Optimalizace

12. Konvexní funkce a konvexní optimalizace

Tomáš Werner, Tomáš Kroupa

FEL ČVUT

Konvexní funkce

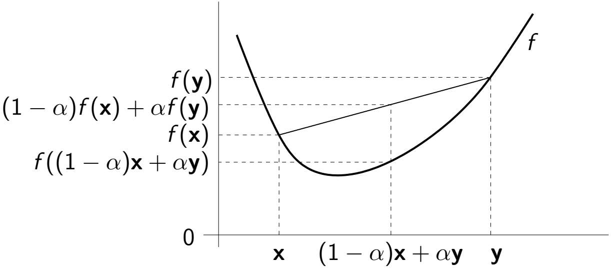
Konvexní funkce

Definice: Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina.

Funkce $f: \mathbb{R}^n \to \mathbb{R}$ je **konvexní** na X, jestliže pro každá $\mathbf{x}, \mathbf{y} \in X$ a $\alpha \in [0, 1]$ je

$$f((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) \leq (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}).$$

Funkce f je **konkávní**, jestliže $-f$ je konvexní.



Ekvivalentní podmínka (**Jensenova nerovnost**):

Pro každé $\mathbf{x}_1, \ldots, \mathbf{x}_k \in X$ a $\alpha_1, \ldots, \alpha_k \geq 0$ kde $\alpha_1 + \cdots + \alpha_k = 1$ platí

$$f(\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k) \leq \alpha_1 f(\mathbf{x}_1) + \cdots + \alpha_k f(\mathbf{x}_k).$$

Příklady konvexních funkcí

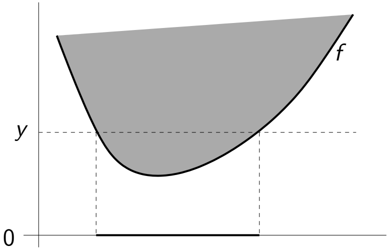
- $f(x) = e^{ax}$ pro $a \in \mathbb{R}$
- $f(x) = |x|^a$ pro $a \geq 1$
- $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ pro $\mathbf{A}$ pozitivně semidefinitní
- $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$
- $f(\mathbf{x}) = \max\{x_1, \ldots, x_n\}$
- libovolná norma

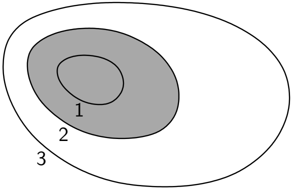
Vztah konvexní funkce a konvexní množiny

- **Epigraf** funkce $f\colon X \to \mathbb{R}$ je množina $\{\, (\mathbf{x}, y) \in X \times \mathbb{R} \mid f(\mathbf{x}) \le y \,\}$.
- **Subkontura** výšky y funkce f je množina $\{\, \mathbf{x} \in X \mid f(\mathbf{x}) \le y \,\}$.

Tvrzení

- Funkce je konvexní, právě když její epigraf je konvexní množina.
- Každá subkontura konvexní funkce je konvexní množina.

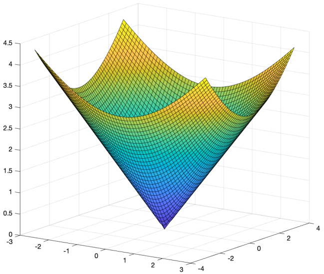


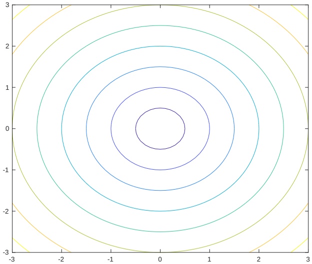


Pozor: Jsou-li všechny subkontury funkce konvexní množiny, funkce nemusí být konvexní.

