Optimalizační metody

Jiří Sgall

4. června 2016

1 Optimalizace

Vznik ve 40. letech, motivace válečná logistika, později teorie her (von Neumann), ekonomie.

Kombinatorická optimalizace

- nejkratší cesty v grafu
- minimální kostry
- párování v grafu
- toky v síti
- (problém obchodního cestujícího)
- (vrcholové pokrytí)

Lineární (matematické) programování

- studium soustav nerovnic a jejich řešení
- optimalizace řešení za podmínek určených soustavou nerovnic

Náplň přednášky

- teorie lineárního programování (dále LP), mnohostěny
- vztah LP a kombinatorické optimalizace
- simplexová metoda, další algoritmy
- párování v grafech
- celočíselné programování
- použití LP pro těžké úlohy
- matroidy

Soustava nerovnic optimalizujeme řešení rovnice omezené soustavou nerovnic

- jedno řešení
- víc řešení
- žádné řešení - žádné přípustné řešení
- žádné řešení - účelová fce neomezená

- rovnice/nerovnice
- proměnné nezáporné/obecné

Převody

rovnice → nerovnice
- substituce (opatrně)
- $A\boldsymbol{x} = \boldsymbol{b} \to A\boldsymbol{x} \leq \boldsymbol{b}$; $A\boldsymbol{x} \geq \boldsymbol{b}$

nerovnice → rovnice $A\boldsymbol{x} \leq \boldsymbol{b} \to A\boldsymbol{x} + z = \boldsymbol{b}$; $z \geq 0$

obecné proměnné → nezáporné proměnné $\boldsymbol{x} \to \boldsymbol{x}^+ - \boldsymbol{x}^-$; $\boldsymbol{x}^+, \boldsymbol{x}^- \geq 0$

2 Úloha lineárního programování

Definice 2.1. **Úloha LP (ve standardním tvaru)** je optimalizační úloha s cílem maximalizovat $\boldsymbol{c}^T\boldsymbol{x}$ za podmínek $A\boldsymbol{x} \leq \boldsymbol{b}$ (porovnání po složkách) přes $\boldsymbol{x} \in \mathbb{R}^n$ pro $A \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{c} \in \mathbb{R}^n$.

Značení:

n počet proměnných

m počet podmínek

Terminologie:

účelová funkce $\boldsymbol{c}^T\boldsymbol{x}$

přípustné řešení $\boldsymbol{x}$ splňující $A\boldsymbol{x} \leq \boldsymbol{b}$

optimum přípustné $\boldsymbol{x}$ s maximální hodnotou $\boldsymbol{c}^T\boldsymbol{x}$

nepřípustná úloha LP neexistuje přípustné řešení

neomezená úloha LP existuje přípustné řešení s libovolně velkou hodnotou $\boldsymbol{c}^T\boldsymbol{x}$

Definice 2.2. **Úloha LP v rovnicovém tvaru** je optimalizační úloha maximalizovat $\boldsymbol{c}^T\boldsymbol{x}$ za podmínek $A\boldsymbol{x} = \boldsymbol{b}$ přes $\boldsymbol{x} \in [0, \infty)^n$ (ve světě LP můžeme psát $\boldsymbol{x} \geq 0$).

Vztah mezi lineární algebrou a LP

lin. algebra	lin. programování
soustavy rovnic	soustavy nerovnic
nad obecným tělesem	nad $\mathbb{R}$
řešení: afinní prostory	řešení: mnohostěny
Gaussova eliminace	simplexová metoda
hodnost, dimenze, ...	dualita

Dopravní problém Máme nějaké zdroje a nějaké cíle. Z každého zdroje se dá dojet do každého cíle, přepravit jednotku zboží mezi zdrojem i a cílem j stojí $c_{i,j}$, zdroje mají kapacitu (a_i), cíle požadavky (b_i).

$$\begin{array}{rl} \text{minimalizuj} & \sum_{i,j} c_{ij}^T x_{ij} \\ \text{pro} & \boldsymbol{x} \geq 0 \\ \text{za podmínek} & \sum_j x_{ij} \leq a_i \qquad \forall i \\ & \sum_i x_{ij} \geq b_j \qquad \forall j \end{array}$$

případně $\boldsymbol{x}_{i,j} \in \mathbb{N}$ (celočíselné programování).

Úloha celočíselného programování (IP - integer programming) je optimalizační úloha s cílem maximalizovat $\boldsymbol{c}^T\boldsymbol{x}$ za podmínek $A\boldsymbol{x} \leq \boldsymbol{b}$, $\boldsymbol{x} \in \mathbb{Z}^n$.

Smíšené celočíselné programování (MIP - mixed IP) některé proměnné $\in \mathbb{Z}$, některé $\in \mathbb{R}$

Binární celočíselné programování $\boldsymbol{x} \in \{0,1\}^n$

Hledání nejkratší cesty je dán orientovaný graf s cenami hran c_e, $e \in G$; s, t vrcholy, hledáme nejkratší cestu z s do t.

$$\begin{array}{rlll} \text{minimalizuj} & \sum c_e x_e & & \\ \text{pro} & x_e \in \{0,1\} & & \\ \text{za podmínek} & \sum_u x_{uv} - \sum_u x_{vu} & = -1 & \text{pro } v = s \\ & & = +1 & \text{pro } v = t \\ & & = 0 & \text{jinak } (\forall v \in V - \{s,t\}) \end{array}$$

Bellman-Ford y_v délka nejkr. cesty $s \to v$

$$\begin{array}{rll} \text{maximalizuj} & y_t - y_s & \\ \text{za podmínek} & y_v - y_u \leq c_{uv} & \text{pro všechny hrany } uv \\ & y_s = 0 & \end{array}$$

Fordův algoritmus nastaví vzdálenosti na nekonečna a pak opravuje hodnoty, pokud jsou porušené nerovnosti.
Belmannův dělá totéž, v pořadí které zaručí polynomiální časovou složitost.

Důkaz: (1) přípustné řešení $\rightarrow y_t \leq$ délka nejkratší cesty (2) existuje přípustné řešení $y_t =$ délka nejkratší cesty viz důkaz Bellman-Forda

Dualita s programem výše.

3 Příklady

3.1 Vrcholové pokrytí

Je dán neorientovaný graf $G = (V, E)$, vrcholové pokrytí je podmnožina vrcholů $U \subseteq V$ taková, že každá hrana z E je incidentní aspoň jednomu vrcholu z U. Zavedeme proměnné pro vrcholy x_v, jejich hodnota je buď 1, pokud je vrchol v ve vrcholovém pokrytí, nebo 0, když ne. Celočíselný program:

$$\begin{aligned} \text{minimalizuj} \quad & \sum_{v \in V} x_v \\ \text{pro} \quad & \boldsymbol{x} \in \{0,1\}^n \\ \text{za podmínek} \quad & x_u + x_v \geq 1 \quad \forall \{u,v\} \in E \end{aligned}$$

Podmínky $x_u + x_v \geq 1$ nám zaručí, že každá hrana bude pokrytá vrcholem a výběr proměnných z $\{0,1\}$ zase celočíselnost. **LP relaxace** této úlohy:

$$\begin{aligned} \text{minimalizuj} \quad & \sum_{v \in V} x_v \\ \text{pro} \quad & 0 \leq x_v \leq 1 \\ \text{za podmínek} \quad & x_u + x_v \geq 1 \quad (\forall u,v)\{u,v\} \in E \end{aligned}$$

Problém je, že optimální řešení této LP relaxace nemusí (a pravděpodobně také nebude) optimálním řešením původní úlohy celočíselného programování. Navíc proměnné budou reálná čísla mezi 0 a 1.

3.2 Nezávislá množina

Je dán neorientovaný graf $G = (V, E)$, nezávislá množina je podmnožina vrcholů $U \subseteq V$ taková, že žádné dva vrcholy z U nejsou spojené hranou. V celočíselném programu opět použijeme proměnné pro vrcholy. x_v je rovno 1, pokud v patří do nezávislé množiny, 0 jinak.

$$\begin{aligned} \text{maximalizuj} \quad & \sum_{v \in V} \boldsymbol{x}_v \\ \text{pro} \quad & \boldsymbol{x} \in \{0,1\}^n \\ \text{za podmínek} \quad & x_u + x_v \leq 1 \quad (\forall u,v)\{u,v\} \in E \end{aligned}$$

Podmínky $x_u + x_v \leq 1$ zaručují, že do nezávislé množiny se může dostat nejvýše jeden z vrcholů incidentních každé hraně. LP relaxace by se provedla jako u vrcholového pokrytí.

Podobně můžeme modelovat i další grafové problémy, lineární programování je vhodné zj. pro ty, kdy jsou omezující podmínky lokální (tedy ne např. TSP, kde je problém zajistit souvislost).

3.3 Minimální perfektní párování v bipartitním grafu

Na vstupu máme bipartitní graf $G = (U, V, E)$ $|U| = |V| = n$, $E \subseteq U \times V$, G má perfektní párování, a váhovou funkci $w : E \to \mathbb{R}_0^+$. A hledáme perfektní párovaní $M \subseteq E$ takové, že $\sum_{e \in M} w(e)$ je minimální.

V celočíselném programu tentokrát zvolíme proměnné pro hrany:

$$\begin{aligned} \text{minimalizuj} \quad & \sum_{e \in E} w(e) x_e \\ \text{pro} \quad & \boldsymbol{x} \in \{0,1\}^n \\ \text{za podmínek} \quad & \sum_{v \in V} x_{uv} = 1 \quad \forall u \in U \\ & \sum_{u \in U} x_{uv} = 1 \quad \forall v \in V \end{aligned}$$

Podmínky zajistí, že do každého vrcholu v U i V vchází právě jedna hrana tedy, že se jedná o perfektní párování. Přímo dostaneme i LP relaxaci této úlohy:

$$\begin{aligned} \text{minimalizuj} \quad & \sum_{e \in E} w(e) x_e \\ \text{pro} \quad & \boldsymbol{x} \geq 0 \\ \text{za podmínek} \quad & \sum_{v \in V} x_{uv} = 1 \quad \forall u \in U \\ & \sum_{u \in U} x_{uv} = 1 \quad \forall v \in V \end{aligned}$$

Tady už nám podmínky nezajistí perfektní párování. Může se stát, že např. dostaneme vrchol stupně 2 takový, že proměnné pro hrany jemu incidentní budou mít obě hodnotu 0.5 nebo dokonce 0.9 a 0.1. Naštěstí pro bipartitní grafy platí následující věta.

Věta 3.1. *Mějme vážený bipartitní graf $G = (V, E)$ a celočíselný program odpovídající hledání minimálního perfektního párování v G. Pak existuje optimální řešení $\boldsymbol{x}$ LP relaxace, které zároveň splňuje $\boldsymbol{x} \in \{0,1\}^m$.*

Důkaz. Každé perfektní párování odpovídá přípustnému řešení celočíselného lineárního programu. Pokud lineární program nemá přípustné řešení, neexistuje perfektní párování.

Pokud je úloha LP relaxace přípustná, pak optimální řešení existuje, protože přípustná řešení tvoří kompaktní množinu (je omezená, protože $0 \leq \boldsymbol{x} \leq 1$). Vezmeme tedy $\boldsymbol{x}$ optimální řešení LP relaxace s minimálním počtem neceločíselných proměnných. Naším cílem je dokázat, že $\boldsymbol{x}$ je celočíselné.

Pokud ne, uvažme neprázdnou podmnožinu hran $E' = \{e \in E \mid x_e \in (0,1)\}$, tj. odpovídající všem proměnným, které nejsou celočíselné. Tvrdíme, že graf G' indukovaný hranami z E' neobsahuje vrcholy stupně 1: Kdyby totiž takový vrchol měl, nesplňoval by podmínku LP, která říká, že součet proměnných incidentních hran je 1.

Dále tvrdíme, že v G' existuje cyklus $C \subseteq E'$, což plyne z předchozího. Protože je G bipartitní, tak $|C| = 2k$ (C je sudé délky). Hrany v cyklu očíslujeme $C = e_1, e_2, \ldots, e_{2k}$.

Uvažme jiné řešení $\boldsymbol{y}$ definované pro nějaké malé reálné ε takto:

$$y_e = \begin{cases} x_e + \varepsilon & e_1, e_3, \ldots, e_{2k-1} \\ x_e - \varepsilon & e_2, e_4, \ldots, e_{2k} \\ x_e & \text{jinak} \end{cases}$$

Protože u proměnných odpovídajícím hranám incidentním s daným vrcholem jednou ε přičteme a jednou odečteme, tak suma zůstane stejná a podmínky LP jsou splněny. Protože x_e pro hrany z C jsou z ostrého intervalu $(0, 1)$, tak pro dostatečně malé ε bude platit i podmínka $0 \leq x_e \leq 1$. Pro dostatečně malé kladné i záporné ε je tedy $\boldsymbol{y}$ přípustné řešení.

Uvažme nyní, jak se změní účelová funkce.

$$\sum_{e \in E} w(e) y_e = \sum_{e \in E} w(e) x_e + \varepsilon \left(\sum_{\text{sudé } e \in C} w(e) x_e - \sum_{\text{liché } e \in C} w(e) x_e \right)$$

Protože $\boldsymbol{x}$ je optimální, musí být výraz ve svorce 0, jinak bychom mohli vhodnou volbou malého kladného nebo záporného ε zlepšit optimum.

Pro maximální ε takové, že $\boldsymbol{y}$ je přípustné řešení, platí pro některou hranu $e \in \{e_1, \ldots, e_{2k}\}$, že $x_e \in \{0, 1\}$, tedy $\boldsymbol{x}$ má méně neceločíselných složek. To je spor s výběrem $\boldsymbol{x}$ jako přípustného řešení s maximálním počtem celočíselných složek.

