Definice 1. Mějme $m, n \in \mathbb{N}$. Řekneme, že zobrazení $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ je *afinní*, pokud platí

$$\boldsymbol{f}(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) = \alpha \boldsymbol{f}(\boldsymbol{x}) + (1-\alpha)\boldsymbol{f}(\boldsymbol{y}) \quad \text{pro všechna } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, \alpha \in \mathbb{R}.$$

Pokuste se sami dokázat, že tato definice se dá ekvivalentně formulovat tak, že existuje *lineární* zobrazení $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$ takové, že je $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{0}) + \boldsymbol{g}(\boldsymbol{x})$ pro každé $\boldsymbol{x} \in \mathbb{R}^n$. (A to je totéž, jako říci, že existuje $\mathbb{A} \in M(m \times n)$ a $\boldsymbol{b} \in \mathbb{R}^m$ takový, že $\boldsymbol{f}(\boldsymbol{x}) = \mathbb{A}\boldsymbol{x} + \boldsymbol{b}$, viz věta !!!. Z toho je vidět, že tato definice souhlasí s definicí pro případ $m = n = 1$ v kapitole !!!.)

Dále si rozmyslete, že pro afinní funkci $f : \mathbb{R}^n \to \mathbb{R}$ a libovolná $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ platí

$$f(\boldsymbol{x} + \boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})\boldsymbol{y}. \tag{1}$$

Budiž $m, n \in \mathbb{N}$, $f : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \dots, m$. V kapitole !!! jsme se zabývali hledáním extrémů funkce f na množině $\{\boldsymbol{x} \in \mathbb{R}^n : g_1(\boldsymbol{x}) = 0, \dots, g_m(\boldsymbol{x}) = 0\}$, přičemž platilo $m < n$. V mnoha úlohách jsou ale vazby zadané pomocí rovností příliš omezující (např. není nutné vyžadovat, abychom plně využili všechny zdroje, pokud by to vedlo k záporné produktivitě). V této kapitole se tedy budeme zabývat následující úlohou:

$$\text{Najděte maximum funkce } f \text{ na množině } M = \{\boldsymbol{x} \in \mathbb{R}^n : g_j(\boldsymbol{x}) \geq 0 \text{ pro } j = 1, \dots, m\}. \tag{2}$$

(Nyní nebudeme vyžadovat, aby platilo $m < n$.) Pokud jsou funkce f a všechny g_j afinní, říká se takovýmto úlohám tradičně lineární programování; pokud jsou obecné, pak se těmto úlohám říká nelineární programování. Některé úlohy tohoto typu jsme již řešili (viz příklad !!!) rozdělením množiny M na její vnitřek a hranici. Pokud jsou ovšem vazby složitější funkce a je jich mnoho, může se stát, že touto metodou dostaneme prakticky obtížně řešitelnou soustavu rovností a nerovností. Nyní si tedy ukážeme, jak řešit úlohu (2) „najednou“.

Nejprve zavedeme některé pomocné pojmy a značení.

Lagrangeovou funkcí[1] úlohy (2) nazveme funkci $L : \mathbb{R}^{n+m} \to \mathbb{R}$ definovanou předpisem

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{g}(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{\lambda} \in \mathbb{R}^m,$$

kde

$$\boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} g_1(\boldsymbol{x}) \\ \vdots \\ g_m(\boldsymbol{x}) \end{pmatrix}, \quad \boldsymbol{x} \in \mathbb{R}^n.$$

Je-li $\boldsymbol{x} \in \mathbb{R}^n$, zápisem $\boldsymbol{x} \geq \boldsymbol{0}$ budeme rozumět, že $x_i \geq 0$ pro všechna $1 \leq i \leq n$. Analogicky zavedeme $\boldsymbol{x} \leq \boldsymbol{0}$. Množinu M tedy můžeme zapsat jako $M = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0}\}$.

Následující věta nám za jistých předpokladů na funkce g_j dává nutnou podmínku, kterou musí splňovat řešení úlohy (2).