Příklad: Epigraf a subkontury pro kužel druhého řádu v $\mathbb{R}^2$

- Epigraf eukleidovské normy v $\mathbb{R}^2$ je $K_2^2 = \{ (\mathbf{x}, y) \in \mathbb{R}^3 \mid \|\mathbf{x}\|_2 \leq y \}$
- Subkontury jsou kruhy



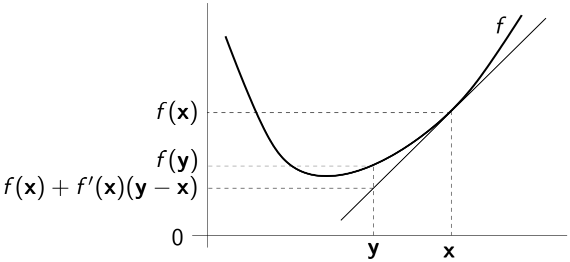


Podmínky na konvexitu pro diferencovatelné funkce

Věta (podmínka prvního řádu)

Diferencovatelná funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je konvexní, právě když pro všechna $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ platí

$$f(\mathbf{y}) \geq f(\mathbf{x}) + f'(\mathbf{x})(\mathbf{x} - \mathbf{y}).$$



Důsledek: Je-li f konvexní diferencovatelná a $f'(\mathbf{x}) = \mathbf{0}$, pak $\mathbf{x}$ je volné globální minimum funkce f.

Věta (podmínka druhého řádu)

Dvakrát diferencovatelná funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je konvexní,
právě když pro každé $\mathbf{x} \in \mathbb{R}^n$ je Hessián $f''(\mathbf{x})$ pozitivně semidefinitní.

Nezáporné lineární kombinace:

Jsou-li $f_1, \ldots, f_k$ konvexní funkce a $\alpha_1, \ldots, \alpha_k \geq 0$, pak funkce $\alpha_1 f_1 + \cdots + \alpha_k f_k$ je konvexní.

Skládání:

- Jestliže $f\colon \mathbb{R} \to \mathbb{R}$ je konvexní a $g\colon \mathbb{R} \to \mathbb{R}$ je konvexní a neklesající, pak $g \circ f$ je konvexní.
- Jestliže $f\colon \mathbb{R}^m \to \mathbb{R}$ je konvexní, pak $f(\mathbf{Ax} + \mathbf{b})$ je konvexní.

Maximum: Nechť $f_i\colon \mathbb{R}^n \to \mathbb{R}$ jsou konvexní funkce pro všechna $i \in I$. Pak

$$f(\mathbf{x}) = \max_{i \in I} f_i(\mathbf{x})$$

je konvexní funkce (předpokládáme, že pro každé $\mathbf{x} \in \mathbb{R}^n$ maximum existuje).

Důkaz: Epigraf funkce f je průnik epigrafů funkcí f_i (není těžké dokázat).
Z toho plyne, že epigraf funkce f je konvexní množina. Tedy f je konvexní funkce.

Příklady:

- $f(\mathbf{x}) = \max\{x_1, \ldots, x_n\}$
- $f(\mathbf{x}) = \max_{i=1}^{k} (\mathbf{a}_i^T \mathbf{x} + b_i)$
- $f(\mathbf{c}) = \max_{\mathbf{x} \in X} \|\mathbf{x} - \mathbf{c}\|$ (kde $X \subseteq \mathbb{R}^n$ je uzavřená a omezená)
- $f(\mathbf{c}) = \max_{\mathbf{x} \in X} \mathbf{c}^T \mathbf{x}$ (kde $X \subseteq \mathbb{R}^n$ je uzavřená a omezená)

Konvexní optimalizační úlohy

Konvexní optimalizační úloha

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

kde $X \subseteq \mathbb{R}^n$ je konvexní množina a $f\colon \mathbb{R}^n \to \mathbb{R}$ je konvexní funkce na množině X.

Věta

Každé lokální minimum funkce f na množině X je globální.

Tvrzení: Optimalizační úloha ve **standardním tvaru**

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

je konvexní, jestliže funkce $f, g_1, \ldots, g_m$ jsou konvexní a funkce $h_1, \ldots, h_\ell$ jsou afinní.

Tato podmínka je postačující ale ne nutná pro konvexitu úlohy. Např. množina

$$X = \{\, (x,y) \in \mathbb{R}^2 \mid (x+y)^2 = 0 \,\}$$

je konvexní, ale funkce $h(x,y) = (x+y)^2$ není afinní.

