Polynomiální algoritmy pro lineární programování

Petr Kolman

Prosinec 2012

Předmluva

Tento text se snaží čtenáři se základními znalostmi lineární algebry, matematické analýzy a lineárního programování elementárním způsobem představit dvě metody pro řešení úloh lineárního programování v polynomiálním čase, totiž elipsoidovou metodu a metodu vnitřního bodu (resp. jednu z jejích mnoha variant ze skupiny metod snižování potenciálu).

Děkuji Martinu Perglovi za výchozí verzi poznámek k elipsoidivé metodě podle běhu přednášky Matematické programování v roce 2006/2007 a Dušanovi Knopovi za elektronickou podobu jeho poznámek k metodě vnitřního bodu z běhu téže přednášky ve školním roce 2008/2009.

Obsah

Kapitola 1

Velmi stručný historický úvod

Soustavám lineárních nerovnic a teorii mnohostěnů, včetně jejich aplikací, věnovali matematikové pozornost přinejmenším od 19. století. Jmenujme alespoň Fouriera, ještě před ním Lagrangea a Bernoulliho, později Farkase a Minkowského. Za počátek nového oboru nazývaného lineární programování se pokládá ale až druhá polovina 40. let 20. století. Matematické základy oboru tehdy položili především von Neumann a trojice matematiků Gale, Kuhn, Tuckera. Jejich hlavním přínosem bylo vybudování teorie duality lineárního programování; von Neumann ve svém přístupu využil výsledky, které získal při práci na jiném novém oboru rozvíjejícím se v té době, teorii her. Ještě před nimi se nezávisle na sobě souvisejícím problémum věnovali v Sovětském svazu Kantorovič a ve Spojených státech Hitchcock; jejich práce zůstaly žel řady let nedoceněny.

Klíčovým momentem pro vývoj lineárního programování byl Dantzigův simplexový algoritmus (publikován 1947), který se na mnoho let stal nezbytnou součástí oboru. Nezanedbatelným podnětem, který rozvoj lineárního programování podstatně ovlivnil a podpořil, byl příchod počítačů, z nichž první přišly na svět právě ve 40. letech. Uspokojivé praktické výsledky simplexového algoritmu byly výborným doporučením pro jeho široké použití v praxi a lineární programování se simplexovým algoritmem se na dlouhou dobu stalo nezbytným arzenálem ekonomů. Za zmínku stojí v této souvislosti skutečnost, že Kantorivič a Koopmans, kteří výrazně přispěli k rozvoji lineárního programování, obdrželi společně v roce 1975 Nobelovu cenu za ekonomii.

Téměř odděleně probíhal vývoj týkající se obecnějších optimalizačních problémů. V tomto světě se dočkaly velkého rozkvětu v 60. letech 20. století postupy známé dnes především pod jménem *metody vnitřního bodu* (Frisch, Fiacco, McCormick). Algoritmy navržené v této době však často trpěly numerickou nestabilitou.

S novým podnětem k dalšímu rozvoji lineárního programování přišla na počátku 70. let teorie složitosti. Nejprve se řada lidí pokoušela dokázat, že simplexový algoritmus pracuje v polynomiálním čase. V roce 1972 ovšem Klee a Minty popsali příklad ukazující exponenciální časovou složitost pro některé varianty simplexového algoritmu; brzy přibyly podobné výsledky i pro jiné varianty algoritmu. To podnítilo zájem o alternativní postupy na řešení úloh lineárního programování. Zlom přichází v roce 1979, kdy ruský matematik Chačijan publikoval v ruštině krátký článek (bez důkazů) o tom, jak lze pomocí tzv. *elipsoidové metody* řešit úlohu lineárního programování v polynomiálním čase; samotná metoda byla navržena již dříve v 70. letech pro problémy konvexního nelineárního programování (Yudin, Nemirovski, Šor). Výsledek měl nebývalý mediální ohlas, žel se ale záhy ukázalo, že k praktickému použití se metoda nehodí. Hledání alternativ simplexového algoritmu pokračovalo.