Dokázali jsme tedy, že optimální řešení $\boldsymbol{x}$ je celočíselné. $\square$

4 Opakování lineární algebry a konvexita

4.1 Afinní prostory

Definice 4.1. L je **afinní prostor** [angl. *affine space*], pokud $L = \boldsymbol{x} + V$ pro nějaké $\boldsymbol{x} \in \mathbb{R}^n$ a V vektorový podprostor $\mathbb{R}^n$.

Definice 4.2. **Afinní obal** [angl. *affine hull*] množiny $X \subseteq \mathbb{R}^n$ je průnik všech afinních prostorů L takových, že $X \subseteq L$.

Definice 4.3. **Afinní kombinace** [angl. *affine combination*] bodů z X je libovolný bod daný výrazem (nebo podle kontextu přímo výraz) $\alpha_0 \boldsymbol{a}_0 + \ldots + \alpha_k \boldsymbol{a}_k$, pro nějaké $k \in \mathbb{N}$, $\boldsymbol{a}_i \in X$, $\alpha_i \in \mathbb{R}$, $\sum \alpha_i = 1$.

Věta 4.4. *Afinní obal X je roven $\{\alpha_0 \boldsymbol{a}_0 + \ldots + \alpha_k \boldsymbol{a}_k \mid k \in \mathbb{N}, \boldsymbol{a}_i \in X, \alpha_i \in \mathbb{R}, \sum \alpha_i = 1\}$. Afinní obal X je afinní prostor.*

Definice 4.5. Body $\boldsymbol{a}_0, \ldots, \boldsymbol{a}_k$ jsou **afinně nezávislé** [angl. *affine independent*], pokud neexistují netriviální $\alpha_0, \ldots, \alpha_k \in \mathbb{R}$ (tj. existuje i tak, že $\alpha_i \neq 0$) tak, že $\alpha_0 \boldsymbol{a}_0 + \cdots + \alpha_k \boldsymbol{a}_k = 0$ a $\sum \alpha_i = 0$.

Pozorování 4.6. *Body $\boldsymbol{a}_0, \ldots, \boldsymbol{a}_k$ jsou afinně nezávislé právě tehdy, když vektory $\boldsymbol{a}_1 - \boldsymbol{a}_0, \boldsymbol{a}_2 - \boldsymbol{a}_0, \ldots, \boldsymbol{a}_k - \boldsymbol{a}_0$ jsou lineárně nezávislé.*

Definice 4.7. **Dimenze** [angl. *dimension*] množiny $X \subseteq \mathbb{R}^n$, $X \neq \emptyset$, označovaná $\dim(X)$, je dimenze afinního obalu X. Pro $X = \emptyset$ dodefinujeme $\dim(X) = -1$.

Pozorování 4.8. *Dimenze množiny $X \neq \emptyset$ je maximální d takové, že v X existují afinně nezávislé body $\boldsymbol{a}_0, \ldots, \boldsymbol{a}_d$.*

4.2 Konvexní množiny

Konvexní množiny definujeme obdobně jako v planimetrii. Množina X je konvexní, jestliže s každými dvěma body je v X i úsečka mezi nimi. Jiný pohled je, že konvexní množiny jsou množiny uzavřené na konvexní kombinace, tj. na afinní kombinace s nezápornými koeficienty. Obdobně jako pro afinní množiny definujeme konvexní obal a konvexní kombinace, poté pro ně dokážeme vztah obdobný jako pro afinní obal a afinní kombinace.

Definice 4.9. Množina $X \subseteq \mathbb{R}^n$ je **konvexní** [angl. *convex*], pokud pro všechna $\boldsymbol{x}, \boldsymbol{y} \in X$ a $t \in [0,1]$ platí $t\boldsymbol{x} + (1-t)\boldsymbol{y} \in X$.

Definice 4.10. Konvexní obal [angl. *convex hull*] množiny $X \subseteq \mathbb{R}^n$ značíme $\operatorname{conv}(X)$ a definujeme jako průnik všech konvexních množin obsahujících X, tedy

$$\operatorname{conv}(X) = \bigcap \{C \mid C \text{ je konvexní a } X \subseteq C\}.$$

Definice 4.11. Konvexní kombinace [angl. *convex combination*] bodů z X je libovolný bod daný výrazem (nebo podle kontextu přímo výraz) $\alpha_0\boldsymbol{a}_0 + \ldots + \alpha_k\boldsymbol{a}_k$, pro nějaké $k \in \mathbb{N}$, $\boldsymbol{a}_i \in X$, $\alpha_i \geq 0$, $\sum \alpha_i = 1$.

Pozorování 4.12. *Nechť Y je množina konvexních kombinací bodů z X. Pak každá konvexní kombinace bodů z Y je konvexní kombinací bodů z X.*

Důkaz. Do vyjádření bodu z jako konvexní kombinace bodů z Y dosadíme vyjádření bodů $y \in Y$ jako konvexní kombinace bodů z X. Nezápornost i součet koeficientů se zachovají. □

Pozorování 4.13. *Pokud všechny body $\boldsymbol{x} \in X$ jsou v konvexní množině C, pak i libovolná konvexní kombinace bodů z X je v C.*

Důkaz. Opakovaně použijeme definici konvexity. □

Pozorování 4.14. *Pokud všechny body $\boldsymbol{x} \in X$ splňují nerovnici $\boldsymbol{a}^T\boldsymbol{x} \leq b$, pak i libovolná konvexní kombinace bodů z X splňuje tuto nerovnici.*

Pokud všechny body $\boldsymbol{x} \in X$ splňují nerovnici $\boldsymbol{a}^T\boldsymbol{x} < b$, pak i libovolná konvexní kombinace bodů z X splňuje tuto nerovnici.

Pokud všechny body $\boldsymbol{x} \in X$ splňují nerovnici $\boldsymbol{a}^T\boldsymbol{x} \leq b$, a bod $\boldsymbol{y} \in X$ splňuje $\boldsymbol{a}^T\boldsymbol{y} < b$, pak i libovolná konvexní kombinace $\boldsymbol{z}$ bodů z X, ve které je koeficient u $\boldsymbol{y}$ kladný, splňuje $\boldsymbol{a}^T\boldsymbol{z} < b$.

Důkaz. Sečteme nerovnice s dosazenými body z konvexní kombinace, každou vynásobenou příslušným koeficientem z konvexní kombinace. □

Věta 4.15. *Konvexní obal X, $\operatorname{conv}(X)$ je roven množině všech konvexních kombinací $Y = \{\alpha_0\boldsymbol{a}_0 + \ldots + \alpha_k\boldsymbol{a}_k \mid k \in \mathbb{N}, \boldsymbol{a}_i \in X, \alpha_i \geq 0, \sum \alpha_i = 1\}$.*

Důkaz. Každá konvexní množina $C \supseteq X$ obsahuje Y podle pozorování 4.13. Tedy průnik všech takových C obsahuje Y, tedy $Y \subseteq \operatorname{conv}(X)$.

Pro opačnou inkluzi stačí uvědomit si, že množina Y je konvexní. To plyne z pozorování 4.12. Zjevně $X \subseteq Y$, a konvexita Y tedy znamená, že Y se vyskytuje mezi množinami C v definici $\operatorname{conv}(X)$. Pak ale $\operatorname{conv}(X) \subseteq Y$. □

Vztah konvexního obalu k dimenzi je složitější. Zatímco každou afinní množinu dimenze d můžeme definovat jako afinní obal $d+1$ bodů, u konvexního obalu počet bodů omezit nejde, může být i nekonečný, např. u kruhu. Platí ale následující věta, která říká, že každý bod $\boldsymbol{x}$ z $\operatorname{conv}(X)$ pro X dimenze d můžeme vyjádřit jako konvexní kombinaci $d+1$ bodů z X; použité body však mohou být různé pro různá $\boldsymbol{x}$.

Věta 4.16 (Carathéodory). *Nechť $X \subseteq \mathbb{R}^n$, $\dim X = d$. Pak $\operatorname{conv}(X) = \{\alpha_0 \boldsymbol{a}_0 + \ldots + \alpha_d \boldsymbol{a}_d \mid \boldsymbol{a}_i \in X, \alpha_i \geq 0, \sum \alpha_i = 1\}$*

Důkaz. Nechť $\boldsymbol{x} \in \operatorname{conv}(X)$. Zapišme $\boldsymbol{x}$ jako konvexní kombinaci $\boldsymbol{x} = \alpha_0 \boldsymbol{a}_0 + \ldots + \alpha_k \boldsymbol{a}_k$, $\alpha_i \geq 0$, $\sum \alpha_i$ tak, že k je minimální. Ukážeme, že $k \leq d$, čímž bude věta dokázána.

Pro spor předpokládejme, že $k > d$. Pak $\boldsymbol{a}_0, \ldots, \boldsymbol{a}_k$ jsou afinně závislé, tedy existují $\beta_i \geq 0$ tak, že $\beta_0 \boldsymbol{a}_0 + \ldots + \beta_k \boldsymbol{a}_k = \boldsymbol{0}$, $\sum \beta_i = 0$, ale pro nějaké i, $\beta_i \neq 0$. Mezi β_i jsou tedy kladná i záporná čísla. Dále vezmeme γ maximální takové, že $(\forall i)(\gamma\beta_i + \alpha_i \geq 0)$. Takové číslo bude existovat, protože množina čísel γ, ze které vybíráme, je neprázdná (obsahuje $\gamma = 0$), omezená (protože některé β_i je záporné) a uzavřená (daná neostrými nerovnostmi). Uvažme kombinaci

$$\sum_{i=0}^{k} (\gamma\beta_i + \alpha_i)\boldsymbol{a}_i = \boldsymbol{x}$$

Vidíme, že tato kombinace je konvexní: Nezápornost koeficientů platí díky výběru γ a $\sum(\gamma\beta_i + \alpha_i) = 1$ plyne z $\sum \alpha_i = 1$ a $\sum \beta_i = 0$. Dále díky maximalitě γ bude existovat i tak, že $\gamma\beta_i + \alpha_i = 0$, tedy $\boldsymbol{x}$ lze vyjádřit jako konvexní kombinaci méně bodů, což je spor s výběrem konvexní kombinace a k. □

5 Konvexní mnohostěny

Konvexní mnohostěn definujeme jako průnik konečně mnoha poloprostorů, tedy jako množinu přípustných řešení nějakého lineárního programu. Proto jsou pro optimalizaci mnohostěny a jejich studium důležité.

Definice 5.1. **Nadrovina** v $\mathbb{R}^n$ je množina všech $\boldsymbol{x}$ takových, že $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}^T\boldsymbol{x} = b\}$ pro $\boldsymbol{a} \in \mathbb{R}^n, \boldsymbol{a} \neq \boldsymbol{0}, b \in \mathbb{R}$.

Definice 5.2. **Poloprostor** v $\mathbb{R}^n$ je množina všech $\boldsymbol{x}$ takových, že $\{\boldsymbol{x} \in \mathbb{R}^n \mid a^T x \leq b\}$ pro $\boldsymbol{a} \in \mathbb{R}^n, \boldsymbol{a} \neq \boldsymbol{0}, b \in \mathbb{R}$.

Definice 5.3. **Konvexní mnohostěn (polyedr)** [angl. *polyhedron*] v $\mathbb{R}^n$ je průnik konečně mnoha poloprostorů. **Omezený konvexní mnohostěn** se také nazývá **polytop** [angl. *polytope*].

Následující věta ukazuje základní vlastnost konvexních množin. Je dobré si uvědomit důležitost všech předpokladů; pokud bychom nepožadovali uzavřenost nebo omezenost, bylo by možné množiny oddělit jen slabě, tj. neležely by uvnitř daného poloprostoru ale mohly by mít neprázdný průnik s hraniční nadrovinou.

Věta 5.4 (o oddělování). *Nechť $C, D \subseteq \mathbb{R}^n$ jsou neprázdné, uzavřené, konvexní a disjunktní a C je omezená. Pak existuje nadrovina $\{\boldsymbol{x} \mid \boldsymbol{a}^T\boldsymbol{x} = b\}$, která silně odděluje C a D, tj. taková, že $C \subseteq \{\boldsymbol{x} \mid \boldsymbol{a}^T\boldsymbol{x} < b\}$ a $D \subseteq \{\boldsymbol{x} \mid \boldsymbol{a}^T\boldsymbol{x} > b\}$.*

Důkaz. Vezmeme $\boldsymbol{c} \in C$, $\boldsymbol{d} \in D$ tak, že jejich vzdálenost $(||c - d||)$ je minimální. Pokud jsou obě množiny C i D omezené, existence dvojice $\boldsymbol{c}, \boldsymbol{d}$ plyne z kompaktnosti množiny $C \times D$ v $\mathbb{R}^{2n}$ a spojitosti normy. Pokud D není omezená, stačí se omezit na dostatečně velkou kouli, aniž bychom ztratili dvojice blízkých bodů. Přesněji to uděláme tak, že vezmeme nějaké $\boldsymbol{d}' \in D$ a bod $\boldsymbol{c}' \in C$, pro který je vzdálenost $||\boldsymbol{c}' - \boldsymbol{d}'||$ maximální (z kompaktnosti C). Označme $\alpha = ||\boldsymbol{c}' - \boldsymbol{d}'||$ a $\beta = \max_{\boldsymbol{c}'' \in C} ||\boldsymbol{c}'' - \boldsymbol{c}'||$. Uvažujme kouli se středem v $\boldsymbol{c}'$ a s poloměrem $\alpha + \beta$, její průnik s množinou D bude tedy kompaktní množina D'. Pro body $\boldsymbol{c}'' \in C$, $\boldsymbol{d}'' \in D \setminus D'$ platí podle trojúhelníkové nerovnosti

$$||\boldsymbol{c}'' - \boldsymbol{d}''|| \geq ||\boldsymbol{c}' - \boldsymbol{d}''|| - ||\boldsymbol{c}'' - \boldsymbol{c}'|| > (\alpha + \beta) - \beta = \alpha = ||\boldsymbol{c}' - \boldsymbol{d}'||.$$

Omezením na D' tedy nemůžeme přijít o dvojici blízkých bodů. Z kompaktnosti najdeme dvojici bodů $\boldsymbol{c} \in C$, $\boldsymbol{d} \in D'$ s minimální vzdáleností a tato dvojice bodů minimalizuje vzdálenost i přes $\boldsymbol{d} \in D$.