Třídy konvexních optimalizačních úloh

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

- Lineární programování (**LP**): f, g_i, h_i afinní
- Kvadratické programování (**QP**): f kvadratická konvexní, g_i, h_i afinní
- Kvadratické programování s kvadratickými omezeními (**QCQP**): f, g_i kvadratické konvexní, h_i afinní
- Programování na kuželu druhého řádu (**SOCP**):
  $f, h_1, \ldots, h_\ell$ afinní, každá funkce g_i má tvar $\|\mathbf{A}\mathbf{x} + \mathbf{b}\|_2 - (\mathbf{c}^T\mathbf{x} + d)$
- Semidefinitní programování (**SDP**)
- Geometrické programování (**GP**): h_i afinní, každá z funkci f, g_i má tvar $\sum_{k=1}^{K} \exp(\mathbf{a}_k^T\mathbf{x} + b_k) - c$

Algoritmy vnitřního bodu (Interior Point Algorithms, IPM)

Chceme minimalizovat konvexní funkci $f : \mathbb{R}^n \to \mathbb{R}$ na konvexní množině $X \subseteq \mathbb{R}^n$.

Pro množinu X zvolíme **bariérovou funkci** $B(\mathbf{x})$ s těmito vlastnostmi:

- konvexní a diferencovatelná na vnitřku X
- konečná uvnitř X, směrem k hranici X se blíží $+\infty$

Pro postupně se zmenšující $\mu > 0$ řešíme posloupnost úloh

$$\min_{\mathbf{x} \in \mathbb{R}^n} \; f(\mathbf{x}) + \mu B(\mathbf{x})$$

- úloha bez omezení, lze řešit (přibližně) např. Newtonovou metodou
- pro $\mu \to 0$ je ekvivalentní původní úloze

Příklad: Pro lineární program

$$\min\{\, \mathbf{c}^T \mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b},\; \mathbf{x} \in \mathbb{R}^n \,\}$$

použijeme logaritmickou bariérovou funkci

$$B(\mathbf{x}) = -\sum_{i=1}^{m} \log(\mathbf{a}_i^T \mathbf{x} - b_i)$$

Příklad: Interface solveru SEDUMI na řešení LP, SOCP, SDP

```
% X = SEDUMI(A,b,c,K) minimizes c'*x subject to A*x = b and x is restricted to
%   a self-dual cone, defined by structure K. The fields of K can be
%   K.f, K.l, K.q, K.s (Free, Linear, Quadratic, Semi-definite).
%
%   (1) K.f is the number of FREE, i.e. UNRESTRICTED primal components.
%       These are ALWAYS the first components in x.
%
%   (2) K.l is the number of NONNEGATIVE components. E.g. if K.f=2, K.l=8
%       then x(3:10) >=0.
%
%   (3) K.q lists the dimensions of LORENTZ (quadratic, second-order cone)
%       constraints. E.g. if K.l=10 and K.q = [3 7] then
%           x(11) >= norm(x(12:13)),
%           x(14) >= norm(x(15:20)).
%       These components ALWAYS immediately follow the K.l nonnegative ones.
%
%   (4) K.s lists the dimensions of POSITIVE SEMI-DEFINITE (PSD) constraints
%       E.g. if K.l=10, K.q = [3 7] and K.s = [4 3], then
%           mat( x(21:36),4 ) is PSD,
%           mat( x(37:45),3 ) is PSD.
%       These components are ALWAYS the last entries in x.
```

CVX:

- Pouze konvexní úlohy (*disciplined convex programming, DCP*)
- Verze pro Matlab, Python, ...
- Příklad: úloha

$$\min\{\, \|\mathbf{Ax} - \mathbf{b}\| \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{Cx} = \mathbf{d},\ \|\mathbf{x}\|_\infty \leq 1 \,\}$$

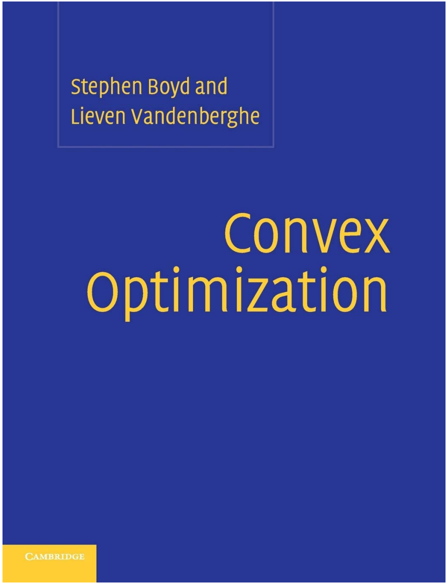
se v CVX (Matlab) modeluje takto:

```
cvx_begin
    variable x(n);
    minimize( norm(A*x-b) );
    subject to
        C*x == d;
        norm(x,Inf) <= 1;
cvx_end
```