Dalším mezníkem v historii lineárního programování byl rok 1984, kdy Karmakar ze společnosti AT&T navrhl algoritmus spadající do skupiny *metod vnitřního bodu* a dokázal, že doba jeho běhu je polynomiální. Jak jsme již zmínili výše, samotná metoda nebyla nová, hlavní Karmarkarův přínos byl v aplikaci metody na úlohu lineárního programování a zejména v analýze časové složitosti algoritmu. Za zmínku stojí, že velmi podobný algoritmus pro lineární programování navrhl (1967) a posléze (1974) dokázal jeho konvergenci k optimálnímu řešení Kantorovičův student, ruský matematik Dikin; jeho práce žel zůstala dlouho nepovšimnuta. Od roku 1984 byly publikovány doslova tisíce článků týkajících se metod vnitřního bodu. Velmi zhruba se dají rozdělit na metody snižování potenciálu (sem patří Karmakarův algoritmus a také algoritmus popsaný v tomto textu), metody sledování centrální cesty (ty se dnes zdají nejlepší pro praktické použití a hojně se používají v softwarových balících pro lineární programování, především algoritmy tzv. Mehrotrova typu) a metody afinních transformací (u těch se většinou dokazuje pouze konvergence k optimálnímu řešení, nikoli polynomiální odhad na časovou složitost); mnohé varianty také najdete pod jménem bariérové metody. Mimo jiné se ukázalo, že tyto metody jsou použitelné i pro některé jiné problémy konvexní optimalizace, především pro semidefinitní programování. Nepřímým důsledkem Karmakarovy práce je i to, že došlo k většímu propojení dvou vědeckých komunit: komunity ekonomů věnujících se lineárnímu programování, a komunity matematiků studujících nelineární optimalizaci.

Na závěr si ještě všimněme některých rozdílů mezi simplexovým algoritmem, elipsoidovou metodou a metodou vnitřního bodu. Simplexový algoritmus chodí po hranici mnohostěnu, totiž po jeho vrcholech. Elipsoidový algoritmus se (kromě posledního kroku) pohybuje vně mnohostěnu. Metoda vnitřního bodu na rozdíl od obou ostatních metod pracuje uvnitř mnohostěnu. Svou podstatou je simplexový algoritmus diskrétním algoritmem, zatímco elipsoidová metoda i metoda vnitřního bodu je iteračním algoritmem.

Pozornosti zvídavého čtenáře doporučujeme ještě následující výsledky (a jeden problém):

- Silně polynomiální algoritmus pro lineární programy s koeficienty 1, -1, 0 – Tardos [8].
- Silně polynomiální algoritmus pro lineární programy v omezené dimenzi.
- Zdůvodnění, proč simplexový algoritmus potřebuje „většinou" pouze polynomiální čas – Spielman a Teng [7].
- Hirschova (polynomiální) hypotéza - do optimálního vrcholu vždy vede „krátká" cesta po hranici mnohostěnu.

Kapitola 2

Elipsoidová metoda

2.1 Náčrt algoritmu

Zadání: Pro daný mnohostěn $P = \{x \mid Ax \leq b\}$ najít $x \in P$, pokud existuje, a v opačném případě rozpoznat, že P je prázdný.

V tomto textu popíšeme tzv. elipsoidovou metodu, což je algoritmus, který daný problém řeší v polynomiálním čase vzhledem k velikosti zadání (podrobnosti níže).

Pro zjednodušení popisu algoritmu budeme prozatím o mnohostěnu P předpokládat:

P1 P je omezený (tj. celý P se vejde do koule o dostatečně velkém průměru),

P2 Je-li P neprázdný, pak je plné dimenze (tj. pro dostatečně malé $\epsilon > 0$ se do P vejde celá koule s průměrem ϵ).

