathbb{R}^n$, jestliže pro všechna $\mathbf{x}, \mathbf{y} \in X$ a každé $\alpha \in [0, 1]$ platí
$$f((1 - \alpha)\mathbf{x} + \alpha\mathbf{y}) \leq (1 - \alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}).$$

Theorem

Je-li funkce f konvexní na konvexní množině X, pak každé lokální minimum funkce f na množině X je globální minimum funkce f na množině X.

Historie optimalizace

"Klasická" doba:

- infinitezimální počet, mat. analýza (Newton, Leibniz)
- variační počet (Newton, Leibniz, Euler)
- podmínky na volné lokální extrémy (Fermat)
- podmínky na lokální extrémy vázané rovnostmi (Lagrange)

Moderní optimalizace (po 2. světové válce, s nástupem počítačů):

- lineární programování:
  - teorie, formulace (Kantorovič, Koopmans)
  - simplexová metoda (Dantzig)
  - dualita, teorie her (von Neumann)
- KKT podmínky na lok. extrémy vázané nerovnostmi (Karush-Kuhn-Tucker)
- moderní kombinatorická optimalizace:
  - řezy a toky v grafu (Ford, Fulkerson)
  - celočíselné lin. programování (Gomory, Chvátal), polyhedrální metody
- polynomiální algoritmus na LP (Chačian), algoritmy vnitřního bodu (Karmarkar)
- semidefinitní programování (SDP)

Viz např. `http://www.mitrikitti.fi/opthist.html`

Taxonomie optimalizačních algoritmů

- iterační (např. gradientní metoda, Newtonova metoda) – neiterační
- deterministické – stochastické
- přesné – přibližné
- lokální – globální (globální optimalizace: teorie a algoritmy pro hledání globálních optim složitých nekonvexních opt. úloh)
- centralizované – distribuované (např. výpočty v počítačových sítích, rozvodných elektrických sítích, na mnoha jádrech procesoru, ...)
- sekvenční – paralelní (překryv s centralizované – distribuované)
- heuristické:
  - lokální hledání (hill climbing)
  - hladové (greedy) algoritmy
  - tabu search
  - motivované přírodou: genetické/evoluční, simulované žíhání, kolonie mravenců