Věta 2 (Harold W. Kuhn, Albert W. Tucker)**.** *Nechť $m, n \in \mathbb{N}$ a $f, g_1, \dots, g_m \in \mathcal{C}^1(\mathbb{R}^n)$. Dále nechť je splněna alespoň jedna z následujících podmínek:*

$$g_1, \dots, g_m \textit{ jsou afinní}, \tag{3}$$

$$g_1, \dots, g_m \textit{ jsou konkávní a existuje } \widehat{\boldsymbol{x}} \in \mathbb{R}^n \textit{ takové, že } g_j(\widehat{\boldsymbol{x}}) > 0 \textit{ pro všechna } 1 \leq j \leq m. \tag{4}$$

Potom ke každému řešení $\boldsymbol{x}^0$ úlohy (2) *existuje $\boldsymbol{\lambda}^0 \in \mathbb{R}^m$ takové, že jsou splněny následující (tzv. Kuhnovy-Tuckerovy) podmínky:*

$$\boldsymbol{g}(\boldsymbol{x}^0) \geq \boldsymbol{0}, \boldsymbol{\lambda}^0 \geq \boldsymbol{0}, (\boldsymbol{\lambda}^0)^T \boldsymbol{g}(\boldsymbol{x}^0) = 0, \tag{5}$$

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^0, \boldsymbol{\lambda}^0) = \left[\frac{\partial L}{\partial x_1}(\boldsymbol{x}^0, \boldsymbol{\lambda}^0), \dots, \frac{\partial L}{\partial x_n}(\boldsymbol{x}^0, \boldsymbol{\lambda}^0) \right] = \boldsymbol{0}. \tag{6}$$

Poznámka 3. Podmínce (4) se říká *Slaterova podmínka*. Zaručuje, že množina M z úlohy (2), na níž hledáme maximum funkce f, je konvexní a má neprázdný vnitřek.

Podmínkám (3) a (4) se říká *podmínky regularity*. Existují i obecnější podmínky regularity, než jsou ty uvedené ve větě 2, například následující podmínka, která připomíná podmínku z věty o Lagrangeových multiplikátorech:

Nechť $\boldsymbol{x}^0$ je řešením úlohy (2). Označme $J = \{j \in \{1, \dots, m\} : g_j(\boldsymbol{x}^0) = 0\}$, tj. množinu indexů těch vazeb, které jsou tzv. „aktivní“ v $\boldsymbol{x}^0$. Potom vektory $\nabla g_j(\boldsymbol{x}^0)$, $j \in J$, musí být lineárně nezávislé.

Pokud není splněna žádná z podmínek regularity, může se stát, že existuje řešení $\boldsymbol{x}^0$ úlohy (2) takové, že neexistuje $\boldsymbol{\lambda}^0 \in \mathbb{R}^m$ splňující Kuhnovy-Tuckerovy podmínky (5), (6). (Rozmyslete si, že tomu tak je v případě, kdy $n = 2, m = 1$, $f(x, y) = -(x-1)^2 - y^2$, $g_1(x, y) = -x^2$, $\boldsymbol{x}^0 = [0, 0]$.)

Důkaz věty 2 provedeme pomocí Farkasova lemmatu, se kterým se lze setkat například v lineárním programování. (Tady možná ještě nějaké kecy!!!)

Lemma 4 (Gyula Farkas, 1902)**.** *Nechť $\mathbb{A} \in M(m \times n)$ a $\boldsymbol{c} \in \mathbb{R}^n$. Soustava rovnic a nerovnic $\boldsymbol{y}^T \mathbb{A} + \boldsymbol{c}^T = \boldsymbol{0}$, $\boldsymbol{y} \geq \boldsymbol{0}$ pro neznámou $\boldsymbol{y} \in \mathbb{R}^m$ má řešení právě když soustava nerovnic $\mathbb{A}\boldsymbol{x} \geq \boldsymbol{0}$, $\boldsymbol{c}^T \boldsymbol{x} > 0$ pro neznámou $\boldsymbol{x} \in \mathbb{R}^n$ nemá žádné řešení.*

[1] Tato funkce pro nás není zcela nový objekt. Pomocí ní lze zformulovat i Lagrangeovu větu o multiplikátorech, viz !!!

Důkaz věty 2. Nechť $\boldsymbol{x}^0$ je řešením úlohy (2). Bez újmy na obecnosti můžeme předpokládat, že $g_j(\boldsymbol{x}^0) = 0$ pro $1 \le j \le k$ a $g_j(\boldsymbol{x}^0) > 0$ pro $k < j \le m$.