YALMIP:

- Konvexní i nekonvexní úlohy (mnohem víc možností než CVX), ale (samozřejmě) jen lokální optima.
- V Matlabu

Kniha o konvexní optimalizaci vhodná pro inženýry



Zdarma ke stažení, spolu se slajdy atd.!

Kvadratické programování (QP)

Funkce f je kvadratická, funkce g_i, h_i jsou afinní.

$$\begin{aligned} \text{min} \quad & \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} \\ \text{za podm.} \quad & \mathbf{C}\mathbf{x} \geq \mathbf{d} \\ & \mathbf{E}\mathbf{x} = \mathbf{f} \end{aligned}$$

Úloha je konvexní, když $\mathbf{A}$ je pozitivně semidefinitní.

Příklady:

- Chybí-li podmínky typu nerovnosti, podmínka optimality je soustava lineárních rovnic (to známe!).
- Řešení přeurčené soustavy s omezeními typu lineárních rovností:

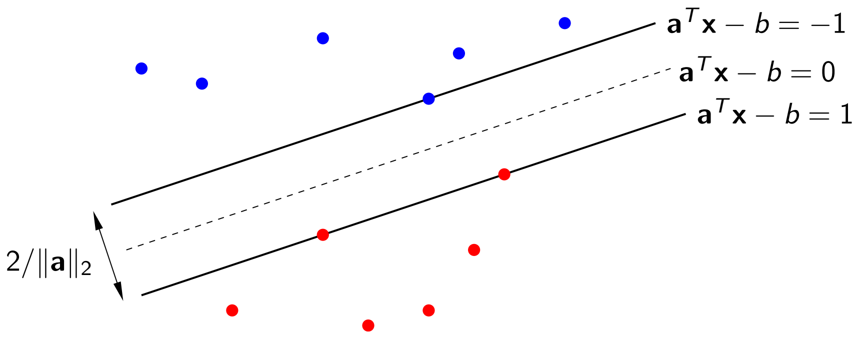
$$\min\{\, \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \mid \mathbf{C}\mathbf{x} = \mathbf{d},\, \mathbf{x} \in \mathbb{R}^n \,\}$$

- Řešení přeurčené lineární soustavy s intervalovými omezeními ('box constraints'):

$$\min\{\, \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \mid \boldsymbol{\ell} \leq \mathbf{x} \leq \mathbf{u},\, \mathbf{x} \in \mathbb{R}^n \,\}$$

- Řešení soustavy lineárních nerovnic s nejmenší normou:

$$\min\{\, \|\mathbf{x}\|_2^2 \mid \mathbf{A}\mathbf{x} \geq \mathbf{b},\; \mathbf{x} \in \mathbb{R}^n \,\}$$



Dány dvojice $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)$, kde $\mathbf{x}_i \in \mathbb{R}^n$ a $y_i \in \{-1, +1\}$.
Hledáme **oddělující nadrovinu**: hledáme $\mathbf{a} \in \mathbb{R}^n$ a $b \in \mathbb{R}$ tak, aby

$$y_i(\mathbf{a}^T\mathbf{x}_i - b) > 0 \quad \forall i = 1, \dots, m$$

Po vydělení $(\mathbf{a}, b)$ vhodným kladným číslem je toto ekvivalentní

$$y_i(\mathbf{a}^T\mathbf{x}_i - b) \geq 1 \quad \forall i = 1, \dots, m$$

Úloha SVM hledá oddělující nadrovinu **maximalizující šířku pásu**, tj. minimalizující $\|\mathbf{a}\|_2^2 = \mathbf{a}^T\mathbf{a}$.

