Optimalizace

1. Co a k čemu je optimalizace

Tom Werner

FEL ČVUT

Co znamená 'optimalizace' v přirozeném jazyce?

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.

Matematická optimalizace (= matematické programování)

Obecná formulace optimalizační úlohy:

- Dána množina $X \subseteq X'$ **přípustných řešení** a **účelová (cílová, kriteriální, ...) funkce** $f\colon X' \to \mathbb{R}$.
- Najdi minumum funkce f na množině X:

$$\min_{x \in X} f(x)$$

Také se říká: minimalizuj $f(x)$ **za podmínky** $x \in X$.

Matematická optimalizace (= matematické programování)

Obecná formulace optimalizační úlohy:

- Dána množina $X \subseteq X'$ **přípustných řešení** a **účelová (cílová, kriteriální, ...) funkce** $f\colon X' \to \mathbb{R}$.
- Najdi minumum funkce f na množině X:

$$\min_{x \in X} f(x)$$

Také se říká: minimalizuj $f(x)$ **za podmínky** $x \in X$.

Názvosloví a značení:

- 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** funkce f na množině X, nebo **optimální řešení** úlohy.
- $f(x^*) = \min_{x \in X} f(x)$ 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** nebo **optima**.

Ekvivalentní formulace

Nejmenší prvek množiny $Y \subseteq \mathbb{R}$ je prvek $y^* \in Y$ splňující $y^* \leq y$ pro všechna $y \in Y$.
Značíme ho

$$y^* = \min Y$$

Ne každá množina $Y \subseteq \mathbb{R}$ má nejmenší prvek!

Ekvivalentní formulace

Nejmenší prvek množiny $Y \subseteq \mathbb{R}$ je prvek $y^* \in Y$ splňující $y^* \leq y$ pro všechna $y \in Y$.
Značíme ho

$$y^* = \min Y$$

Ne každá množina $Y \subseteq \mathbb{R}$ má nejmenší prvek!

Minimální hodnota funkce $f\colon X' \to \mathbb{R}$ na množině $X \subseteq X'$ je nejmenší prvek množiny

$$f(X) = \{\, f(x) \mid x \in X \,\} \subseteq \mathbb{R}$$

Tedy

$$\min_{x \in X} f(x) = \min f(X) = \min\{\, f(x) \mid x \in X \,\}$$

Neformální dělení optimalizačních úloh

- **kombinatorická optimalizace**: X konečná (ale obvykle obrovská)
  ($X \subseteq \{0, 1\}^n$ nebo X obsahuje permutace, textové řetězce, grafy, konfigurace Rubikovy kostky, ...).

- **spojitá optimalizace**: $X \subseteq \mathbb{R}^n$ ("spojité proměnné"), X je řešení soustavy (ne)rovnic
  (hlavní náplň kursu)

- **variační počet**: X je množina (obvykle diferencovatelných) funkcí $\varphi: D \to \mathbb{R}$, kde $D \subseteq \mathbb{R}^n$ je pevné.
  Účelová funkce $f: X \to \mathbb{R}$ se pak nazývá **funkcionál**.

Příklad kombinatorické optimalizace: Problém obchodního cestujícího

Máme n měst.

Mezi každou dvojicí měst i, j je silnice o známé délce $d(i, j) \geq 0$.

Když nějaká dvojice měst i, j není spojena silnicí, je $d(i, j) = +\infty$.

Najdi nejkratší trasu, která navštíví každé město právě jednou a vrátí se nazpět do výchozího města.

Příklad kombinatorické optimalizace: Problém obchodního cestujícího

Máme n měst.
Mezi každou dvojicí měst i, j je silnice o známé délce $d(i, j) \geq 0$.
Když nějaká dvojice měst i, j není spojena silnicí, je $d(i, j) = +\infty$.
Najdi nejkratší 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

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

- NP-těžká úloha:
  - P$\neq$NP $\implies$ neexistuje algoritmus řešící NP-těžké úlohy v polynomiálním čase (velikosti úlohy, zde n).
  - teorie výpočetní složitosti (computational complexity theory)

Příklad spojité optimalizace: Bod na křivce nejbližší danému bodu

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

- $X = \{\,(x, y) \in \mathbb{R}^2 \mid xy = 1,\ x \geq 0\,\}$
- $f(x, y) = \sqrt{(x - x_0)^2 + (y - x_0)^2}$

Příklad variačního počtu: Nejkratší křivka spojující dva body

Najdi nejkratší křivku spojující dva dané body $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$.

Příklad variačního počtu: Nejkratší křivka spojující dva body

Najdi nejkratší křivku spojující dva dané body $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$.