Pokud $k = 0$, pak (ze spojitosti funkcí $g_1, \dots, g_m$) je $\boldsymbol{x}^0$ vnitřním bodem množiny M, tedy musí být i stacionárním bodem funkce f a snadno ověříme, že volbou $\boldsymbol{\lambda}^0 = \boldsymbol{0}$ jsou obě podmínky (5) a (6) splněny.

Nechť tedy $k > 0$. Označme $\mathbb{A} = \left(\nabla \boldsymbol{g}_j(\boldsymbol{x}^0)\right)_{1 \le j \le k}$, tedy je $\mathbb{A} \in M(k \times n)$. Důkaz nejprve provedeme za obecnější (i když málo názorné) podmínky regularity:

$$\nabla f(\boldsymbol{x}^0)\boldsymbol{h} \le \boldsymbol{0} \text{ pro všechna } \boldsymbol{h} \in \mathbb{R}^n \text{ taková, že } \mathbb{A}\boldsymbol{h} \ge \boldsymbol{0}. \tag{7}$$

Tato podmínka ovšem znamená, že označíme-li $\boldsymbol{c} = (\nabla f(\boldsymbol{x}^0))^T$, pak soustava nerovnic $\mathbb{A}\boldsymbol{h} \ge \boldsymbol{0}$, $\boldsymbol{c}^T\boldsymbol{h} > 0$ nemá žádné řešení. Podle lemmatu 4 tedy existuje $\boldsymbol{y} \in \mathbb{R}^k$, $\boldsymbol{y} \ge \boldsymbol{0}$ takové, že $\nabla f(\boldsymbol{x}^0) + \boldsymbol{y}^T\mathbb{A} = \boldsymbol{0}$. Nyní si stačí uvědomit, že položíme-li $\lambda_j^0 = y_j$ pro $1 \le j \le k$ a $\lambda_j^0 = 0$ pro $k < j \le m$, je tím splněna podmínka (6). První dvě nerovnosti v podmínce (5) jsou zaručeny tím, že $\boldsymbol{x}^0 \in M$ a volbou $\boldsymbol{\lambda}^0$. Poslední nerovnost potom platí, protože pro $1 \le j \le k$ je $g_j(\boldsymbol{x}^0) = 0$, zatímco pro $k < j \le m$ je $\lambda_j^0 = 0$.

Nyní ukážeme, že podmínka (3), případně (4), implikuje podmínku (7).

Nechť tedy platí (3), tj. $g_1, \dots, g_m$ jsou afinní. Vezměme libovolný vektor $\boldsymbol{h} \in \mathbb{R}^n$ takový, že $\mathbb{A}\boldsymbol{h} \ge \boldsymbol{0}$. Podle (1) máme $g_j(\boldsymbol{x}^0 + t\boldsymbol{h}) = g_j(\boldsymbol{x}^0) + t\nabla g_j(\boldsymbol{x}^0)\boldsymbol{h} \ge t\nabla g_j(\boldsymbol{x}^0)\boldsymbol{h} \ge 0$ pro $t \ge 0$ a $1 \le j \le k$. Ze spojitosti funkcí g_j plyne existence $t_0 > 0$ takového, že $g_j(\boldsymbol{x}^0 + t\boldsymbol{h}) > 0$ pro $t \in (0, t_0)$ a $k < j \le m$. (Pro tato j je totiž $g_j(\boldsymbol{x}^0) > 0$.) Dohromady nám to dává, že $\boldsymbol{x}^0 + t\boldsymbol{h} \in M$ pro všechna $t \in (0, t_0)$. Funkce f nabývá v bodě $\boldsymbol{x}^0$ na M maxima, platí tedy $f(\boldsymbol{x}^0 + t\boldsymbol{h}) \le f(\boldsymbol{x}^0)$ pro $t \in (0, t_0)$, z čehož dostáváme $\nabla f(\boldsymbol{x}^0)\boldsymbol{h} = \lim_{t \to 0+} \frac{f(\boldsymbol{x}^0 + t\boldsymbol{h}) - f(\boldsymbol{x}^0)}{t} \le 0$, tedy platí (7).