Kvadratické programování s kvadratickými omezeními (QCQP)

Funkce f, g_i jsou kvadratické, h_i jsou afinní.

$$\begin{aligned} \min \quad & \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} \\ \text{za podm.} \quad & \mathbf{x}^T\mathbf{C}_i\mathbf{x} + \mathbf{d}_i^T\mathbf{x} \le e_i, \qquad i = 1, \ldots, m \\ & \mathbf{F}\mathbf{x} = \mathbf{g} \end{aligned}$$

Úloha je konvexní, když matice $\mathbf{A}, \mathbf{C}_1, \ldots, \mathbf{C}_m$ jsou positivně semidefinitní.

Příklad: Nejmenší koule obsahující zadané body $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \max_{i=1}^{m} \|\mathbf{x} - \mathbf{a}_i\|_2$$

To je ekvivalentní QCQP úloze

$$\begin{aligned} \min \quad & y \\ \text{za podm.} \quad & \|\mathbf{x} - \mathbf{a}_i\|_2^2 \le y \end{aligned}$$

Programování na kuželu druhého řádu (SOCP)

$$
\begin{aligned}
\min \quad & \mathbf{g}^T \mathbf{x} \\
\text{za podm.} \quad & \|\mathbf{A}_i \mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{c}_i^T \mathbf{x} + d_i, \qquad i = 1, \ldots, m \\
& \mathbf{E}\mathbf{x} = \mathbf{f}
\end{aligned}
$$

SOCP podmínky jdou napsat jako

$$
\begin{bmatrix} \mathbf{A}_i \\ \mathbf{c}_i \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{b}_i \\ d_i \end{bmatrix} \in K_2^{m_i}
$$

kde $\mathbf{A}_i \in \mathbb{R}^{m_i \times n}$ a

$$
K_2^{m_i} = \{ (\mathbf{x}, y) \in \mathbb{R}^{m_i} \times \mathbb{R} \mid \|\mathbf{x}\|_2 \leq y \}
$$

je epigraf eukleidovské normy $\| \cdot \|_2$ (**kužel druhého řádu**) na $\mathbb{R}^{m_i}$.

- Když $\mathbf{A}_i = \mathbf{0}$ pro všechna i, SOCP se redukuje na LP.
- Když $\mathbf{c}_i = \mathbf{0}$ pro všechna i, SOCP se redukuje na QCQP.

Příklad: Geometrický medián (Fermat-Weberův problém)

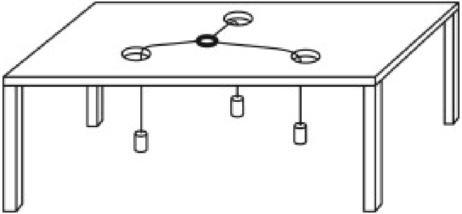
Pro zadané body $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ najděte minimum funkce

$$f(\mathbf{x}) = \sum_{i=1}^{m} \|\mathbf{x} - \mathbf{a}_i\|_2$$

SOCP formulace:

$$\begin{aligned} \min \quad & y_1 + \cdots + y_m \\ \text{za podmínek} \quad & \|\mathbf{x} - \mathbf{a}_i\|_2 \leq y_i, \quad i = 1, \ldots, m. \end{aligned}$$

Existuje na to speciální algoritmus (Weiszfeldův).



Varignon frame

Semidefinitní programování (SDP)

Věta

Množina všech positivně semidefinitních matic pevného rozměru $n \times n$ je konvexní kužel.

Důkaz: Pro každé $\mathbf{x} \in \mathbb{R}^n$ a $\alpha, \beta \geq 0$ platí $\mathbf{x}^T(\alpha\mathbf{A} + \beta\mathbf{B})\mathbf{x} = \alpha\mathbf{x}^T\mathbf{A}\mathbf{x} + \beta\mathbf{x}^T\mathbf{B}\mathbf{x} \geq 0$.

Úloha SDP:

$$\begin{aligned} \text{min} \quad & \langle \mathbf{C}, \mathbf{X} \rangle \\ \text{za podmínek} \quad & \langle \mathbf{A}_i, \mathbf{X} \rangle = b_i, \quad i = 1, \dots, m \\ & \mathbf{X} \in \mathbb{R}^{n \times n} \text{ je positivně semidefinitní} \end{aligned}$$

Tedy minimalizujeme lineární funkci p.s.d. matice za podmínek lineárních rovností.