Pro vybrané $\boldsymbol{c} \in C$ a $\boldsymbol{d} \in D$ označme $\boldsymbol{a} = \boldsymbol{d} - \boldsymbol{c}$ a $b = \boldsymbol{a}^T \left(\frac{\boldsymbol{c}+\boldsymbol{d}}{2}\right)$. Pak $\boldsymbol{a}^T\boldsymbol{d} - \boldsymbol{a}^T\boldsymbol{c} = ||\boldsymbol{a}||^2 > 0$ a tedy platí

$$\boldsymbol{a}^T\boldsymbol{c} < b < \boldsymbol{a}^T\boldsymbol{d}.$$

Z toho, že $||\boldsymbol{c} - \boldsymbol{d}||$ je minimální plyne, že pro všechna $\boldsymbol{c}'' \in C$ platí $\boldsymbol{a}^T\boldsymbol{c}'' \leq \boldsymbol{a}^T\boldsymbol{c}$. To nejlépe nahlédneme geometricky. V opačném případě by trojúhelník $\boldsymbol{c}''\boldsymbol{c}\boldsymbol{d}$ měl u $\boldsymbol{c}$ ostrý úhel, a proto $\boldsymbol{x} = \boldsymbol{c} + \varepsilon(\boldsymbol{c}'' - \boldsymbol{c})$ pro dostatečně malé $\varepsilon > 0$ bude blíž k $\boldsymbol{d}$ než $\boldsymbol{c}$. Navíc z konvexity je $\boldsymbol{x} \in C$, dostali jsme tedy spor s minimalitou $||\boldsymbol{c} - \boldsymbol{d}||$.

Obdobně pro všechna $\boldsymbol{d}'' \in D$ platí $\boldsymbol{a}^T\boldsymbol{d}'' \geq \boldsymbol{a}^T\boldsymbol{d}$. Celkově dostáváme

$$\boldsymbol{a}^T\boldsymbol{c}'' \leq \boldsymbol{a}^T\boldsymbol{c} < b < \boldsymbol{a}^T\boldsymbol{d} \leq \boldsymbol{a}^T\boldsymbol{d}''.$$

a tedy nadrovina $\{\boldsymbol{x} \mid \boldsymbol{a}^T\boldsymbol{x} = b\}$ silně odděluje C a D. □

Následující věta říká, že omezené konvexní mnohostěny jsou přesně konvexní obaly konečných množin bodů, dává tedy alternativní definici omezeného konvexního mnohostěnu. Přechod od množiny nerovností k množině bodů a naopak není samozřejmý, také množina bodů může být podstatně větší než množina nerovností a naopak. Ve skutečnosti je tato věta obdobou duality lineárního programování. Tomu taky odpovídá určitá koncepční náročnost důkazu. Jedna implikace je vcelku přímočará indukce podle dimenze mnohostěnu. V druhé implikaci se na nerovnosti omezující konvexní obal množiny bodů budeme dívat jako na body a šikovně použijeme první už dokázanou implikaci.

Věta 5.5 (Minkowski-Weyl)**.** *Nechť $X \subseteq \mathbb{R}^n$, pak X je omezený konvexní mnohostěn, právě když existuje $V \subseteq \mathbb{R}^n$ tak, že V je konečná a $X = \mathrm{conv}(V)$.*

Důkaz. **(1) Je dán mnohostěn X, hledáme množinu bodů V:** Budeme postupovat indukcí podle dimenze $d = \dim(X)$. Pokud $d \leq 0$, je X prázdná nebo jednoprvková, a vezmeme $V = X$.

Nechť $d \geq 1$ a L je afinní obal X, tedy L je afinní prostor dimenze d. Označme nerovnice z popisu X označíme jako $\boldsymbol{a}_i^T\boldsymbol{x} \leq b_i$, $i = 1, \ldots, m$. Dále označme $P_i = \{\boldsymbol{x} \mid \boldsymbol{a}_i^T\boldsymbol{x} \leq b_i\}$ a $R_i = \{\boldsymbol{x} \mid \boldsymbol{a}_i^T\boldsymbol{x} = b_i\}$, tedy poloprostor a hraniční nadrovinu odpovídající podmínce i. Podmínka může celé X obsahovat v hraniční polorovině (pak i $L \subseteq R_i$); pro indukci nás zajímají právě ostatní podmínky. Označme $X_i = X \cap R_i$ a $M = \{i \mid X_i \neq X\}$. Pro $i \in M$ je $\dim(X_i) \leq \dim L \cap R_i \leq d - 1$, protože L neni obsažen v R_i a afinní podprostor $L \cap R_i$

pak musí mít dimenzi menší než L. Pro $i \in M$ tedy najdeme dle indukčního předpokladu konečnou množinu V_i takovou, že $X_i = \operatorname{conv}(V_i)$. Položíme $V = \bigcup_{i \in M} V_i$.

Zbývá dokázat, že $X = \operatorname{conv}(V)$. Vzhledem k tomu, že $V_i \subseteq X_i \subseteq X$, máme $V \subseteq X$ a $\operatorname{conv}(V) \subseteq X$. Nyní pro $\boldsymbol{x} \in X$ dokážeme, že $x \in \operatorname{conv}(V_i)$. Zvolme přímku p tak, že $\boldsymbol{x} \in p$ a $p \subseteq L$ (to lze, protože $\dim(L) \geq 1$). Uvažme průnik $p \cap X$. Ten je dán průnikem se všemi poloprostory P_i, a protože X je omezený, je průnik úsečka s koncovými body $\boldsymbol{y} \in X_i$ a $\boldsymbol{z} \in X_j$ pro nějaké $i, j \in M$. (Přesněji, protože pro $i \notin M$ je p obsažena v R_i, máme $p \cap X = p \cap \bigcap_{i \in M} P_i = \bigcap_{i \in M}(p \cap P_i)$, a členy posledního průniku jsou polopřímky v p, případně celá p. Protože $p \cap X$ je omezený, je to úsečka s koncovými body, které jsou koncové body některé z polopřímek a tedy leží v R_i pro nějaké $i \in M$.) Nyní ovšem $\boldsymbol{x} \in \operatorname{conv}\{\boldsymbol{y}, \boldsymbol{z}\}$, $\boldsymbol{y} \in X_i = \operatorname{conv}(V_i) \subseteq \operatorname{conv}(V)$ a $\boldsymbol{y} \in X_j = \operatorname{conv}(V_j) \subseteq \operatorname{conv}(V)$, a tedy $\boldsymbol{x} \in \operatorname{conv}(V)$, což jsme potřebovali dokázat.

(2) Je dána konečná množina V, dokazujeme, že $\operatorname{conv}(V)$ je konvexní mnohostěn: Intuitivní idea důkazu je následující. Vezmeme všechny lineární nerovnosti, které jsou splněny pro všechny body V. Z věty o oddělování plyne, že jejich průnik je $\operatorname{conv}(V)$, protože každý bod mimo $\operatorname{conv}(V)$ lze od $\operatorname{conv}(V)$ oddělit nadrovinou, tedy nerovností. Ukážeme, že množina nerovností ve vhodné reprezentaci je mnohostěn, a proto z nich podle implikace dokázané v části (1) můžeme vybrat konečně mnoho tak, že všechny ostatní z nich plynou. Tyto vybrané nerovnosti pak popisují $\operatorname{conv}(V)$ jako konvexní mnohostěn. Drobná technická komplikace je v tom, že potřebujeme, aby mnohostěn reprezentující nerovnosti byl omezený; vzhledem k možnosti přenásobení nerovností můžeme předpokládat, že všechny koeficienty jsou omezené.

Nerovnice reprezentujeme vektory v $\mathbb{R}^{n+1}$, kde vektor $\begin{pmatrix} \boldsymbol{a} \\ b \end{pmatrix}$ reprezentuje nerovnici $\boldsymbol{a}^T \boldsymbol{x} \leq b$. Označme

$$Q = \left\{ \begin{pmatrix} \boldsymbol{a} \\ b \end{pmatrix} \in \mathbb{R}^{n+1} \;\middle|\; \boldsymbol{a} \in [-1,1]^n \wedge b \in [-1,1] \wedge (\forall \boldsymbol{v} \in V)(\boldsymbol{a}^T \boldsymbol{v} \leq b) \right\}.$$

Množina Q je omezená díky podmínkám pro $\boldsymbol{a}$ a b. Navíc je Q dána konečně mnoha neostrými lineárními nerovnostmi: $2n$ nerovností $-1 \leq a_i \leq 1$, dvě nerovnosti $-1 \leq b \leq 1$, a pro každý z konečně mnoha bodů $\boldsymbol{v} \in V$ jedna nerovnost $\boldsymbol{a}^T \boldsymbol{v} \leq b$ (uvědomme si, že nyní $\boldsymbol{a}$ a b mají roli proměnných a $\boldsymbol{v}$ jsou koeficienty; nerovnici bychom také mohli zapsat $\boldsymbol{v}^T \boldsymbol{a} - b \leq 0$). Ukázali jsme, že množina Q je tedy omezený konvexní mnohostěn.

Podle části (1) tedy existuje konečná množina $W \subseteq Q$ taková, že $Q = \operatorname{conv}(W)$. Uvědomme si, že pokud bod $\boldsymbol{x}$ splňuje nerovnosti reprezentované vektory v W, pak splňuje i všechny nerovnosti reprezentované vektory v Q: vektor v Q je konvexní kombinací vektorů z W, a tedy odpovídající nerovnost získáme jako součet nerovností z W přenásobených koeficienty v konvexní kombinaci.

Chceme dokázat, že $\operatorname{conv}(V) = Y$ pro $Y = \left\{ \boldsymbol{x} \;\middle|\; \left(\forall \begin{pmatrix} \boldsymbol{a} \\ b \end{pmatrix} \in W\right)(\boldsymbol{a}^T \boldsymbol{x} \leq b) \right\}$. Z definice je Y průnikem konečně mnoha poloprostorů, a tedy $\operatorname{conv}(V)$ je pak konvexní mnohostěn; $\operatorname{conv}(V)$ je také omezený, protože V je konečná.

Inkluze $\operatorname{conv}(V) \subseteq Y$: Platí $V \subseteq Y$ z definice Q a tedy $\operatorname{conv}(V) \subseteq Y$, protože Y je konvexní.

Inkluze $\operatorname{conv}(V) \supseteq Y$: Nechť $x \notin \operatorname{conv}(V)$. Podle věty o oddělování existuje $\boldsymbol{a} \in \mathbb{R}^n$ a $b \in \mathbb{R}$ tak, že $\boldsymbol{a}^T \boldsymbol{x} > b$ a pro všechna $\boldsymbol{v} \in V$ platí $\boldsymbol{a}^T \boldsymbol{v} < b$. Koeficienty nerovnice $\boldsymbol{a}$ a b můžeme přeškálovat tak, že jejich absolutní hodnoty jsou nejvýš 1. Pak $\begin{pmatrix} \boldsymbol{a} \\ b \end{pmatrix} \in Q$ podle definice Q.

Bod $\boldsymbol{x}$ tedy nesplňuje jednu z nerovnic reprezentovanou v Q. Podle pozorování za definicí W pak také bod $\boldsymbol{x}$ nesplňuje jednu z nerovnic reprezentovanou v W a tedy $\boldsymbol{x} \notin Y$. Tím je inkluze dokázána. $\square$

6 Stěny mnohostěnů

Stěny konvexních mnohostěnů jsou geometricky definovány jako průnik s nadrovinou takovou, že celý mnohostěn leží v jednom poloprostoru touto nadrovinou definovaném. Z hlediska optimalizace rovnice nadroviny také definuje lineární funkci, která je v polytopu maximalizována právě na všech bodech stěny. Zatímco konvexní mnohostěny odpovídají množinám přípustných řešení lineárního programu, stěny mnohostěnů jsou tedy množiny optimálních řešení lineárních programů.

Definice 6.1. Nechť P je konvexní mnohostěn a $\boldsymbol{c} \in \mathbb{R}^n$, $t \in \mathbb{R}$. Jestliže $(\forall \boldsymbol{x} \in P)(\boldsymbol{c}^T\boldsymbol{x} \leq t)$ a $(\exists \boldsymbol{x} \in P)(\boldsymbol{c}^T\boldsymbol{x} = t)$, pak množinu $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{c}^T\boldsymbol{x} = t\}$ nazýváme **tečná nadrovina** [angl. *supporting hyperplane*], a množinu $\{\boldsymbol{x} \in P \mid c^T x = t\}$ nazýváme **stěna** P [angl. *face*].

Stěnami nazýváme také množiny $\emptyset$ a P. Stěnu F, pro kterou platí $F \neq P$ a $F \neq \emptyset$, nazýváme **vlastní stěna** [angl. *proper face*].

Stěna konvexního mnohostěnu je sama konvexním mnohostěnem, v popisu mnohostěnu stačí přidat nerovnici $\boldsymbol{c}^T\boldsymbol{x} \geq t$.

Stěny mnohostěnu můžeme klasifikovat podle jejich dimenze a podle toho zavést názvosloví. Stěny podle naší definice zahrnují např. u třídimenzionálních mnohostěnů i vrcholy a hrany, nejen to, čemu jsme zvyklí říkat stěny v obecné řeči. Náš intuitivní pojem stěny nahrazuje a zobecňuje pojem fasety.

Definice 6.2. Stěny podle dimenze klasifikujeme takto:

Vrchol [angl. *vertex*] konvexního mnohostěnu P je stěna dimenze 0.

Hrana [angl. *edge*] konvexního mnohostěnu P je stěna dimenze 1.

Faseta [angl. *facet*] konvexního mnohostěnu P je stěna dimenze $\dim(P) - 1$.