Hrubý náčrt algoritmu: Najdeme dostatečně velký elipsoid E_0 takový, aby obsahoval celý mnohostěn P (předpoklad P1 zaručuje, že takový elipsoid existuje) a opakovaně provádíme následující. Pokud střed z_k elipsoidu E_k leží uvnitř P, našli jsme požadované řešení. V opačném případě víme, že z_k porušuje nějakou nerovnost ze systému $Ax \leq b$, řekněme i-tou nerovnost $a_i x \leq b_i$. Vezmeme poloprostor odpovídající porušené nerovnosti a posuneme ho tak, aby jeho hranice procházela bodem z_k; označme H_k tento posunutý poloprostor, tedy $H_k = \{x \mid a_i x \leq a_i z_k\}$. Pak celý mnohostěn P leží v průniku elipsoidu E_k s poloprostorem H_k. Najdeme nový (menší) elipsoid E_{k+1} obsahující $E_k \cap H_k$ a krok opakujeme.

Jak poznáme, že je P prázdný a že můžeme přestat s konstrukcí nových elipsoidů? Využijeme náš druhý předpoklad P2. Je-li P neprázdný, je možno na základě A a b spočítat dolní odhad na objem P. Pokud zajistíme, že objem elipsoidu E_{k+1} je vždy výrazně menší než objem elipsoidu E_k, pak po dostatečném počtu kroků budeme vědět, že neprázdný P by se již díky svému objemu do elipsoidu E_l, pro dostatečně velké l, nevešel a tudíž P musí být prázdný.

Abychom dostatečně popsali elipsoidovou metodu (za výše uvedených předpokladů P1 a P2), potřebujeme zodpovědět tyto otázky:

1. Jak na počátku volit elipsoid E_0 tak, aby $P \subseteq E_0$?

2. Jak ze starého elipsoidu E_k spočítat nový E_{k+1}, neboli jak volit parametry nového, aby obklopoval, co má, a přitom byl co do objemu podstatně menší?

3. Kolik iterací je třeba provést k nalezení $x \in P$ nebo k rozpoznání, že P je prázdné?

Dále bude třeba ukázat, jak se předpokladů P1 a P2 zbavit. Než se ale těmto otázkám budeme věnovat, zavedeme a připomeneme některé základní potřebné pojmy a skutečnosti.

2.2 Terminologie a opakovaní z lineární algebry

Opakovaně budeme pracovat s vektory $\mathbf{0} = (0, 0, \ldots, 0)^T$ a $e_1 = (1, 0, \ldots, 0)^T$. Porovnání vektorů (případně matic) provádíme po složkách: pro vektory $x, y \in \mathbb{R}^n$ píšeme $x \leq y$, pokud $x_i \leq y_i$ pro $i = 1, \ldots, n$. Jednotkovou matici značíme I.

Definice 1 *Symetrickou matici $A \in \mathbb{R}^{n \times n}$ nazýváme positivně definitní, jestliže $\forall x \neq 0$ platí $x^T A x > 0$.*

Tvrzení 2 *Pro pozitivně definitní matici A platí:*

- *Všechna vlastní čísla matice A jsou kladná,*
- *inverzní matice A^{-1} je také pozitivně definitní,*
- *existuje právě jedna pozitivně definitní matice Q taková, že $A = Q^T \cdot Q$; takovou matici Q značíme $A^{1/2}$.*

Pro nás bude důležitá poslední vlastnost.

Definice 3 *Množinu*

$$E(a, A) = \{x \in \mathbb{R}^n : (x - a)^T A^{-1} (x - a) \leq 1\} ,$$

kde A je pozitivně definitní matice, budeme nazývat elipsoidem se středem v a a určeným maticí A.

Příklad 4 $E(\boldsymbol{0}, I) = \{x \in \mathbb{R}^n \mid x^T x \leq 1\}$ *– jednotková koule se středem v počátku.*

Definice 5 *Zobrazení $T : \mathbb{R}^n \to \mathbb{R}^n$ nazýváme* afinní zobrazení, *pokud je dáno předpisem $T(x) = Qx + s$, kde $Q \in \mathbb{R}^{n \times n}$ je matice a $s \in \mathbb{R}^n$ je vektor (tj. T je složením lineárního zobrazení a posunutí).*

Tvrzení 6 *Každý elipsoid je afinním obrazem jednotkové koule: $E(a, A) = A^{1/2} \cdot E(\boldsymbol{0}, I) + a$.*