- Pro diagonální matice $\mathbf{A}_i, \mathbf{C}$ se SDP redukuje na LP.
- SOCP lze reprezentovat pomocí SDP (neuvádíme).
- Souhrn:

$$\text{LP} \subset \text{QP} \subset \text{QCQP} \subset \text{SOCP} \subset \text{SDP}$$

Co s nekonvexními úlohami?

Některé nekonvexní úlohy jsou ‘snadné’ ...

Příklad: Minimalizace (ne nutně p.s.d.) kvadratické formy na sféře:

$$\begin{aligned} \min \quad & \mathbf{x}^T \mathbf{A} \mathbf{x} \\ \text{za podmínky} \quad & \mathbf{x}^T \mathbf{x} = 1 \end{aligned}$$

Příklad: QCQP s jediným omezením:

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{za podmínky} \quad & g(\mathbf{x}) = 0 \quad (\text{příp. } g(\mathbf{x}) \leq 0) \end{aligned}$$

kde $f, g\colon \mathbb{R}^n \to \mathbb{R}$ jsou dané (ne nutně konvexní) kvadratické funkce.

... ale většina ne

Příklad: Minimalizace (ne nutně p.s.d.) kvadratické formy na hyperkrychli:

$$\begin{aligned} \min \quad & \mathbf{x}^T \mathbf{A} \mathbf{x} \\ \text{za podmínky} \quad & \mathbf{x} \in [-1, 1]^n \end{aligned}$$

Pro $\mathbf{A} = -\mathbf{I}$ má úloha 2^n lokálních minim (vrcholy hyperkrychle).
Pro obecné $\mathbf{A}$ je úloha NP-těžká (důkaz neuvádíme).

Příklad: Celočíselné lineární programování

$$\begin{aligned} \min \quad & \mathbf{c}^T \mathbf{x} \\ \text{za podmínek} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b} \\ & \mathbf{x} \in \{0, 1\}^n \end{aligned}$$

Podmínky $x_i \in \{0, 1\}$ lze napsat jako nekonvexní kvadratické rovnosti $x_i(1 - x_i) = 0$.

Konvexní relaxace nekonvexní úlohy

Je-li $X \subseteq Y \subseteq \mathbb{R}^n$, pak

$$\underbrace{\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in X \,\}}_{\text{původní úloha}} \quad \geq \quad \underbrace{\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in Y \,\}}_{\text{relaxovaná úloha}}$$

Nahradíme původní (složitou, nekonvexní) množinu X její (jednoduchou, konvexní) nadmnožinou Y.

Viděli jsme dříve pro 0-1 LP, kde

$$X = \{\, \mathbf{x} \in \{0,1\}^n \mid \mathbf{Ax} \geq \mathbf{b} \,\}$$
$$Y = \{\, \mathbf{x} \in [0,1]^n \mid \mathbf{Ax} \geq \mathbf{b} \,\}$$

Užitečnost:

- Poskytne dolní mez na hodnotu globálního minima nerelaxované úlohy.
- Projekce opt. řešení relaxované úlohy na množinu X je aproximací opt. řešení původní úlohy.

Co s nekonvexními úlohami?

Obecně jsou dva přístupy:

- **Globální optimalizace:** nalezení globálního optima, obecně (worst-case) v exponenciálním čase
  - hierarchie stále těsnějších konvexních relaxací (Sherali-Adams hierarchy pro 0-1 LP)
  - metoda sečných nadrovin (cutting-plane method)
  - metoda větví a mezí (branch-and-bound)

- Nalezení přibližného optima (tzv. sub-optima) v krátkém čase.
  Heuristiky (každá použitelná jen na určitý typ úloh):
  - **Lokální optimalizace:** najde lokální (v nějakém smyslu) optimum
    (gradientní sestup, optimalizace po blocích souřadnic, majorize-minimize methods, ...)
  - konv. relaxace + projekce na původní přípustnou množinu
  - přírodou inspirované heuristiky (evoluční/genetické algoritmy, simulované žíhání, kolonie mravenců, ...)
  - continuation/homotopy methods
  - stochastické metody