Jestliže tečná nadrovina definuje vlastní stěnu P, musí netriviálně protínat afinní obal P. Z toho plyne, že dimenze každé vlastní stěny je menší než $\dim(P)$. Fasety jsou tedy vlastní stěny maximální dimenze.

Nechť V_{ext} je množina bodů v P, které nemůžeme vyjádřit jako konvexní kombinaci jiných bodů z P; někdy se nazývají extremální body. Následující věta říká, že extremální body jsou právě vrcholy mnohostěnu. Navíc pro omezený mnohostěn existuje ještě třetí možná definice množiny vrcholů, a to jako jediné minimální množiny, jejíž konvexní obal je P. Větu dokážeme jen pro omezené mnohostěny.

Věta 6.3. *Nechť $P \subseteq \mathbb{R}^n$ je konvexní mnohostěn. Označme V množinu všech vrcholů P a $V_{ext} = \{\boldsymbol{x} \in P \mid \boldsymbol{x} \notin \operatorname{conv}(P \setminus \{\boldsymbol{x}\})\}$. Pak $V = V_{ext}$.*

Pokud je navíc mnohostěn P omezený, tak $P = \operatorname{conv}(V)$.

Důkaz. (Pro omezený mnohostěn P.)

Označme jako V_0 nějakou konečnou množinu splňující $\operatorname{conv}(V_0) = P$, a navíc takovou, že V_0 je mezi těmito množinami minimální vzhledem k inkluzi. Z Minkowského-Weylovy věty víme, že taková konečná množina existuje, a množinu minimální k inkluzi dostaneme postupným odebíráním prvků (s použitím konečnosti!). Existence V_0 je krok důkazu, kde využíváme omezenost mnohostěnu P.

Důkaz povedeme tak, že dokážeme tři inkluze $V \subseteq V_{ext} \subseteq V_0 \subseteq V$. Z toho plyne, že tři uvažované množiny jsou si rovné; navíc i to, že minimální množina V_0 je určena jednoznačně.

Inkluze $V \subseteq V_{ext}$: Nechť $\boldsymbol{z} \in V$. Z definice vrcholu víme, že existují $\boldsymbol{c} \in \mathbb{R}^n$ a $t \in \mathbb{R}$ takové, že $\boldsymbol{c}^T\boldsymbol{z} = t$ a $(\forall x \in P \setminus \{\boldsymbol{z}\})(\boldsymbol{c}^T\boldsymbol{x} < t)$. Pak podle pozorování 4.14 každé $\boldsymbol{y} \in \operatorname{conv}(P \setminus \{\boldsymbol{z}\})$ splňuje $\boldsymbol{c}^T\boldsymbol{y} < t$. Z toho plyne, že $\boldsymbol{z} \notin \operatorname{conv}(P \setminus \{\boldsymbol{z}\})$ a tedy $z \in V_{ext}$.

Inkluze $V_{ext} \subseteq V_0$: Dokážeme opačnou implikaci. Nechť $\boldsymbol{z} \notin V_0$. Pak $\boldsymbol{z} \in P = \operatorname{conv}(V_0) \subseteq \operatorname{conv}(P \setminus \{\boldsymbol{z}\})$ a tedy $\boldsymbol{z} \notin V_{ext}$.

Inkluze $V_0 \subseteq V$: Nechť $\boldsymbol{z} \in V_0$, cílem je dokázat, že $\boldsymbol{z}$ je vrchol P. Nechť $D := \operatorname{conv}(V_0 \setminus \{\boldsymbol{z}\})$. Pak $\{\boldsymbol{z}\}$ a D jsou disjunktní uzavřené konvexní množiny. Podle věty o oddělování máme $\boldsymbol{c} \in \mathbb{R}^n$ a $r \in \mathbb{R}$ takové, že $(\forall \boldsymbol{x} \in D)(\boldsymbol{c}^T\boldsymbol{x} < r$ a $\boldsymbol{c}^T\boldsymbol{z} > r$. Zvolíme $t = \boldsymbol{c}^T\boldsymbol{z}$. Dokážeme, že $L = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{c}^T\boldsymbol{x} = t\}$ je tečná nadrovina mnohostěnu P, která protíná P v jediném bodě $\boldsymbol{z}$.

Podle volby t platí pro všechna $\boldsymbol{x} \in V_0 \setminus \{\boldsymbol{z}\}$ ostrá nerovnost $\boldsymbol{c}^T\boldsymbol{x} < r < t$ a $\boldsymbol{c}^T\boldsymbol{z} = t$. Tedy pro všechna $\boldsymbol{x} \in V_0$ platí neostrá nerovnost $\boldsymbol{c}^T\boldsymbol{x} \leq t$, dále $\boldsymbol{z} \in L$ a tedy L je tečná nadrovina mnohostěnu $\operatorname{conv}(V_0) = P$. Podle pozorování 4.14 také pro všechna $\boldsymbol{y} \in \operatorname{conv}(V_0) \setminus \{\boldsymbol{z}\} = P \setminus \{\boldsymbol{z}\}$ platí ostrá nerovnost $\boldsymbol{c}^T\boldsymbol{y} < t$, protože bod $\boldsymbol{y} \neq \boldsymbol{z}$ při vyjádření jako konvexní kombinace bodů z V_0 musí mít kladný koeficient u některého bodu z $V_0 \setminus \{\boldsymbol{z}\}$, a pro něj neostrost platí ostře. Tím jsme dokázali, že $P \cap L = \{\boldsymbol{z}\}$ a tedy $\boldsymbol{z}$ je vrchol P. □

Následující dvě věty ukazují, že stěny mnohostěnů se v jistém smyslu chovají dle naší intuice: U třídimenzionálního mnohostěnu je průnik faset hrana mnohostěnu, jeho vrchol nebo prázdná množina, tedy opět stěna mnohostěnu. Stejně tak hrana či vrchol fasety jsou i stěnou mnohostěnu.

Z technického hlediska v obou větách potřebujeme ze dvou tečných nadrovin vyrobit jinou tečnou nadrovinu, což uděláme volbou vhodné lineární kombinace nerovností. Je dobře si uvědomit, že geometricky tato konstrukce odpovídá pootočení jedné nadroviny kolem průniku obou nadrovin. Druhou větu dokážeme pro omezené mnohostěny.

Věta 6.4. *Průnik stěn P je stěna P.*

Důkaz. Pokud průnik stěn je prázdný, věta platí triviálně. Stejně tak pokud jedna ze stěn je rovna P. Ve zbývajících případech jsou obě stěny vlastní.

Nechť tedy stěna E je dána tečnou nadrovinou $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{c}^T\boldsymbol{x} = t\}$ a stěna F je dána tečnou nadrovinou $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{d}^T\boldsymbol{x} = r\}$, všechny body $\boldsymbol{x} \in P$ splňují $\boldsymbol{c}^T\boldsymbol{x} \leq t$ a $\boldsymbol{d}^T\boldsymbol{x} \leq r$.

Ukážeme, že nadrovina $L = \{\boldsymbol{x} \in \mathbb{R}^n \mid (\boldsymbol{c}+\boldsymbol{d})^T\boldsymbol{x} = t + r\}$, tj. nadrovina daná součtem obou rovnic, je tečná nadrovina. Pro $\boldsymbol{x} \in E \cap F$ platí obě rovnosti, tedy i $(\boldsymbol{c}+\boldsymbol{d})^T\boldsymbol{x} = t+r$, a proto $E \cap F \subseteq L$. Součtem odpovídajících nerovnic ověříme, že všechny body $\boldsymbol{x} \in P$ splňují $(\boldsymbol{c}+\boldsymbol{d})^T\boldsymbol{x} \leq t + r$, L je tedy tečná nadrovina. Navíc všechny body $\boldsymbol{x} \in P \setminus (E \cap F)$ splňují jednu z nerovnic ostře, a tedy i $(\boldsymbol{c}+\boldsymbol{d})^T\boldsymbol{x} < t + r$. Dokázali jsme, že $P \cap L = E \cap F$ a tedy $E \cap F$ je stěna P. □

Věta 6.5 (Stěna stěny je stěna)**.** *Nechť $E \subseteq F$, F stěna konvexního mnohostěnu P. Pak E je stěna F, právě když E je stěna P.*

Důkaz. (Pro omezený mnohostěn P.)

Věta triviálně platí, jestliže $E = \emptyset$, $F = P$ nebo $E = F$. V ostatních případech jde ve znění věty o vlastní stěny.

Nejprve předpokládejme, že E je stěna P, a dokažme, že E je stěna F. Nechť $E = P \cap L$, kde L je tečná nadrovina. Pak $E \subseteq F \cap L \subseteq P \cap L = E$, množiny jsou si tedy rovné a E je stěna F definovaná stejnou tečnou nadrovinou L. (Nadrovina L je tečná nadrovina i pro F, protože $F \subseteq P$ a tedy i všechny body F leží ve stejném poloprostoru daném L.)

Nyní předpokládejme, že E je stěna F, a dokažme, že E je stěna P, pro omezený mnohostěn P. Nechť F je dáno tečnou nadrovinou jako $F = \{\boldsymbol{x} \in P \mid \boldsymbol{c}^T\boldsymbol{x} = t\}$ a pro všechna $\boldsymbol{x} \in P$ platí $\boldsymbol{c}^T\boldsymbol{x} \leq t$. Nechť E je dáno tečnou nadrovinou F, tedy $E = \{\boldsymbol{x} \in F \mid \boldsymbol{d}^T\boldsymbol{x} = r\}$ a pro všechna $\boldsymbol{x} \in F$ platí $\boldsymbol{d}^T\boldsymbol{x} \leq r$.

Uvažme nadrovinu $L_\alpha = \{\boldsymbol{x} \in \mathbb{R}^n \mid (\alpha\boldsymbol{c} + \boldsymbol{d})^T\boldsymbol{x} = \alpha t + r\}$ pro $\alpha > 0$. Ukážeme, že pro dostatečně velké α je L_α tečná nadrovina P a $E = L_\alpha \cap P$. Protože body $\boldsymbol{x} \in E$ leží v obou výchozích tečných nadrovinách, leží i v L_α pro všechna $\alpha > 0$. Body $\boldsymbol{x} \in F \setminus E$ splňují $\boldsymbol{c}^T\boldsymbol{x} = t$ a $\boldsymbol{d}^T\boldsymbol{x} < r$, a tedy i

$$(\alpha\boldsymbol{c} + \boldsymbol{d})^T\boldsymbol{x} < \alpha t + r \tag{6.1}$$

pro všechna $\alpha > 0$. Zbývá zvolit vhodné $\alpha > 0$ takové, že pro všechny body $\boldsymbol{x} \in P \setminus F$ platí (6.1).

Nechť V je množina vrcholů P, z omezenosti P víme, že $P = \operatorname{conv}(V)$. Pro $\boldsymbol{x} \in V \cap F$ platí

$$(\alpha\boldsymbol{c} + \boldsymbol{d})^T\boldsymbol{x} \leq \alpha t + r \tag{6.2}$$

podle předchozího odstavce. Pro $\boldsymbol{v} \in V \setminus F$ platí $\boldsymbol{c}^T\boldsymbol{x} < t$ a proto pro dostatečně velké α platí (6.1). Zvolíme α takové, že (6.1) platí pro všechna $\boldsymbol{x} \in V \setminus F$. Pak tedy (6.2) platí pro všechny vrcholy $\boldsymbol{x} \in V$. Vezměme nyní $\boldsymbol{x} \in P \setminus F$ a vyjádřeme ho jako konvexní kombinaci vrcholů V. Alespoň jeden z koeficientů u vrcholu z $V \setminus F$ bude kladný, posčítáním přenásobených nerovností (6.1) a (6.2) pro vrcholy tedy dostaneme nerovnost (6.1) pro $\boldsymbol{x}$. Tím jsme dokončili důkaz tvrzení $L_\alpha \cap P = E$ pro zvolené α a věta je dokázána. □

7 Minimální popis mnohostěnu

Mějme neprázdný konvexní mnohostěn zadaný následujícím popisem:

$$P = \{\boldsymbol{x} \in \mathbb{R} \mid A'\boldsymbol{x} = \boldsymbol{b}' \wedge A''\boldsymbol{x} \leq \boldsymbol{b}''\} \tag{7.1}$$

tj. pomocí m' rovnic a m'' nerovnic. Zajímá nás, kdy je takový popis nejjednodušší možný, tedy v nějaké „normální formě“, a jak takový popis souvisí s geometrickými vlastnostmi mnohostěnu. Na rozdíl od předchozích kapitol zde uvedené důkazy fungují i pro neomezené mnohostěny.

Definice 7.1. Popis (7.1) je **minimální popis mnohostěnu** P pokud (i) nemůžeme vynechat žádnou rovnost nebo nerovnost bez změny P a (ii) nemůžeme změnit žádnou nerovnost na rovnost bez změny P.

Minimalizovat část popisu s rovnicemi umíme pomocí lineární algebry. Má-li soustava lineárních rovnic řešení, můžeme vždy vybrat $\operatorname{rank}(A')$ nezávislých rovnic a ostatní vynechat. V minimálním popisu tedy platí $\operatorname{rank}(A') = m'$.

První pozorování ukazuje, že dimenze P je dána pouze systémem rovnic v minimálním popisu.

Pozorování 7.2. *Nechť $\boldsymbol{z} \in P$ splňuje $A''\boldsymbol{z} < \boldsymbol{b}''$ (tj. v každé nerovnici platí ostrá nerovnost). Pak*

- $\dim(P) = n - \operatorname{rank}(A')$
- *$\boldsymbol{z}$ neleží v žádné vlastní stěně P.*

Pro minimální popis P navíc takové $\boldsymbol{z}$ existuje, je-li $P \neq \emptyset$.