Nyní nechť platí (4). Chceme ukázat, že pak platí i (7). Udělejme to sporem. Ať tedy neplatí (7), tj. existuje $\boldsymbol{h} \in \mathbb{R}^n$ takový, že $\mathbb{A}\boldsymbol{h} \ge \boldsymbol{0}$ a $\nabla f(\boldsymbol{x}^0)\boldsymbol{h} > 0$. Bohužel nemůžeme postupovat stejně jako v afinním případě, neboť zde se může stát, že vektory $\boldsymbol{x}^0 + t\boldsymbol{h}$ leží mimo množinu M pro všechna $t \in \mathbb{R}$, $t \ne 0$. Ale vše ještě není ztraceno. Zkoumejme konvexní kombinace vektorů $\boldsymbol{x}^0 + \boldsymbol{h}$ a $\widehat{\boldsymbol{x}}$, neboli vektory $\boldsymbol{x}^\alpha = \alpha\widehat{\boldsymbol{x}} + (1 - \alpha)(\boldsymbol{x}^0 + \boldsymbol{h})$, kde $\alpha \in (0, 1)$. Máme

$$\nabla f(\boldsymbol{x}^0)(\boldsymbol{x}^\alpha - \boldsymbol{x}^0) = \nabla f(\boldsymbol{x}^0)(\alpha\widehat{\boldsymbol{x}} + (1 - \alpha)(\boldsymbol{x}^0 + \boldsymbol{h}) - \boldsymbol{x}^0) = \alpha\nabla f(\boldsymbol{x}^0)(\widehat{\boldsymbol{x}} - \boldsymbol{x}^0) + (1 - \alpha)\nabla f(\boldsymbol{x}^0)\boldsymbol{h}.$$

Protože $\nabla f(\boldsymbol{x}^0)\boldsymbol{h} > 0$, existuje $\alpha_0 \in (0, 1)$ takové, že $\nabla f(\boldsymbol{x}^0)(\boldsymbol{x}^{\alpha_0} - \boldsymbol{x}^0) > 0$. Definujeme-li funkci $\varphi\colon \mathbb{R} \to \mathbb{R}$ předpisem $\varphi(t) = f(\boldsymbol{x}^0 + t(\boldsymbol{x}^{\alpha_0} - \boldsymbol{x}^0))$ (tj. funkce φ je řezem funkce f bodem $\boldsymbol{x}^0$ ve směru $\boldsymbol{x}^{\alpha_0} - \boldsymbol{x}^0$), platí podle věty o derivaci složené funkce

$$\varphi'(0) = \nabla f(\boldsymbol{x}^0)(\boldsymbol{x}^{\alpha_0} - \boldsymbol{x}^0) > 0. \tag{8}$$

Podobně dostaneme pro $1 \le j \le k$, že

$$\nabla g_j(\boldsymbol{x}^0)(\boldsymbol{x}^{\alpha_0} - \boldsymbol{x}^0) = \alpha_0\nabla g_j(\boldsymbol{x}^0)(\widehat{\boldsymbol{x}} - \boldsymbol{x}^0) + (1 - \alpha_0)\nabla g_j(\boldsymbol{x}^0)\boldsymbol{h}.$$