Důkaz. $E(a, A) = \{x \in \mathbb{R}^n : (x - a)^T A^{-1} (x - a) \leq 1\} = \{x \in \mathbb{R}^n : (x - a)^T \cdot A^{-1/2^T} \cdot A^{-1/2} \cdot (x - a) \leq 1\}$. Označme $y = A^{-1/2} \cdot (x - a)$. Potom $A^{1/2} y + a = x$. Dosadíme-li do formule popisující elipsoid, získáme (uvědomme si, že $A^{-1/2}$ je také positivně definitní matice, tudíž je symetrická a $A^{-1/2} = A^{-1/2^T}$):

$$E(a, A) = \{(A^{1/2} y + a) \in \mathbb{R}^n : y^T y \leq 1\} = A^{1/2} \cdot E(\mathbf{0}, I) + a .$$

□

Pro $T \subseteq \mathbb{R}^n$ budeme $vol(T)$ označovat objem množiny T.

Tvrzení 7 *Pro afinní zobrazení $T : \mathbb{R}^n \to \mathbb{R}^n$ dané předpisem $T(x) = Qx + t$ a množinu $X \subseteq \mathbb{R}^n$ platí:*

$$vol(T(X)) = |\det Q| \cdot vol(X) \ .$$

V případě elipsoidů dostaneme rovnost:

$$vol(E(a, A)) = |\det(A^{1/2})| \cdot vol(E(\boldsymbol{0}, I)).$$

Pozorování 8 *Pro objem elipsoidu $E(a, A)$ platí:*

$$|\det(A^{1/2})| \cdot n^{-n} \leq \ vol(E(a, A)) \leq |\det(A^{1/2})| \cdot 2^n \ .$$

Odhady se opírají o objem krychle vepsané, resp. opsané.

Definice 9 Velikost zápisu. *Pro celé číslo k značíme $\langle k \rangle = 1 + \lceil \log_2(|k| + 1) \rceil$, což je počet bitů potřebných pro zápis k včetně znaménka. Podobně pro racionální číslo $r = \frac{p}{q}$, kde p a q jsou nesoudělná, značíme $\langle r \rangle = \langle p \rangle + \langle q \rangle$. Pro racionální vektor $v \in \mathbb{Q}^n$ značíme $\langle v \rangle = \sum_{i=1}^{n} \langle v_i \rangle$ a pro matici $A \in \mathbb{Q}^{m \times n}$ obdobně $\langle A \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} \langle a_{i,j} \rangle$.*

Je užitečné uvědomit si, že pro každé přirozené číslo r platí nerovnost $|r| \leq 2^{\langle r \rangle - 1} - 1$.

2.3 Počáteční elipsoid

Začneme připomenutím známé nerovnosti poskytující horní odhad na velikost determinantu matice.

Lemma 10 (Hadamardova nerovnost) *Pro každou čtvercovou matici C platí:*

$$|\det C| \leq \Pi_{i=1}^{n} ||C_i|| \ ,$$

kde C_i označuje i-tý sloupec matice C a $||u||$ velikost vektoru u.

V důkazu hlavního výsledku této kapitolky bude naším nástrojem následující lemma.

Lemma 11 *Nechť C je celočíselná matice $n \times n$. Pak $|\det C| \leq 2^{\langle C \rangle - n^2}$*

Důkaz.

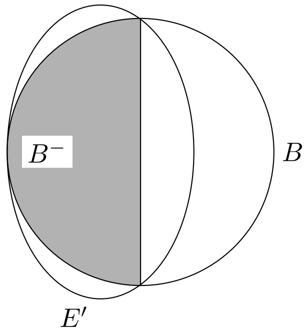
$$1 + |\det C| \leq 1 + \Pi_{i=1}^{n} ||C_i|| \leq \Pi_{i=1}^{n} (1 + ||C_i||) \leq \Pi_{i=1}^{n} 2^{||C_i|| - n} = 2^{\langle C \rangle - n^2} \ .$$