Důkaz. Nechť L je afinní prostor definovaný soustavou rovnic $A'\boldsymbol{x} = \boldsymbol{b}'$. Nechť $\varepsilon > 0$ a $B = \{\boldsymbol{x} \in L \mid ||\boldsymbol{x} - \boldsymbol{z}|| \leq \varepsilon\}$ je koule v L se středem $\boldsymbol{z}$ a poloměrem ε. Protože $\boldsymbol{z}$ splňuje všechny nerovnosti ostře, pro dostatečně malé $\varepsilon > 0$ je $B \subseteq P$. Koule B má dimenzi stejnou jako L, tedy $n - rank(A')$, protože v ní můžeme najít odpovídající počet afinně nezávislých bodů. Například tak, že vezmeme bázi příslušného vektorového prostoru a jejich malé násobky přičteme k $\boldsymbol{z}$.

Pokud $\boldsymbol{z}$ leží v nadrovině R, která neobsahuje celé P, pak v kouli B najdeme body na obou stranách nadroviny R. Proto R není tečná nadrovina a nedefinuje stěnu. Bod $\boldsymbol{z}$ tedy není v žádné vlastní stěně P.

V minimálním popisu pro každou nerovnici existuje $\boldsymbol{z}^{(i)} \in P$, takové, že $A''_{i*}\boldsymbol{z}^{(i)} < b''_i$. Vezmeme z těžiště těchto bodů, tj. $\boldsymbol{z} = \frac{1}{m''}\sum_{i=1}^{m''} \boldsymbol{z}^{(i)}$. Platí, že $\boldsymbol{z} \in P$, protože $\boldsymbol{z}$ je konvexní kombinací bodů z P. Navíc každá nerovnost platí ostře, protože pro jeden z bodů platí ostře a jeho váha je nenulová. (Je-li matice A'' prázdná, tj. v popisu nejsou žádné nerovnice, vezmeme $z \in P$ libovolné.) □

Následující věta a její důsledky dávají do souvislosti nerovnosti v minimálním popisu mnohostěnu s jeho stěnami v geometrickém smyslu. Součástí věty je tvrzení, že každá stěna leží v některé fasetě. To s minimálním popisem zdánlivě nesouvisí, ale uvědomme si, že dosud jsme ani nevěděli, že mnohostěn musí mít fasetu (pokud vůbec má vlastní stěnu, tj. pokud není roven nějakému afinnímu prostoru).

Věta 7.3. *V minimálním popisu (7.1) nerovnosti odpovídají vzájemně jednoznačně fasetám mnohostěnu P.*

Každá vlastní stěna mnohostěnu je stěnou některé fasety.

Důkaz. Pro $k = 1, \ldots, m''$ označme $R_k = \{\boldsymbol{x} \in \mathbb{R}^n \mid A''_{k*}\boldsymbol{x} = b''_k\}$ a $F_k = P \cap R_k$, kde A''_{k*} označuje k-tý řádek matice A''. Tedy F_k je podmnožina P definovaná tak, že k-tou nerovnici nahradíme rovnicí; v tomto popisu bude dána $m' + 1$ rovnicemi a $m'' - 1$ nerovnicemi. (Tento popis F_k nemusí být minimální.)

Nejprve ukážeme, že F_k je faseta pro dané k. Protože nadrovina R_k odpovídá původní nerovnosti, je to tečná nadrovina a F_k je stěna. Z minimality popisu (7.1) plyne, že existuje bod $\boldsymbol{x}$, který porušuje k-tou nerovnost, tj. $A''_{k*}\boldsymbol{x} > b''_k$, ale všechny ostatní podmínky z minimálního popisu splňuje. Z pozorování 7.2 plyne existence bodu $\boldsymbol{z} \in P$, který splňuje všechny nerovnosti (7.1) ostře. Jako konvexní kombinaci bodů $\boldsymbol{x}$ a $\boldsymbol{z}$ získáme bod $\boldsymbol{y}^{(k)} \in P$, který splňuje $A''_{k*}\boldsymbol{y}^{(k)} = b''_k$ a všechny ostatní nerovnosti splňuje ostře (pro existenci $\boldsymbol{y}$ použijeme také spojitost funkce $A''_{k*}\boldsymbol{y}$ na úsečce spojující $\boldsymbol{x}$ a $\boldsymbol{z}$ a pro ostré splnění nerovností to, že $\boldsymbol{z}$ nemůže mít v konvexní kombinaci váhu 0, protože $\boldsymbol{z}$ není v P). Je tedy $\boldsymbol{y}^{(k)} \in F_k$ a z pozorování 7.2 použitého pro popis F_k s $m' + 1$ rovnicemi plyne, že $\dim(F_k) \geq n - (m' + 1) = \dim(P) - 1$. Z minimality popisu (7.1) plyne, že F_k je vlastní stěna, tedy $\dim(F_k) < \dim(P)$; pak nutně $\dim(F_k) = \dim(P) - 1$ a F_k je faseta. Navíc vybraný bod $\boldsymbol{y}^{(k)}$ neleží v F_i pro $i \neq k$, a tedy fasety F_i jsou různé.

Nechť $E \subseteq P$ je vlastní stěna P. Ukážeme, že E je podmnožinou některé fasety F_i; z toho podle věty 6.5 plyne, že E je stěna F_i. Pokud neexistuje i tak, že $E \subseteq F_i$, tak pro každou nerovnost existuje bod $\boldsymbol{z}^{(i)} \in E$, který i-tou nerovnost splňuje ostře. Jejich těžiště $\boldsymbol{z} = \frac{1}{m''} \sum_{i=1}^{m''} \boldsymbol{z}^{(i)}$ leží v $E \subseteq P$ a splňuje všechny nerovnosti ostře, což je podle pozorování 7.2 ve sporu s tím, že $\boldsymbol{z}$ leží ve vlastní stěně.

Zbývá ukázat, že P nemá jiné fasety než F_i. Nechť E je faseta P. Podle předchozího odstavce je E stěna některé fasety F_i. Protože E i F_i jsou fasety P, je $\dim(E) = \dim(F_i) = \dim(P) - 1$. Tedy E nemůže být vlastní stěna F_i a $E = F_i$. □

Důsledek 7.4. *Pokud* $\dim(P) = n$, *tak existuje jediný minimální popis mnohostěnu* P *až na násobky nerovnic.*

Důkaz. Minimální popis obsahuje nerovnici pro každou fasetu. Faseta má dimenzi $n - 1$, proto je odpovídající tečná nadrovina rovna afinnímu obalu fasety, a tedy dána jednoznačně (až na násobek rovnice). Směr nerovnosti je také určený jednoznačně, protože existuje bod mnohostěnu, který neleží na fasetě.

Minimální popis neobsahuje žádnou rovnici vzhledem k plné dimenzi P. □

Důsledek 7.5. *Každá vlastní stěna konvexního mnohostěnu je průnikem jeho faset.*

Důkaz. Z věty víme, že každá vlastní stěna je obsažena v nějaké fasetě, která je dána popisem, kde jednu nerovnici změníme na rovnici. Z tohoto popisu získáme minimální vynecháním podmínek nebo změnou nerovnic na rovnice. Tento posup iterujeme, až dostaneme popis dané stěny. V postupu nevznikají nové nerovnice, pouze je zaměňujeme za rovnice tečných nadrovin, které odpovídají původním fasetám. Výsledek je tedy roven průniku mnohostěnu se všemi tečnými nadrovinami použitých faset, a tedy roven průniku těchto faset. (Protože stěna je vlastní, použili jsme aspoň jednu fasetu.) □

8 Simplexová metoda

Nyní předvedeme praktickou metodu na řešení úloh lineárního programování. Skrývá se za ní jednoduchá intuice. Hledáme řešení pouze ve vrcholech mnohostěnu, a pohybujeme se vždy z vrcholu do jiného vrcholu spojeného s ním hranou a to tak, že účelová funkce se nikdy nezhorší. Jde tedy o algoritmus lokálního prohledávání. Z konvexity mnohostěnu plyne, že se takto ke globálnímu optimu vždy můžeme dostat, tedy že není možné, že se ocitneme v lokálním optimu a nebudeme moci pokračovat. Vrcholům mnohostěnu odpovídají bázická přípustná řešení.

Definice 8.1. Bázické přípustné řešení lineárního programu je takové přípustné řešení, které splňuje n lineárně nezávislých podmínek s rovností. (Počítáme všechny rovnice, těsně splněné nerovnice a těsně splněné podmínky nezápornosti. Podmínky považujeme za lineárně nezávislé, jestliže jsou lineárně nezávislé vektory koeficientů u proměnných.)

Ukazuje se jako výhodné pracovat s maximalizační úlohou lineárního programování v rovnicovém tvaru:

$$\begin{aligned} \text{maximalizuj} \quad & \boldsymbol{c}^T\boldsymbol{x} \\ \text{pro} \quad & \boldsymbol{x} \geq 0 \\ \text{za podmínek} \quad & A\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

kde $A \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ a $\boldsymbol{c} \in \mathbb{R}^n$. Pro popis simplexové metody předpokládáme, že máme takový lineární program a také, že soustava rovnic má řešení a má plnou hodnost, tj. $\operatorname{rank}(A) = m$. Předpoklad o hodnosti matice snadno zajistíme tak, že metodami lineární algebry vynecháme závislé rovnice. Nalezení počátečního přípustného řešení se budeme věnovat později.

Podmínky nezápornosti v rovnicovém tvaru kromě jiného zaručují, že mnohostěn přípustných řešení a také každá jeho stěna mají vrchol. Je tedy dostatečné hledat optimum mezi bázickými přípustnými řešeními. Ta mají navíc pro úlohy v rovnicovém tvaru jednoduchý popis a řadu pěkných vlastností.

Označme $A_X = A_{*X}$ podmatici A se sloupci indexovanými X pro $X \subseteq \{1, \dots, n\}$.

Definice 8.2. Pro úlohu lineárního programování v rovnicovém tvaru definujeme:

Báze je množina indexů proměnných $B \subseteq \{1, \dots n\}$ taková, že A_B je regulární.

Bázické řešení odpovídající bázi B je řešení soustavy $A\boldsymbol{x} = \boldsymbol{b}$ takové, že $x_i = 0$ pro $i \notin B$.

Přípustná báze je taková, že odpovídající bázické řešení je přípustné.

Všimněme si, že báze má vždycky m prvků a bázi odpovídá jediné řešení. (Využíváme předpoklad, že $\operatorname{rank}(A) = m$.)

Lemma 8.3. *Nechť $\boldsymbol{x}$ je přípustné řešení a K je množina $K = \{i \mid x_i \neq 0\}$. Pak $\boldsymbol{x}$ je bázické právě tehdy, když A_K má lineárně nezávislé sloupce.*

Důkaz. Jestliže $\boldsymbol{x}$ je bázické řešení odpovídající bázi B, tak $K \subseteq B$.

Jestliže A_K má $|K| \leq m$ nezávislých sloupců, A má m lineárně nezávislých sloupců a podle věty o výměně doplníme K na bázi B. □

Pozorování 8.4. *Přípustná bázická řešení jsou právě vrcholy mnohostěnu přípustných řešení.*

Důkaz. Nechť B je přípustná báze a $i \notin B$. Přidáním podmínky $x_i = 0$ pro se množina přípustných řešení změní na stěnu (ne nutně vlastní) původního mnohostěnu přípustných řešení, protože původní mnohostěn splňuje $x_i \geq 0$ a $x_i = 0$ tedy definuje tečnou nadrovinu. (Průnik s touto nadrovinou je vždy neprázdný, protože B je přípustná báze.) Průnikem těchto stěn přes všechna $i \notin B$ tedy získáme stěnu původního mnohostěnu, protože průnik stěn je stěna. Tato stěna je jednoprvková, protože z definice báze obsahuje jediné bázické řešení odpovídající B, je to tedy nutně vrchol.

Naopak každý vrchol můžeme získat jako jediné řešení poté, co některé rovnice změníme na nerovnice. To ukážeme z minimálního popisu mnohostěnu: Nejprve nerovnice lineárního programu upravíme na minimální popis tím, že některé nerovnice změníme na rovnice nebo vynecháme. (Také vynecháme některé rovnice, ale to není podstatné.) Z minimálního popisu mnohostěnu víme, že vrchol je průnikem faset a každou fasetu získáme změněním nejaké nerovnice na rovnici. Jediné nerovnice jsou $x_i \geq 0$, nově jsme tedy přidali jen podmínky $x_i = 0$ pro nějakou množinu proměnných. Zbylé sloupce matice A jsou lineárně nezávislé, protože soustava má jediné řešení. Lze je tedy doplnit na bázi podle předchozího lemmatu. □

Upravíme $Ax = b$ tak, že $A_B = I_m$. Tento stav si budeme udržovat.

Definice 8.5. Simplexová tabulka určená bází B je soustava $m + 1$ lineárních rovnic pro proměnné $x_1, \dots x_n, z$, která má stejná řešení jako soustava $Ax = b$, $z = c^T x$ a navíc splňuje $A_B = I_m$

Zapisujeme a značíme $x_B = p + Qx_N$, kde N jsou všechny nebázické proměnné ($N = \{1, \dots, n\} \setminus B$), $z = z_0 + r^T x_N$ a matice koeficientů u x_B je I_m. Kde $p \in \mathbb{R}^m$, $Q \in \mathbb{R}^{m \times (n-m)}$, $z_0 \in \mathbb{R}$, $r \in \mathbb{R}^{n-m}$.

Pozorování 8.6. *Báze B je přípustná (= odpovídající řešení je přípustné) $\Leftrightarrow$ $p \geq 0$.*

Pozorování 8.7. *Řešení odpovídající přípustné bázi B je optimální $\Leftarrow$ $r \leq 0$.*

Pozorování 8.8. (*$\forall$ bázi B*) $p = A_B^{-1} b$, $Q = -A_B^{-1} A_N$, $z_0 = c_B^T A_B^{-1} b$, $r = c_N - (c_B^T A_B^{-1} A_N)^T$

Důkaz. $Ax = b$ upravíme na $A_B^{-1} Ax = A_B^{-1} b$, x rozdělíme na x_B a x_N, $z = c^T x = c_B^T x_B + c_N^T x_N = c_B^T \cdot (A_B^{-1} b - A_B^{-1} A_N x_N) + c_N^T x_N = c_B^T A_B^{-1} b + (c_N^T - c_B^T A_B^{-1} A_N) \cdot x_N$. □

Lemma 8.9. *Pokud je úloha v rovnicovém tvaru omezená a přípustná, pak má optimální bázické řešení.*

Důkaz. Ke každému řešení existuje bázické řešení se stejnou nebo vyšší hodnotou účelové funkce. □

8.1 Pivotovací krok

Mějme bázické proměnné $B = \{j_1, \dots j_m\}$, nebázické proměnné $N = \{l_1, \dots l_{n-m}\}$. B' je nová báze tak, že $B' = B \cup \{l_t\} \setminus \{j_s\}$.