- X je množina všech diferencovatelných funkcí (předpokládáme $x_1 \neq x_2$)
$$\varphi\colon [x_1, x_2] \to \mathbb{R}$$
splňujících $\varphi(x_1) = y_1$ 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.

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), ...
- **porozumění/zpracování signálu** (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
- ...

Obecná úloha spojité optimalizace

Zopakujme: v optimalizaci hledáme minimum funkce $f\colon X' \to \mathbb{R}$ na množině $X \subseteq X'$.

Ve **spojité optimalizaci** je $X' = \mathbb{R}^n$ a X je množina řešení $\mathbf{x} = (x_1, \ldots, x_n)$ soustavy nerovnic a rovnic

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

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

Obecná úloha spojité optimalizace

Zopakujme: v optimalizaci hledáme minimum funkce $f\colon X' \to \mathbb{R}$ na množině $X \subseteq X'$.
Ve **spojité optimalizaci** je $X' = \mathbb{R}^n$ a X je množina řešení $\mathbf{x} = (x_1, \dots, x_n)$ soustavy nerovnic a rovnic

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

kde $n, m, \ell \in \mathbb{N}$ a $f, g_1, \dots, g_m, h_1, \dots, h_\ell\colon \mathbb{R}^n \to \mathbb{R}$.

Úloha se zapisuje také jako

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

nebo vektorově

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

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

Názvosloví ve spojité optimalizaci

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

- $\mathbf{x} = (x_1, \ldots, x_n)$ jsou **proměnné** úlohy
- f je účelová funkce
  - když f chybí (je konstantní), máme **úlohu na přípustnost**
- $g_i(\mathbf{x}) \leq 0$ a $h_i(\mathbf{x}) = 0$ jsou **omezující podmínky (omezení)**
  - Když omezení chybí ($m = \ell = 0$, tedy $X = \mathbb{R}^n$), máme **úlohu bez omezení**
- $g_i(\mathbf{x}) \leq 0$ jsou **omezení typu nerovnosti** a $h_i(\mathbf{x}) = 0$ jsou **omezení typu rovnosti**
- omezení (nerovnost) $g_i(\mathbf{x}) \leq 0$ 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

Obtížnost optimalizačních úloh

"Téměř všechny" optimalizační úlohy jsou NP-těžké (nebo hůř).

- "prokletí dimenzionality", "kombinatorická exploze", ...
- derivace často nepomohou:
  - podmínka stacionarity vede na soustavu nelineárních (ne)rovnic
  - příliš mnoho stacionárních bodů / lokálních minim
- "Spojitá" formulace úloh zahrnuje i "kombinatorické" úlohy.

  Např. úloha (Maximum Cut)

$$
\begin{aligned}
\min \quad & \textstyle\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \\
\text{za podmínek} \quad & x_i^2 = 1, \quad i = 1, \dots, n \\
& x_i \in \mathbb{R}, \quad i = 1, \dots, n
\end{aligned}
$$

  vypadá spojitě, ale množina X má 2^n izolovaných bodů.

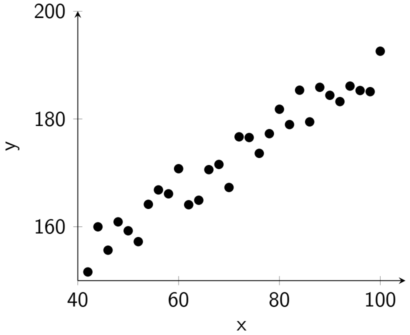
Klíčová role konvexity: **konvexní úlohy** jsou obvykle "snadné"!

Příklady úloh spojité optimalizace

Lineární nejmenší čtverce: Prokládáme body přímkou

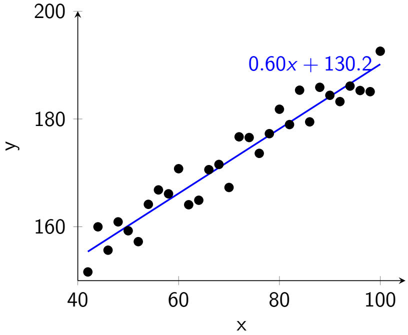
Odhadujeme funkční závislost výšky y [cm] na váze x [kg] člověka z m měření $(x_i, y_i) \in \mathbb{R}^2$:

Hledáme přímku, která "co nejlépe" proloží dané body.



Lineární nejmenší čtverce: Prokládáme body přímkou

Odhadujeme funkční závislost výšky y [cm] na váze x [kg] člověka z m měření $(x_i, y_i) \in \mathbb{R}^2$:
Hledáme přímku, která "co nejlépe" proloží dané body.