V úpravách jsme kromě Hadamardovy nerovnosti použili pro velikost n-složkového racionálního vektoru x odhad $||x|| \leq 2^{\langle x \rangle - n} - 1$, odvozený takto:

$$1 + ||x|| \leq 1 + \sum_{i=1}^{n} |x_i| \leq \Pi_{i=1}^{n} (1 + |x_i|) \leq \Pi_{i=1}^{n} (1 + (2^{\langle x_i \rangle - 1} - 1)) \leq \Pi_{i=1}^{n} 2^{\langle x_i \rangle - 1} = 2^{\langle x \rangle - n} \ .$$

□

Lemma 12 *Je-li $P = \{x \mid Cx \leq d\} \subseteq \mathbb{Q}^n$ omezený mnohostěn a C a d jsou celočíselné, pak všechny vrcholy P jsou obsaženy v kouli se středem v počátku a poloměrem $R = \sqrt{n} \cdot 2^{\langle C, d \rangle + n \log n}$.*



Obrázek 2.1: Hledané E'.

Důkaz. Uvažme libovolný vrchol $\mathbf{v}$ polytopu P. Z teorie mnohostěnů víme (např. [1]), že pak existuje podsystém nerovností $C'x \leq d'$ (vybraný ze systému nerovností $Cx \leq d$) takový, že $\mathbf{v}$ je jediným řešením $C'x = d'$. Z Cramerova pravidla dostáváme: $v_i = \frac{\det(C'_i)}{\det(C')}$, kde C'_i je matice vzniklá z C' nahrazením i-tého sloupce vektorem d'. Protože matice C' je celočíselná a regulární, máme $|\det(C')| \geq 1$. Pro horní odhad velikosti v_i užijeme následující horní odhad velikosti determinantu. K dokončení důkazu Lemmatu 12 si všimneme, že $\langle C'_i \rangle \leq \langle C, d \rangle$. Zbývá upočítat délku vektoru o všech složkách velikosti nejvýše $2^{\langle C,d \rangle + n \log n}$: $||\mathbf{v}|| \leq \sqrt{\sum_{i=1}^{n} (2^{\langle C,d \rangle + n \log n})^2} = \sqrt{n} \cdot 2^{\langle C,d \rangle + n \log n}$, což je hledaný poloměr R. $\square$

Jako počáteční elipsoid E_1 použijeme tedy kouli $E(0, R \cdot I)$.

2.4 Iterační krok: výpočet E_{k+1} z E_k

2.4.1 Snadný případ: $E_k = E(\mathbf{0}, I)$ a $H_k = \{x \mid x_1 \leq 0\}$

V této části předpokládáme, že $E_k = E(\mathbf{0}, I)$ a $H_k = \{x \mid x_1 \leq 0\}$. Pro zjednodušení zápisu (ušetření indexů) budeme používat značení B místo E_k (jedná se o jednotkovou kouli kolem počátku) a H místo H_k. Máme tedy $B = \{x \in \mathbb{R}^n | x^T x \leq 1\}$, $H = \{x \in \mathbb{R}^n | e_1^T x \leq 0\}$ a označme $B^- = B \cap H$ „zápornou" polovinu jednotkové koule se středem v počátku, to jest

$$B^- = \{x \in \mathbb{R}^n | x^T x \leq 1, x_1 \leq 0\}.$$

Hledáme elipsoid E' obsahující B^-, který bude „malý". Přesněji, chceme, aby $\mathrm{vol}(E') \leq q \cdot \mathrm{vol}(E)$ pro nějakou konstantu $q < 1$. Protože množina B^- je symetrická kolem osy x_1, budeme hledat E' ve tvaru

$$E' = \{x \in \mathbb{R}^n | (x - z)^T Z^{-1} (x - z) \leq 1\},$$

kde $z = t \cdot e_1$ pro $t \in \mathbb{R}^-$ a Z je diagonální matice. Vezměme $0 < p < 1$ a $d > 1$ a diagonální matici Z takovou, že

$$Z^{-1} = \begin{pmatrix} \frac{1}{p^2} & & & & \\ & \frac{1}{d^2} & & & \\ & & \frac{1}{d^2} & & \\ & & & \ddots & \\ & & & & \frac{1}{d^2} \end{pmatrix}.$$