Vstupující proměnná: l_t taková $r_{l_t} > 0$ (z nich vybereme pomocí pivotovacího pravidla).

Vystupující proměnná: $q_{st} < 0$ a $-\frac{p_s}{q_{st}}$ je minimální z $-\frac{p_i}{q_{it}}$ pro $i \in B$, $q_{it} < 0$.

Úprava tabulky: Převedeme vstupující proměnnou x_{l_t} vlevo a dosadíme, nebo to popíšeme formálně - tedy přičtením $-\frac{p_s}{q_{st}} \cdot q_{it}$ násobku s-tého řádku k i-tému.

Pokud **neexistuje vstupující proměnná**: $\Rightarrow$ konec, B dává optimální řešení.

Pokud **neexistuje vystupující proměnná**: $\Rightarrow$ konec, ale úloha je neomezená.

Pivotovací pravidla:

- maximální r_{l_t} (Dantzig)
- Spočítat maximální výslednou hodnotu účelové funkce, tedy maximální z_0, což ale může být náročné.
- Užít kompromis: **nejstrmější hranu**: spočítáme si maximální $c^t \cdot \frac{x \text{ nové } - x \text{ staré}}{||x \text{ nové } - x \text{ staré } ||}$, kde "$x$ nové" zvolíme například takové, že $x_{l_t} = 1$. Tento postup je praktický, protože většinou trvá malý počet kroků.
- **Blandovo pravidlo**: Vybereme nejmenší možný přípustný index l_t a k němu vybereme nejmenší možný index j_s. Tento postup necyklí.
- **Lexikografické pravidlo**: Vezmeme libovolnou vstupující proměnnou t; pro každé s, kandidáta na vystupující proměnnou spočítáme vektor $\left(\frac{q_{s1}}{q_{st}}, \frac{q_{s2}}{q_{st}}, \dots \frac{q_{s \cdot (n-m)}}{q_{st}}\right)$ a vybereme s s lexikograficky nejmenším vektorem. Tento postup necyklí.
- l_t vybereme **náhodně** mezi těmi, které můžeme použít.

Hledání počátečního řešení:

$$\begin{aligned} \text{maximalizuj} \quad & c^T x \\ \text{pro} \quad & x \geq 0 \\ \text{za podmínek} \quad & Ax = b \end{aligned}$$

Pomocná úloha: úprava taková, aby b bylo nezáporné, pak:

$$\begin{aligned} \text{maximalizuj} \quad & -(x_{n+1}, \dots x_{n+m}) \\ \text{pro} \quad & \bar{x} \geq 0,\ \bar{x} \in \mathbb{R}^{n+m} \\ \text{za podmínek} \quad & \bar{A}\bar{x} = b,\ \bar{A} \in \mathbb{R}^{m \times (n+m)},\ \bar{A} = (A|I_m) \end{aligned}$$

$(0, \dots 0, b_1, \dots b_m)$ je bázické řešení $\bar{A}\bar{x} = b$
$(x_1, \dots x_n, 0, \dots 0)$ řešení pomocné úlohy $\Rightarrow$ řešení původní úlohy.

9 Dualita lineárního programování

Získáme duální proměnné y_i pro každou primární podmínku $A_{i.}$ (i-tý řádek A):

$$\begin{aligned} y_i \in \mathbb{R} \quad & \text{pro} \quad \sum_{j=1}^{n} a_{ij} x_j = b_i \\ y_i \geq 0 \quad & \text{pro} \quad \sum_{j=1}^{n} a_{ij} x_j \leq b_i, \text{pokud primární maximalizuje} \\ y_i \leq 0 \quad & \text{pro} \quad \sum_{j=1}^{n} a_{ij} x_j \leq b_i, \text{pokud primární minimalizuje} \end{aligned}$$

Duální podmínka pro každou primární proměnnou x_j:

$$\begin{aligned} \sum_{i=1}^{m} a_{ij} y_i = c_j \quad & \text{pro} \quad x_j \in \mathbb{R} \\ \sum_{i=1}^{m} a_{ij} y_i \geq c_j \quad & \text{pro} \quad x_j \geq 0, \text{pokud primární maximalizuje} \\ \sum_{i=1}^{m} a_{ij} y_i \leq c_j \quad & \text{pro} \quad x_j \geq 0, \text{pokud primární minimalizuje} \end{aligned}$$

Účelová funkce

$$\begin{aligned} \min \sum_{i=1}^{m} b_i y_i \quad & \text{pokud primární je } \max c^T x \\ \max \sum_{i=1}^{m} b_i y_i \quad & \text{pokud primární je } \min c^T x \end{aligned}$$

Pozorování 9.1. *Duální LP k duálnímu LP je původní LP.*

Příklad 1. Speciální případy pro $a \in \mathbb{R}^{m\times n}, c \in \mathbb{R}^n, b \in \mathbb{R}^m$:

$$(\mathrm{P}) \quad \begin{array}{ll} \max & c^T x \\ & x \in \mathbb{R}^n \\ & Ax \le b \end{array} \qquad (\mathrm{D}) \quad \begin{array}{ll} \min & b^T y \\ & y \ge 0 \\ & A^T y = c \end{array}$$

$$(\bar{\mathrm{P}}) \quad \begin{array}{ll} \max & c^T x \\ & x \ge 0 \\ & Ax \le b \end{array} \qquad (\bar{\mathrm{D}}) \quad \begin{array}{ll} \min & b^T y \\ & y \ge 0 \\ & A^T y \ge c \end{array}$$

Věta 9.2 (slabá o dualitě)**.** *Nechť x a y jsou přípustná řešení (P) a (D) pak $c^T x \le b^T y$.*

Důkaz.

$$\begin{aligned} y^T Ax &\le y^T b \quad \text{z (P)}, y \ge 0 \\ y^T Ax &= c^T x \quad \text{z (D)} \end{aligned}$$

□

Věta 9.3 (o dualitě)**.** *Pro úlohy (P) a (D) nebo $(\bar{P})$ a $(\bar{D})$ nastává jedna z možností:*

- *ani (P) ani (D) nemá přípustné řešení,*
- *jedna z nich je neomezená a druhá nepřípustná,*
- *(P), (D) mají přípustné řešení; pak navíc existují optimální řešení x^* a y^* a platí: $c^T x^* = b^T y^*$.*

Pozn: soustava rovnic $Ax = b$ má řešení $\Leftrightarrow$ $A^T y = 0$ a $b^T y < 0$ nemá řešení.

10 Farkasovo lemma

Farkasovo lemma je důležitá věta, která říká, že z lineárních rovnic a nerovnic můžeme získat spor jedině tak, že je přenásobíme koeficienty, u nerovnic nezápornými, sečteme, a dostaneme nepravdivou nerovnost $0 < 0$.

Farkasovo lemma se dá dokázat více způsoby. My ho dokážeme pomocí takzvané Fourier-Motzkinovy eliminace. To je metoda, kterou lze ze soustavy nerovnic s n proměnnými získat soustavu nerovnic s $n-1$ proměnnými tak, že se zachová řešitelnost soustavy, navíc nová soustava má nerovnice, které jsou součtem vždy dvou přenásobených nerovnic původní soustavy. Tj. z řešitelné soustavy dostaneme řešitelnou, z neřešitelné neřešitelnou. Tuto redukci použijeme pro indukční krok v důkazu netriviální implikace lemmatu.

Pro propojení technického znění lemmatu s výše uvedenou intuicí si uvědomme, že $\boldsymbol{y}$ označuje váhy jednotlivých nerovnic, tj. koeficienty, kterými násobíme nerovnice soustavy před sčítáním. Podmínky na $\boldsymbol{y}$ pak říkají, že (i) váhy nerovnic jsou nezáporné, (ii) koeficienty lineárních členů výsledné nerovnice jsou 0 u všech proměnných a (iii) absolutní člen výsledné nerovnice je záporný.

Věta 10.1 (Farkasovo lemma pro nerovnice)**.** *Nechť $A \in \mathbb{R}^{m\times n}$ a $\boldsymbol{b} \in \mathbb{R}^n$. Soustava $A\boldsymbol{x} \le \boldsymbol{b}$ má řešení, právě když neexistuje $\boldsymbol{y} \in \mathbb{R}^m$ tak, že (i) $\boldsymbol{y} \ge 0$, (ii) $A^T\boldsymbol{y} = 0$ a (iii) $\boldsymbol{b}^T\boldsymbol{y} < 0$.*

Důkaz. Jestliže existuje řešení $\boldsymbol{x}$ a zároveň požadované $\boldsymbol{y}$, pak sečteme nerovnosti soustavy $A\boldsymbol{x} \leq \boldsymbol{b}$ s nezápornými koeficienty $\boldsymbol{y}$. Formálně $0 = \boldsymbol{y}^T A\boldsymbol{x} \leq \boldsymbol{y}^T \boldsymbol{b} < 0$, a dostáváme spor. Dokázali jsme tedy implikaci, že pokud soustava $A\boldsymbol{x} \leq \boldsymbol{b}$ má řešení, tak požadované $\boldsymbol{y}$ neexistuje.

Pro důkaz opačné implikace předpokládáme, že soustava $A\boldsymbol{x} \leq \boldsymbol{b}$ nemá řešení a najdeme požadované $\boldsymbol{y}$. Postupujeme indukcí podle n.

Pro $n = 0$ neexistence řešení znamená, že soustava obsahuje nerovnici $0 \leq \boldsymbol{b}_i$ s $\boldsymbol{b}_i < 0$. Zvolíme $\boldsymbol{y}_i = 1$ a $\boldsymbol{y}_j = 0$ pro $j \neq i$.

Pro $n \geq 1$ nejprve upravíme soustavu do vhodného tvaru. Bez újmy na obecnosti předpokládáme, že nerovnice mají u x_n koeficienty $a_{in} \in \{-1, 0, 1\}$. Toho můžeme dosáhnout přenásobením jednotlivých nerovnic $1/|a_{in}|$ (pro $a_{in} \neq 0$) a tato transformace nezmění existenci $\boldsymbol{y}$, protože jednotlivá y_i naopak přenásobíme $|a_{in}|$. Navíc předpokládáme, že řádky matice jsou seřazeny dle znaménka v posledním sloupci. Je-li $\bar{A}$ matice A bez posledního sloupce a $\bar{\boldsymbol{x}}$ vektor $\boldsymbol{x}$ bez poslední složky, můžeme pak soustavu označit (S1) a pro vhodné m' a m'' psát takto:

$$\begin{aligned} \bar{A}_{i*}\bar{\boldsymbol{x}} + \boldsymbol{x}_n &\leq b_i & \qquad & i = 1, \ldots, m' \\ \bar{A}_{i*}\bar{\boldsymbol{x}} - \boldsymbol{x}_n &\leq b_i & & i = 1 + m', \ldots, m'' \\ \bar{A}_{i*}\bar{\boldsymbol{x}} &\leq b_i & & i = 1 + m'', \ldots, m, \end{aligned}$$

Kombinací řádků pomocné soustavy vytvoříme novou soustavu (S2):

$$\begin{aligned} (\bar{A}_{i*} + \bar{A}_{j*})\bar{\boldsymbol{x}} &\leq b_i + b_j & \qquad & i = 1, \ldots, m', \quad j = 1 + m', \ldots, m'' \\ \bar{A}_{i*}\bar{\boldsymbol{x}} &\leq b_i & & i = 1 + m'', \ldots, m, \end{aligned}$$

Plán důkazu je takový, že ukážeme, že pokud (S1) nemá řešení, pak ani (S2) nemá řešení, pak z indukčního předpokladu existuje $\boldsymbol{y}'$ pro soustavu (S2) a konečně z tohoto $\boldsymbol{y}'$ zkonstruujeme $\boldsymbol{y}$ pro soustavu (S1).

Pro první krok dokážeme opačnou implikaci, tedy z řešení $\boldsymbol{x}'$ soustavy (S2) zkonstruujeme řešení soustavy (S1). To uděláme tak, že zvolíme $\boldsymbol{x}_i = \boldsymbol{x}'_i$ pro $i = 1, \ldots, n-1$ a $\boldsymbol{x}_n = \min_{i \in \{1,\ldots,m'\}}(b_i - \bar{A}_{i*}\boldsymbol{x}')$. První skupina nerovnic (S1) platí, neboť $x_n \leq \boldsymbol{b}_i - \bar{A}_{i*}\boldsymbol{x}'$ podle volby x_n. Zároveň existuje $i \in \{1, \ldots, m'\}$ takové, že $x_n = b_i - \bar{A}_{i*}\boldsymbol{x}'$, tuto rovnost můžeme pro každé $j \in \{m'+1, \ldots, m''\}$ odečíst od nerovnice v (S2) $\bar{A}_{i*} + \bar{A}_{j*})\bar{\boldsymbol{x}} \leq b_i + b_j$, čímž získáme $\bar{A}_{j*}\bar{\boldsymbol{x}} - \boldsymbol{x}_n \leq b_j$ a tedy i druhá skupina nerovnic (S1) platí. Třetí skupina nerovnic (S1) plyne z (S2) triviálně.