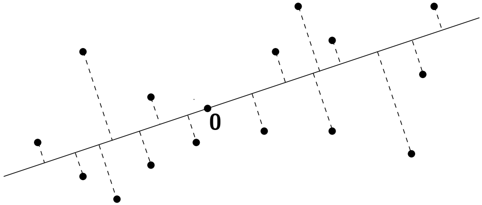


- Vztah má být lineární (přesněji afinní) funkce, tj. $y = \theta_1 + \theta_2 x$
- Formulace ve smyslu **nejmenších čtverců**: na množině $\mathbb{R}^2$ minimalizuj (bez omezení) funkci

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

  (součet čtverců **svislých** vzáleností bodů od přímky)
- f je konvexní kvadratická funkce dvou proměnných
- Podmínka na optimum: **normální rovnice** $\mathbf{A}^T \mathbf{A}\boldsymbol{\theta} = \mathbf{A}^T \mathbf{y}$

SVD: Prokládáme body přímkou/rovinou apod., ale jinak



- Dáno n bodů v $\mathbb{R}^m$, tvořící sloupce matice $\mathbf{A} \in \mathbb{R}^{m \times n}$
- Najdi podprostor dané dimenze $k < m$ tak, aby součet čtverců vzdáleností bodů k němu byl minimální.
- Řešení: pomocí spektrálního rozkladu matice $\mathbf{A}\mathbf{A}^T$, tj. pomocí SVD matice $\mathbf{A}$
- Nejde nijak převést na řešení soustavy lineárních rovnic!

Lineární programování: 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.

Lineární programování: 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**):

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

- Optimální řešení je $x_1 \doteq 0.12$, $x_2 \doteq 0.03$, $x_3 = 0$ za cenu 3.59 Kč.

Lineární programování: 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**):

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

- 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.

Těžiště

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

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

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

Těžiště

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

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

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

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

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

- f je kvadratická funkce n proměnných, je konvexní a diferencovatelná
- Rozpadá se na n nezávislých úloh:

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_j - a_{ij})^2 = \sum_{j=1}^{n} \underbrace{\sum_{i=1}^{m} (x_j - a_{ij})^2}_{f_j(x_j)}$$

- optimální řešení je **těžiště** $\mathbf{x}^* = \frac{1}{m} \sum_{i=1}^{m} \mathbf{a}_i$.

Medián, geometrický medián

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

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

- f je konvexní po částech afinní funkce, není diferencovatelná v bodech 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
- optimální řešení je **medián** čísel $a_1, \ldots, a_m$

Medián, geometrický medián

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

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

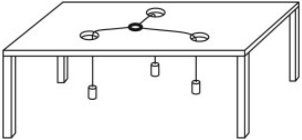
- f je konvexní po částech afinní funkce, není diferencovatelná v bodech a_i
- optimální řešení je **medián** čísel $a_1, \dots, a_m$

Fermat-Weberův problém:

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

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

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



Varignon frame

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.

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.

Shlukování

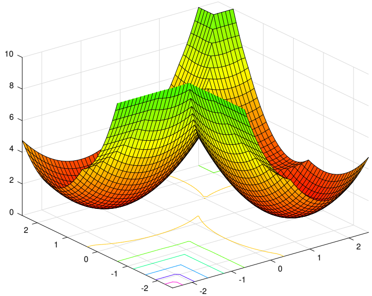
Máme m bodů $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$. Rozmísti dalších k bodů $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ tak, aby průměrná vzdálenost bodu $\mathbf{a}_i$ k nejbližšímu bodu $\mathbf{x}_j$ byla co nejmenší.

- 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 $n = 1$, $m = 3$, $k = 2$, $a_1 = -1$, $a_2 = 1$, $a_3 = 2$:



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

O čem bude B0B33OPT

1. Použití lineární algebry v optimalizaci
    - Lineární úloha nejmenších čtverců
    - PCA
    - Maticové rozklady (QR, spektrální, Cholesky, SVD)
2. Analýza a numerické iterační metody
    - Derivace vektorových a maticových výrazů
    - Podmínky optimality na volné lokální extrémy
    - Iterační numerické metody na hledání lok. minim (gradientní, Newtonova, Gauss-Newtonova)
    - Lok. extrémy vázané rovnostmi (metoda Lagranegeových multiplikátorů)
3. Lineární programování
    - Formulace úloh LP
    - Něco o algoritmech na LP
    - Dualita v LP
4. Konvexní optimalizace
    - Konvexní množiny a funkce
    - Třídy konvexních opt. úloh