Pak elipsoid E' bude ve směru první souřadnice smrštěný (ve srovnání s koulí o stejném počátku) a ve směru všech ostatních souřadnic roztažený. Chceme, aby se tento náš nový elipsoid E' toho původního $E = B$ dotýkal v bodech $-e_1$ a e_2 (viz obr. 2.1) a ze všech takových chceme ten s nejmenším objemem. Z první podmínky na dotek dostáváme následující omezení:

$$(-1-t)e_1^T \begin{pmatrix} \frac{1}{p^2} & & & & \\ & \frac{1}{d^2} & & & \\ & & \frac{1}{d^2} & & \\ & & & \ddots & \\ & & & & \frac{1}{d^2} \end{pmatrix} (-1-t)e_1 = 1$$

a tedy:

$$\frac{(1+t)^2}{p^2} = 1 \ .$$

Ze druhé podmínky na dotek dostáváme:

$$(-t, 1, 0, ..., 0) \begin{pmatrix} \frac{1}{p^2} & & & & \\ & \frac{1}{d^2} & & & \\ & & \frac{1}{d^2} & & \\ & & & \ddots & \\ & & & & \frac{1}{d^2} \end{pmatrix} \begin{pmatrix} -t \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 1 \ ,$$

tedy

$$\frac{t^2}{p^2} + \frac{1}{d^2} = 1 \ .$$

Z výše uvedených výrazů vyjádříme $1/p^2$ a $1/d^2$ jako funkce proměnné t:

$$\frac{1}{p^2} = \frac{1}{(1+t)^2}, \quad \frac{1}{d^2} = \frac{1+2t}{(1+t)^2} \qquad .$$

Nyní máme dvě podmínky na tři proměnné. Třetí podmínku získáme z požadavku na minimální objem elipsoidu E', pro který podle Tvrzení 7 platí $vol(E') = \sqrt{|\det(Z)|}vol(B)$, což je rovno

$$vol(E') = pd^{n-1}vol(B) = \frac{(1+t)^n}{\sqrt{(1+2t)^{n-1}}}vol(B) \ .$$

Snadno ověříme, že $vol(E')$ jako funkce proměnné t nabývá svého minima pro

$$t = \frac{-1}{n+1} \ .$$

Dosazením za p, d a t dostáváme, že hledaný elipsoid E' je tvaru

$$E' = \left\{ x \in \mathbb{R}^2 \mid (x + \frac{1}{n+1}e_1)^T \begin{pmatrix} \frac{(n+1)^2}{n^2} & 0 & \dots & 0 \\ 0 & \frac{n^2-1}{n^2} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \frac{n^2-1}{n^2} \end{pmatrix} (x + \frac{1}{n+1}e_1) \leq 1 \right\} \ .$$

Pro kontrolu ověříme, že elipsoid E' vskutku obsahuje celou „zápornou polovinu" koule B, tedy že $B^- \subseteq E'$. Víme, že pro $x \in B^-$ platí $x^T x \leq 1$, $x_1 \leq 0$. Rozepsáním podmínky v zápisu E' dostaneme

$$\left(\frac{n+1}{n}\right)^2 (x_1 + \frac{1}{1+n})^2 + \left(\frac{n^2-1}{n^2}\right) \sum_{i=2}^{n} x_i^2 \leq 1 \ .$$

Podle předpokladů $\sum_{i=1}^{n} x_i^2 \leq 1$ a proto také

$$\left(\frac{n^2-1}{n^2}\right) \sum_{i=2}^{n} x_i^2 \leq \frac{n^2-1}{n^2}(1 - x_1^2) \ .$$

Levá straně nápadně připomíná omezení v zápisu elipsoidu E'. Všimneme si dále, že

$$\left(\frac{n+1}{n}\right)^2 \left(x_1 + \frac{1}{n+1}\right)^2 = \left(\frac{n+1}{n}\right)^2 x_1^2 + 2\frac{n+1}{n^2} x_1 + \frac{1}{n^2} \ .$$

Poslední dvě nerovnosti (resp. nerovnost a rovnost) sečteme. S užitím $x_1^2 + x_1 \leq 0$ pro $-1 \leq x_1 \leq 0$ dostaneme kýžené $(x + \frac{1}{n+1})^T Z^{-1} (x + \frac{1}{n+1})^T \leq 1$.