Pro poslední krok důkazu vezměme dle indukčního předpokladu $\boldsymbol{y}'$ pro soustavu (S2). Vektor $\boldsymbol{y}'$ udává nezápornou kombinaci nerovnic (S2), která vede ke sporu. Vzhledem k tomu, že každá nerovnice (S2) je součtem dvou nebo jedné nerovnic (S1), snadno získáme spor také jako nezápornou kombinaci nerovnic (S1); koeficienty této kombinace jsou požadovaný vektor $\boldsymbol{y}$. Postupujme teď o něco přesněji. Vektor $\boldsymbol{y}'$ má složky odpovídající nerovnicím (S2), tedy jednak složky y'_{ij} pro každou dvojici $i = 1, \ldots, m'$ a $j = 1 + m', \ldots, m''$ a jednak složky y'_i pro každé $i = 1 + m'', \dots, m$. Zvolíme $\boldsymbol{y}$ takto:

$$
\begin{aligned}
y_i &= \sum_{j=m'+1}^{m''} y'_{ij}, && \text{pro} \quad i = 1, \dots, m' \\
y_j &= \sum_{i=1}^{m''} y'_{ij}, && \text{pro} \quad j = m'+1, \dots, m'' \\
y_i &= y'_i && \text{pro} \quad i = m''+1, \dots, m.
\end{aligned}
$$

Platí $\boldsymbol{y} \geq \boldsymbol{0}$, protože $\boldsymbol{y}' \geq \boldsymbol{0}$. Není těžké ověřit, že kombinace nerovnic (S1) s koeficienty $\boldsymbol{y}$ dává stejnou nerovnici jako kombinace nerovnic (S2) s koeficienty $\boldsymbol{y}'$. Podle indukčního předpokladu je to nerovnice dávající spor. □

Věta 10.2 (Farkasovo lemma)**.** *Nechť $A \in \mathbb{R}^{m \times n}$ a $\boldsymbol{b} \in \mathbb{R}^n$. Soustava $A\boldsymbol{x} = \boldsymbol{b}$ má nezáporné řešení, právě když neexistuje y takové, že $A^T\boldsymbol{y} \geq 0$ a zároveň $\boldsymbol{y}^T\boldsymbol{b} < 0$.*

Důkaz. Použijeme Farkasovo lemma pro nerovnice na ekvivalentní soustavu $2m + n$ nerovnic $A'\boldsymbol{x} \leq \boldsymbol{b}'$ daných takto.

$$
A' = \begin{pmatrix} A \\ -A \\ -I_n \end{pmatrix} \in \mathbb{R}^{(2m+n) \times n} \qquad b' = \begin{pmatrix} b \\ -b \\ 0 \end{pmatrix} \in \mathbb{R}^{(2m+n)}
$$

Soustava $A\boldsymbol{x} = \boldsymbol{b}$ má nezáporné řešení, právě když $A'\boldsymbol{x} \leq \boldsymbol{b}'$ má řešení. Jestliže soustava $A'\boldsymbol{x} \leq \boldsymbol{b}'$ nemá řešení, pak existuje $\boldsymbol{y}' \geq 0$ takové, že $A'^T\boldsymbol{y}' = 0$ a $\boldsymbol{b}'^T\boldsymbol{y}' < 0$. Z něj zkonstruujeme $\boldsymbol{y}$ tak, že vezmeme $y_i = y'_i - y'_{i+m}$ pro $i = 1, \dots, m$. □

Věta 10.3. *Nechť $A\boldsymbol{x} \leq \boldsymbol{b}$ má řešení. Pak každé řešení $A\boldsymbol{x} \leq \boldsymbol{b}$ splňuje $\boldsymbol{c}^T\boldsymbol{x} \leq \delta$, právě když existuje $\boldsymbol{y} \geq 0$ takové, že $A^T\boldsymbol{y} = \boldsymbol{c}$ a $\boldsymbol{b}^T\boldsymbol{y} \leq \delta$.*

Důkaz.
"$\Leftarrow$"... $\boldsymbol{y}$ dává koeficienty, kterými násobíme nerovnice $A\boldsymbol{x} \leq \boldsymbol{b}$, dostaneme $\boldsymbol{c}^T\boldsymbol{x} = \boldsymbol{y}^T A x \leq \boldsymbol{b}^T\boldsymbol{y} \leq \delta$
"$\Rightarrow$"... neexistuje $\boldsymbol{y} \Longrightarrow$ najdu řešení $A\boldsymbol{x} \leq \boldsymbol{b}$ nesplňující $\boldsymbol{c}^T\boldsymbol{x} \leq \delta$
Neexistuje $\boldsymbol{y} \geq 0$ a $\lambda \geq 0$:

$$
\begin{aligned}
A^T\boldsymbol{y} &= \boldsymbol{c} && \cdot \boldsymbol{z} \\
b^T\boldsymbol{y} + \lambda &= \delta && \cdot \mu
\end{aligned}
$$

Podle Farkasova lemmatu existuje $\boldsymbol{z}$ a μ takové, že platí $A\boldsymbol{z} + \boldsymbol{b}\mu \geq 0$ (podmínka pro koeficienty proměnných $\boldsymbol{y}$), $\mu \geq 0$ (podmínka pro koeficienty proměnné λ) a $\boldsymbol{c}^T\boldsymbol{z} + \delta\mu < 0$ (podmínka pro pravou stranu).

- $\mu > 0$

$$
\begin{aligned}
\boldsymbol{x} &:= -\tfrac{1}{\mu}\boldsymbol{z} \\
A\boldsymbol{x} &\leq \boldsymbol{b} \\
\boldsymbol{c}^T\boldsymbol{x} &> \delta
\end{aligned}
$$

- $\mu = 0$

$$\begin{aligned} \mathbf{x_0} \text{ řešení } \mathbf{Ax_0} &\leq \boldsymbol{b} && (i) \\ A\boldsymbol{z} &\geq 0 && (ii) \\ \boldsymbol{c}^T\boldsymbol{z} &< 0 \\ \boldsymbol{x} &= \mathbf{x_0} - \mathbf{M} \cdot \mathbf{z} \end{aligned}$$

z (i) a (ii) vyplývá: $A\boldsymbol{x} \leq \boldsymbol{b}$
Pro $M \gg 0$ dostáváme $\boldsymbol{c}^T\boldsymbol{x} = \boldsymbol{c}^T \cdot \mathbf{x_0} - \mathbf{Mc^T} \cdot \mathbf{z} \to \infty$
(pro libovolné δ vždy najdu dostatečně velké M) □

10.1 Dualita

Označme si speciální případy úloh LP (a úlohy k nim duální):

$$(P) \quad \begin{aligned} \text{maximalizuj} \quad & \boldsymbol{c}^T\boldsymbol{x} \\ \text{pro} \quad & \boldsymbol{x} \in \mathbb{R}^n \\ \text{za podmínek} \quad & A\boldsymbol{x} \leq \boldsymbol{b} \end{aligned} \qquad (D) \quad \begin{aligned} \text{minimalizuj} \quad & \boldsymbol{b}^T\boldsymbol{y} \\ \text{pro} \quad & \boldsymbol{y} \geq 0 \\ \text{za podmínek} \quad & A^T\boldsymbol{y} = \boldsymbol{c} \end{aligned}$$

$$(\bar{P}) \quad \begin{aligned} \text{maximalizuj} \quad & \boldsymbol{c}^T\boldsymbol{x} \\ \text{pro} \quad & \boldsymbol{x} \geq 0 \\ \text{za podmínek} \quad & A\boldsymbol{x} \leq \boldsymbol{b} \end{aligned} \qquad (\bar{D}) \quad \begin{aligned} \text{minimalizuj} \quad & \boldsymbol{b}^T\boldsymbol{y} \\ \text{pro} \quad & \boldsymbol{y} \geq 0 \\ \text{za podmínek} \quad & A^T\boldsymbol{y} \geq \boldsymbol{c} \end{aligned}$$

Věta 10.4 (o dualitě)**.** *Pro (P) a (D) $[(\overline{P})a(\overline{D})]$ platí jedna z možností:*
(1) (P) i (D) jsou nepřípustné
(2) Jedna je neomezená, druhá nepřípustná
(3) Existují optima $\mathbf{x}^*$ a $\mathbf{y}^*$ úloh (P) a (D) splňující $\boldsymbol{c}^T\mathbf{x}^* = \mathbf{b^T y}^*$

Důkaz. Nechť (P) má řešení, $\delta = \sup\{\boldsymbol{c}^T\boldsymbol{x} \mid A\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \in \mathbb{R}^n\}$
Slabá věta o dualitě: $\delta \leq \boldsymbol{b}^T\boldsymbol{y}$ pro každé přípustné řešení (D)
Pokud $\delta < \infty$, tak podle **Věty 2.** existuje $\boldsymbol{y} \geq 0$: $A^T\boldsymbol{y} = \boldsymbol{c}$, $\boldsymbol{b}^T\boldsymbol{y} \leq \delta$
Takže $\boldsymbol{y}$ je optimální řešení (D). □

Příklad 2. kdy případ *(1)* může nastat:

$$\begin{aligned} \text{maximalizuj} \quad & x_2 + x_3 \\ \text{za podmínek} \quad & \begin{aligned} x_1 &\leq -1 \\ -x_1 &\leq -1 \\ x_2 - x_3 &\leq 0 \end{aligned} \end{aligned} \quad \overset{dualita}{\rightsquigarrow} \quad \begin{aligned} \text{minimalizuj} \quad & -y_1 - y_2 \\ \text{za podmínek} \quad & \begin{aligned} y_1 - y_2 &\geq 0 \\ y_3 &\geq 1 \\ -y_3 &\geq 1 \end{aligned} \end{aligned}$$

Věta 10.5 (o komplementaritě)**.** *Nechť $\boldsymbol{x}$ a $\boldsymbol{y}$ jsou přípustná řešení (P) a (D). Pak jsou obě optimální, právě když platí:*
$(*)$ $(\forall j \in \{1, \ldots, m\})(y_j = 0 \vee a^{(j)}x = b_j)$ *($a^{(j)}$ je j-tý řádek a)*
Nechť $\boldsymbol{x}$ a $\boldsymbol{y}$ jsou přípustné řešení $(\overline{P})$ a $(\overline{D})$.
Pak jsou obě optimální $\iff$ platí (*) a (**)
$(**)$ $(\forall i \in \{1, \ldots, n\})(x_i = 0 \vee \sum_{j=1}^{m} a_{ij}y_j = c_j)$

Důkaz. Pro (P) a (D):
$\boldsymbol{y}^T\boldsymbol{b} - \boldsymbol{c}^T\boldsymbol{x} = \boldsymbol{y}^T \cdot (\boldsymbol{b} - A\boldsymbol{x}) = \sum_{j=1}^{n} y_j \cdot (\boldsymbol{b} - a^{(j)}\boldsymbol{x}) \geq 0$ ... rovnost nastává $\iff$ platí $(*)$
Pro $(\overline{P})$ a $(\overline{D})$:
$\boldsymbol{y}^T\boldsymbol{b} - \boldsymbol{c}^T\boldsymbol{x} = \boldsymbol{y}^T\boldsymbol{b} - \boldsymbol{y}^T A\boldsymbol{x} + \boldsymbol{y}^T A\boldsymbol{x} - \boldsymbol{c}^T\boldsymbol{x} = \boldsymbol{y}^T \cdot (\boldsymbol{b} - A\boldsymbol{x}) + (\boldsymbol{y}^T A - \boldsymbol{c}^T) \cdot \boldsymbol{x} \geq 0$ ... rovnost nastává $\iff$ platí $(*)$ a $(**)$. □

11 Perfektní párování minimální ceny

V této kapitole předvedeme polynomiální algoritmus pro hledání prefektního párování minimální ceny, tedy pro následující problém:

> PERFEKTNÍ PÁROVÁNÍ MINIMÁLNÍ CENY
> *Vstup*: Neorientovaný graf $G = (V, E)$, ceny hran $c_e \geq 0$ pro $e \in E$.
> *Výstup*: Perfektní párování, tj. $M \subseteq E$ takové, že $(\forall v) deg_M(v) = 1$, kde $deg_M(v)$ je stupeň vrcholu v v M.
> *Cíl*: Minimalizovat $\sum_{e \in M} c_e$.

Jde o slavný Edmondsův kvítkový algoritmus (blossom algorithm) ze šedesátých let, který znamenal nejen průlom v teorii párování a lineárního programování, ale také podstatně přispěl k pozdějšímu zavedení pojmu třídy P jako abstrakci efektivních algoritmů.

Protože jde o poměrně složitý algoritmus, předvedeme ho ve třech krocích. Nejprve předvedeme verzi pro bipartitní grafy, na které uvidíme základní myšlenku primárně duálního algoritmu; také budeme potřebovat myšlenku alternujících stromů pro hledání zlepšující cesty, která je základem kombinatorických algoritmů pro hledání maximálního párování. Poté předvedeme neváženou verzi kvítkového algoritmu pro obecné grafy, kde oproti bipartitnímu párování potřebujeme pracovat s kontrakcemi lichých cyklů. Nakonec obě myšlenky propojíme netriviálním způsobem do vážené verze pro obecné grafy.

V definici problému si všimněme, že omezení na nezáporné ceny není podstatné: Vzhledem k tomu, že všechna párování mají stejně hran, můžeme ke všem cenám přičíst konstantu bez změny optimálních řešení. Naopak pro primárně-duální algoritmus bude toto omezení šikovné pro hledání počátečního duálního řešení.

Pokud chceme hledat maximální párování místo perfektního, můžeme použít snadnou redukci: Použijeme úplný graf a hranám neobsaženým v původním grafu dáme stejnou dostatečně velkou cenu; případně také přidáme vrchol, aby byl počet vrcholů sudý. Optimální řešení pak nutně obsahuje minimální počet nových hran, v původním grafu je to tedy maximální párování minimální ceny.