Protože funkce g_j jsou konkávní a $g_j(\boldsymbol{x}^0) = 0$ pro $1 \le j \le k$, platí (věta 5.42!!!) $\nabla g_j(\boldsymbol{x}^0)(\widehat{\boldsymbol{x}} - \boldsymbol{x}^0) \ge g_j(\widehat{\boldsymbol{x}}) - g_j(\boldsymbol{x}^0) = g_j(\widehat{\boldsymbol{x}}) > 0$. (Poslední nerovnost plyne z předpokladu (4).) Pro $1 \le j \le k$ je ale také $\nabla g_j(\boldsymbol{x}^0)\boldsymbol{h} \ge 0$ (viz (7)), takže dohromady dostáváme $\nabla g_j(\boldsymbol{x}^0)(\boldsymbol{x}^{\alpha_0} - \boldsymbol{x}^0) > 0$. Definujeme-li pro $1 \le j \le k$ funkce $\psi_j : \mathbb{R} \to \mathbb{R}$ předpisem $\psi_j(t) = g_j(\boldsymbol{x}^0 + t(\boldsymbol{x}^{\alpha_0} - \boldsymbol{x}^0))$, je opět podle věty o derivaci složené funkce $\psi_j'(0) = \nabla g_j(\boldsymbol{x}^0)(\boldsymbol{x}^{\alpha_0} - \boldsymbol{x}^0) > 0$. To ovšem podle definice derivace znamená, že existuje $t_1 > 0$ takové, že $\psi_j(t) > \psi_j(0)$ pro všechna $t \in (0, t_1)$, $1 \le j \le k$. Dále, ze spojitosti funkcí g_j, $k < j \le m$, plyne existence $t_2 \in (0, t_1)$ takového, že $g_j(\boldsymbol{x}^0 + t(\boldsymbol{x}^{\alpha_0} - \boldsymbol{x}^0)) > 0$ pro $t \in (0, t_2)$, $k < j \le m$. (Uvědomme si, že pro $k < j \le m$ je $g_j(\boldsymbol{x}^0) > 0$.) A konečně, opět z definice derivace a nerovnosti (8), existuje $t_0 \in (0, t_2)$ pro které je $\varphi(t_0) > \varphi(0)$. Dohromady tedy dostáváme, že v bodě $\boldsymbol{x}^1 = \boldsymbol{x}^0 + t_0(\boldsymbol{x}^{\alpha_0} - \boldsymbol{x}^0)$ je $g_j(\boldsymbol{x}^1) > 0$ pro všechna $1 \le j \le m$, tedy $\boldsymbol{x}^1 \in M$, a dále $f(\boldsymbol{x}^1) > f(\boldsymbol{x}^0)$, což je spor s faktem, že f nabývá maxima na M v bodě $\boldsymbol{x}^0$.

□

Nyní si ukážeme, že za předpokladů konkávnosti jsou Kuhnovy-Tuckerovy podmínky podmínkami postačujícími pro extrém.

Věta 5. *Nechť $m, n \in \mathbb{N}$ a $f, g_1, \dots, g_m \in \mathcal{C}^1(\mathbb{R}^n)$. Dále nechť $f, g_1, \dots, g_m$ jsou konkávní na $\mathbb{R}^n$. Jestliže vektory $\boldsymbol{x}^0 \in \mathbb{R}^n$ a $\boldsymbol{\lambda}^0 \in \mathbb{R}^m$ splňují Kuhnovy-Tuckerovy podmínky* (5) *a* (6)*, pak $\boldsymbol{x}^0$ je řešením úlohy* (2).

Důkaz. Z podmínky (5) dostáváme $\boldsymbol{x}^0 \in M$. Z předpokladu konkávnosti snadno plyne, že funkce $\boldsymbol{x} \mapsto L(\boldsymbol{x}, \boldsymbol{\lambda}^0)$ je konkávní na $\mathbb{R}^n$.[2] Protože podle (6) je $\boldsymbol{x}^0$ jejím stacionárním bodem, je bodem maxima této funkce na $\mathbb{R}^n$ (viz !!!). Jinými slovy, pro každé $\boldsymbol{x} \in \mathbb{R}^n$ platí

$$f(\boldsymbol{x}) + (\boldsymbol{\lambda}^0)^T\boldsymbol{g}(\boldsymbol{x}) \le f(\boldsymbol{x}^0) + (\boldsymbol{\lambda}^0)^T\boldsymbol{g}(\boldsymbol{x}^0) = f(\boldsymbol{x}^0).$$

(Poslední rovnost plyne z podmínky (5).) Je-li $\boldsymbol{x} \in M$, je ovšem $(\boldsymbol{\lambda}^0)^T\boldsymbol{g}(\boldsymbol{x}) \ge 0$, a proto $f(\boldsymbol{x}) \le f(\boldsymbol{x}^0)$. Tím jsme dokázali, že bod $\boldsymbol{x}^0$ je bodem maxima funkce f na M.

□

(dokonce to je nutná podmínka lokálního maxima)

Možná by bylo lepší nahradit ty funkce φ,ψ_j pojmem derivace ve směru? (Beztak se musí nějak na konci toho afinního případu ještě doodůvodnit to, že ta limita se rovná tomu gradientu.)

obrázek pro ty afinní??

[2] To proto, že nezáporný násobek konkávní funkce je konkávní a součet konkávních funkcí je konkávní funkce. Rozmyslete si to podrobněji.