V používaných lineárních programech se budou používat proměnné x_e, které v celočíselném řešení indikují, zda $e \in M$. Pro množinu hran $F \subseteq E$ označíme $x(F) = \sum_{e \in F} x_e$. Pro $S \subseteq V$ označíme $\delta(S) = \{e \in E \mid |e \cap S| = 1\}$, tj. $\delta(S)$ je množina hran, které jdou z vrcholů S ven, nazývá se hranice S nebo řez daný S. Budeme zkracovat $\delta(\{v\})$ na $\delta(v)$. Je tedy speciálně $x(\delta(v)) = \sum_{e: v \in e} x_e$, což odpovídá stupni vrcholu v. Neorientovanou hranu $\{u, v\}$ budeme zkráceně označovat uv.

11.1 Perfektní bipartitní párování minimální ceny

Nejprve se omezíme na bipartitní grafy. Primární lineární program pro perfektní bipartitní párování už známe:

Primární LP:

$$\begin{aligned} \text{minimalizuj} \quad & \sum_{e \in E} c_e \cdot x_e \\ \text{pro} \quad & x_e \geq 0 \\ \text{za podmínek} \quad & x(\delta(v)) = \sum_{e: v \in e} x_e = 1 \end{aligned}$$

Celočíselné přípustné řešení odpovídá perfektnímu párování $M = \{e \in E \mid x_e = \cancel{0}\}$ a účelová funkce jeho ceně. Víme, že matice programu je totálně unimodulární, a program tedy má celočíselné optimum. Mohli bychom lineární program vyřešit a tím najít optimum. Naším cílem je ale primárně duální algoritmus, který je jednak efektivnější a jednak půjde zobecnit na nebipartitní grafy. K tomu potřebujeme duální lineární program a podmínky komplementarity:

Duální LP:

$$\begin{aligned} \text{maximalizuj} \quad & \sum y_v \\ \text{pro} \quad & y_v \in \mathbb{R} \\ \text{za podmínek} \quad & \forall uv \in E : y_u + y_v \leq c_{uv} \end{aligned}$$

Komplementarita: Jestliže pro $uv \in E$ platí $x_{uv} > 0$, pak $y_u + y_v = c_{uv}$.

V primárně duálním algoritmu udržujeme duálně přípustné řešení $\boldsymbol{y}$ a párování M. Párování M nemusí být perfektní; budeme ho postupně zvětšovat a teprve ve chvíli, kdy je M perfektní, tak se odpovídající $\boldsymbol{x}$ stane primárně přípustným řešením. Budeme však zachovávat podmínky komplementarity. Jakmile algoritmus najde perfektní párování M a duálně přípustné $\boldsymbol{y}$ splňující podmínky komplementarity, víme, že M je optimální řešení.

Pro dané $\boldsymbol{y}$ definujme

$$E_= = \{uv \in E \mid y_u + y_v = c_{uv}\}.$$

Podmínky komplementarity nyní říkají, že $M \subseteq E_=$.

Algoritmus začne s nulovým $\boldsymbol{y}$, což je přípustné řešení, protože ceny jsou nezáporné. Pro dané $\boldsymbol{y}$ se snaží najít párování v množině $E_=$ pomocí standardního kombinatorického algoritmu, který hledá střídavé cesty z nějakého nespárovaného vrcholu, tj. cesty, na kterých se střídají hrany mimo M a hrany z M. Pokud najde takovou cestu do jiného nespárovaného vrcholu, zvětší párování. Střídavý strom je standardní datová struktura, která reprezentuje všechny vrcholy, do kterých existuje střídavá cesta.

Definice 11.1. Nechť $T = (V(T), E(T), r)$ je strom s kořenem r, který je podgrafem G. Označme $A(T)$ množinu vrcholů T v liché vzdálenosti od r a $B(T)$ množinu vrcholů T v sudé vzdálenosti od r. (Tedy $r \in B(T)$; $A(T)$ a $B(T)$ tvoří rozklad $V(T)$.)

Strom T nazýváme **střídavý strom**, jestliže každý vrchol $u \in A(T)$ má právě jednoho syna v a navíc $uv \in M$.

Všimněme si, že všechny vrcholy T kromě r jsou spárované s jiným vrcholem v T a tyto hrany párování jsou zároveň hranami T. Také platí $|B(T)| = |A(T)| + 1$.

Pokud je $\boldsymbol{y}$ optimální, najdeme perfektní párování v $E_=$ (to ovšem ještě musíme dokázat). Pokud $\boldsymbol{y}$ optimální není, perfektní párování neexistuje. To se v algoritmu projeví tak, že do střídavého stromu nemůžeme přidat další hranu. (Např. hned na začátku je $E_=$ prázdná, jestliže jsou všechny ceny hran kladné.) Pak musíme $\boldsymbol{y}$ změnit opatrně tak, aby hrany v M a

$E(T)$ zůstaly v $E_=$. To bude platit, pokud y_v zvětšíme o ε pro $v \in A(T)$ a zmenšíme o stejné ε pro $v \in B(T)$, protože součty na hranách T zůstanou zachovány. Zároveň se tímto zvýší duální účelová funkce o ε. Musíme ovšem zvolit ε takové, aby podmínky duálního lineárního programu odpovídající hranám vedoucím z T mimo T zůstaly splněny, zároveň chceme aby některá z těchto podmínek byla nově splněna s rovností, abychom získali novou hranu $E_=$, kterou můžeme přidat do T.

Algoritmus: Perfektní párování minimální ceny v bipartitních grafech

(1) Inicializace: $\boldsymbol{y} = \mathbf{0}$, $M = \emptyset$.

(2) Inicializace stromu:
Jestliže M je prefektní párování, pak vystup M.
Jinak zvol $r \in V$ libovolný vrchol nespárovaný v M a $T = (\{r\}, \emptyset, r)$

(3) Budování stromu:
Zvol libovolnou $e = vw \in E_=$, $v \in B(T)$, $w \notin V(T)$, jestliže existuje, jinak goto (4).
Jestliže w je v M spárovaný s vrcholem z, přidej do T vrcholy w, z a hrany vw, wz; goto (3).
Jinak zvětšíme párování za pomoci střídavé cesty z r do w (z M uber všechny hrany na cestě v T z r do v a přidej do M hranu vw a všechny hrany na cestě, které v něm nebyly); goto (2).

(4) Změna $\boldsymbol{y}$:
$\epsilon = min\{c_{vw} - y_v - y_w \mid vw \in E, v \in B(T), w \notin V(T)\}$
Pokud $\epsilon = \infty$ (tj. minimum je přes prázdnou množinu), vystup "neexistuje perfektní párování".

$$y_v := \begin{cases} y_v + \epsilon & v \in B(T) \\ y_v - \epsilon & v \in A(T) \\ y_v & v \notin T \end{cases}$$

goto (3)

Věta 11.2. *Algoritmus nalezne perfektní párování minimální ceny v polynomiálním čase.*

Důkaz. Invarianty algoritmu jsou dva: (i) $\boldsymbol{y}$ je duálně přípustné a (ii) $M \cup E(T) \subseteq E_=$. Ověříme jejich zachování.

Po inicializaci algoritmu a každé inicializaci stromu invarianty zjevně platí (s použitím předpokladu nezáporných cen pro přípustnost $\boldsymbol{y} = \mathbf{0}$). Při budování stromu do $M \cup E(T)$ přidáváme jedině hranu vw, která je v $E_=$ dle podmínky. Do $E(T)$ přidáváme i hranu wz, která však už je v M a tedy v $E_=$; obdobně zlepšení párování podle střídavé cesty do M přidá jen hrany z $E_=$. V těchto krocích se $\boldsymbol{y}$ nemění, přípustnost se tedy zachová.

Při změně $\boldsymbol{y}$ se může zvýšit $y_u + y_v$ jen pro hrany uv, kde jeden vrchol je v $B(T)$ a druhý mimo $V(T)$, pro ty je podmínka duálního LP zachována díky volbě ε. Díky bipartitnosti grafu neexistují hrany s oběma vrcholy v $B(T)$. Pro hrany $uv \in E(T)$ i pro hrany M se $y_u + y_v$ nezmění, zůstanou tedy v $E_=$. Invarianty jsou tedy zachovány.

Pokud algoritmus skončí v kroku (2), z invariantů plyne, že výsledné perfektní párování splňuje podmínky komplementarity a je tedy optimální. Pokud algoritmus skončí v kroku (4), je upravené $\boldsymbol{y}$ přípustné pro libovolně velké ε; duální úloha je tedy neomezená, primární je nutně nepřípustná a tedy graf G nemá perfektní párování.

Zbývá dokázat, že algoritmus se zastaví po polynomiálně mnoha krocích. (Je lehké ověřit, že každý krok se dá implementovat v polynomiálním čase.) Především si uvědomme, že po každém kroku (4) následuje krok (3), protože za hranu vw můžeme vybrat tu hranu, která v kroku (4) definovala v minimu ε. V každém kroku (3) se zvětší M nebo T. Párování M se nikdy nezmenšuje, zvětšit se může maximálně n-krát, kde n je počet vrcholů v jedné partitě grafu. Strom T se zmenší pouze při zvětšení M, mezi zvětšeními M se T může zvětšit maximálně n-krát. Celkem tedy krok (2) proběhne maximálně n-krát a kroky (3) a (4) maximálně n^2-krát. Po tomto množství kroků se algoritmus nutně zastaví. □

Poznamenejme, že důkazem, že se algoritmus zastaví, jsme zároveň dokázali, že primární LP má celočíselné optimum (pokud je přípustný).

11.2 Perfektní párování v obecných grafech

Mějme G' .. G multigraf s kontrahovanými lichými cykly.

Definice 11.3. G/C .. ztotožníme vrcholy C do nového pseudovrcholu z, hrany $\{u, v\}$, $u, v \in C$ vynecháme a hrany $\{u, v\}$, $u \in C$, $v \notin C$ nahradíme $\{z, v\}$.

Pozorování 11.4. *Nechť C lichý cyklus, M' perfektní párování v G/C. Pak existuje $M \supseteq M'$ perfektní párování v G.*

Blossom algorithm

(i) $M = \emptyset$, $G' = G$

(ii) r nespárovaný $\rightarrow$ nový T (když neexistuje $\rightarrow$ párování)

(iii) $\{u, v\} \in E$, $u \in B(T)$, $v \notin T \rightarrow$ zvětšíme M nebo T

(iv) $\{u, v\} \in E$, $u, v \in B(T)$
P_u, P_v jsou cesty v T ke společnému předchůdci
$C = P_u \cup P_v \cup \{\{u, v\}\}$
$G' = G'/C$, $T = T/C$, $M = M/C$

Lze naimplementovat v čase $\mathcal{O}(mn \log n)$.

11.3 Perfektní párování minimální ceny v obecném grafu

Definice 11.5. $D \subseteq E$ je **lichý řez**, pokud $D = \delta(S)$ pro nějaké $S \subseteq V$, $|S|$ je lichá.

Primární LP:

$$\begin{array}{rrll}
\text{minimalizuj} & \displaystyle\sum_{e \in E} c_e \cdot x_e & & \\
\text{pro} & x_e \geq 0 & & \\
\text{za podmínek} & (\forall v \in V) & x(\delta(v)) & = 1 \\
 & (\forall S \subseteq E, |S| \text{ lichá}) & x(\delta(S)) & \geq 1
\end{array}$$

$D = \delta(S)$ ... lichý řez

Duální LP:

$$\begin{array}{rl} \text{maximalizuj} & \sum y_v + \sum Y_D \\ \text{pro} & Y_D \geq 0 \\ \text{za podmínek} & (\forall e = uv \in E)\bar{c}_e = c_e - y_u + y_v + \sum\limits_{D \ni uv} Y_D \geq 0 \end{array}$$

Definice 11.6. Redukovaná cena $\bar{c}_e = c_e - y_u - y_v - \sum Y_D$, kde $e \in \{u, v\}$

Komplementarita:

(i) $e \in M \implies \bar{c}_e = 0$

(ii) $Y_D > 0 \implies |M \cap D| = 1$

$E_= = \{e \in E \mid \bar{c}_e = 0\}$ hledáme $M, T \subseteq E_=$
Pracujeme v G' s kontrahovanými cykly, $u \in G'$ pseudovrchol, máme podmínku $y_u \geq 0$.

Algoritmus

(i) $y \equiv 0$, $M = \emptyset$, $G' = G$

(ii) r nespárovaný v M
$T = \{\{r\}, \emptyset\}$
$T, M \subseteq E_=$
pokud r neexistuje $\to$ M je perfektní $\to$ optimum

(iii) Nechť $\exists e = \{v, w\} \in E_=$, $v \in B(T)$, $w \notin T$ $\to$ budujeme M, T

(iv) $\{u, v\} \in E_=$: $u, v \in B(T)$: $C = P_u \cup P_v \cup \{u, v\}$
$G' = G'/C$ nový pseudovrchol $z \in B(T)$ .. sudá vzdálenost od kořene
nová duální proměnná y_z s podmínkou $y_z \geq 0$
$y_z = 0$
$(\forall e) e = \{a, b\}$, $a \in C$, $b \notin C$, $c_e = c_e - y_a$

(v) $z \in A(T)$ je pseudovrchol, $y_z = 0$
expandujeme z na cyklus C
do T a M přidáme sudou cestu z C
$(\forall e) e = \{a, b\}$, $a \in C$, $b \notin C$, $c_{ab} = c_{ab} + y_a$

(vi) změníme duální řešení y takto:

$$y_v := \begin{cases} y_v + \epsilon & v \in B(T) \\ y_v - \epsilon & v \in A(T) \\ y_v & v \notin T \end{cases}$$

kde

$$\epsilon \text{ je minimum z } \begin{cases} \bar{c}_{uv} & u \in B(T), v \notin T, uv \in E \\ \frac{\bar{c}_{uv}}{2} & u, v \in B(T), uv \in E \\ y_v & v \in A(T) \text{ je pseudovrchol} \end{cases}$$

Algoritmus končí, když nalezne perfektní párování. Pokud je ϵ neomezené párování neexistuje.

Věta 11.7. *Algoritmus nalezne perfektní párování minimální ceny v obecném grafu v polynomiálním čase.*