Zbývá ukázat, že nový elipsoid bude mít objem výrazně menší, než elipsoid původní.

Lemma 13

$$\frac{vol(E')}{vol(B)} \leq e^{\frac{-1}{2(n+1)}}.$$

Důkaz. Připomeňme, že $vol(E') = \sqrt{\det Z} \cdot vol(B)$. Proto

$$\frac{\mathrm{vol}(E')}{\mathrm{vol}(B)} = \sqrt{\det Z} = \sqrt{\frac{n^2}{(n+1)^2}} \cdot \sqrt{\left(\frac{n^2}{n^2-1}\right)^{n-1}} = \frac{n}{n+1} \cdot \left(\frac{n^2}{n^2-1}\right)^{\frac{n-1}{2}} =$$

$$(1 - \frac{1}{n+1}) \cdot (1 + \frac{1}{n^2-1})^{\frac{n-1}{2}} \leq e^{-\frac{1}{n+1} + \frac{n-1}{2} \cdot \frac{1}{n^2-1}} = e^{-\frac{1}{n+1} + \frac{1}{2(n+1)}} = e^{\frac{-1}{2(n+1)}}.$$

Pro pořádek dodejme, že ve výpočtu jsme použili odhad $1 + x \leq e^x$. □

2.4.2 Obecný případ

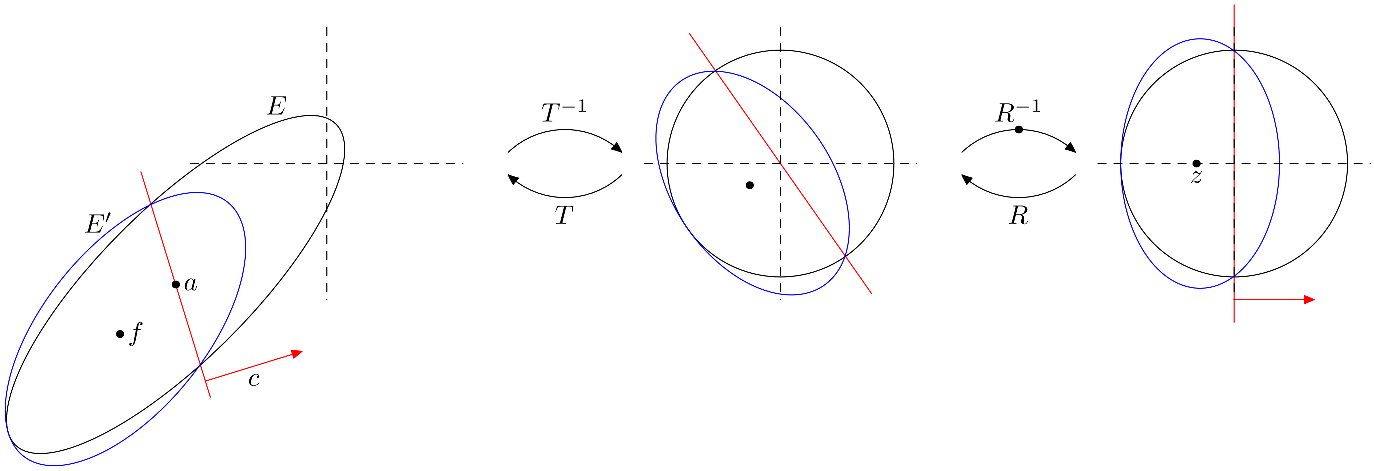
Připomeňme si situaci. Je dán elipsoid E se středem v a určený pozitivně definitní maticí A. Dále je dán poloprostor H s normálovým vektorem c, jehož hranice prochází středem a elipsoidu E; poloprostor H odpovídá porušené nerovnosti. Víme, že všechna přípustná řešení se vyskytují v poloprostoru H. Označme $Q = A^{1/2}$. Formálně tedy:

$$E = E(a, A) = \{x \in \mathbb{R}^n | (x-a)^T A^{-1} (x-a) \leq 1\} = Q \cdot E(0, I) + a,$$

$$H = \{x \in \mathbb{R}^n | c^T x \leq c^T a\} = \{x \in \mathbb{R}^n | c^T (x-a) \leq 0\}$$

Hledáme elipsoid $E' = E(f, C) \supseteq E \cap H$ s co nejmenším objemem.

Znovu využijeme skutečnosti, že každý elipsoid je afinním obrazem jednotkové koule. Pro daný elipsoid $E = E(a, A)$ označme T afinní zobrazení, které jednotkovou kouli se



Obrázek 2.2: Obrázek demonstruje, jaká zobrazení použijeme.

středem v počátku zobrazí na E; s pomocí T^{-1} zobrazíme naopak náš elipsoid E na jednotkovou kouli. Toto zobrazení T^{-1} aplikujeme také na poloprostor $H = \{x \mid c^T(x - a) \leq 0\}$. Všimněme si, že poloprostor $T^{-1}(H)$ typicky nebude totožný s poloprostorem $H_0 = \{x \in \mathbb{R}^n \mid x_1 \leq 0\}$, ale pro vhodnou rotaci R kolem počátku jistě $T^{-1}(H) = R(H_0)$. Abychom mohli použít výsledek z předešlé sekce, zobrazíme poloprostor $T^{-1}(H)$ pomocí R^{-1} na H_0 (viz obr. 2.2). Pak jistě

$$R^{-1}(T^{-1}(E)) \cap R^{-1}(T^{-1}(H)) \subseteq E(\frac{-1}{n+1}e_1, Z) \ , \tag{2.1}$$

a tudíž elipsoid $T(R(E(\frac{-1}{n+1}e_1, Z)))$ obsahje $E \cap H$. Chceme proto spočítat zobrazení T a R.

V dalších výpočtech využijeme o zobrazení T následující vlastnosti:

- $T(y) = Q \cdot y + a$
- $T^{-1}(x) = Q^{-1} \cdot (x - a)$
- $T^{-1}(H) = \{T^{-1}(x) | c^T(x - a) \leq 0\} = \{y \in \mathbb{R}^n | c^T(T(y) - a) \leq 0\} = \{y | c^T Q y \leq 0\} = \{y | (Qc)^T y \leq 0\}$.

Po rotaci R vyžadujeme jedinou vlastnost, totiž

$$R^{-1}\left(\frac{Qc}{||Qc||}\right) = e_1 \ . \tag{2.2}$$

Všimněme si, že v $\mathbb{R}^2$ je rotace R určena jednoznačně, ale v prostorech vyšších dimenzí už nikoli (což nám ovšem nevadí, jak za chvíli uvidíme). Nyní použijeme složení zobrazení T a R k popsání E' (pro jednoduchost požíváme R pro označení zobrazení i označení matice totoho zobrazení). Připomeňme si ještě z lineární algebry, že pro matici R rotace platí $RR^T = I$.

Věta 14 *Pro elipsoid $E' = E(f, C)$ se středem v f a určeném maticí C, kde*

$$\begin{aligned} f &= a - \frac{1}{n+1}\frac{Ac}{\sqrt{c^T Ac}} \ , \\ C &= \frac{n^2}{n^2-1} \cdot (A - \frac{2}{n+1}\frac{Acc^T A^T}{c^T Ac}) \ , \end{aligned}$$

platí:

$$E(a, A) \cap \{x | c^T(x-a) \leq 0\} \quad \subseteq \quad E(f, C) \ , \tag{2.3}$$

$$\frac{vol(E(f, C))}{vol(E(a, A))} \quad \leq \quad e^{\frac{-1}{2(n+1)}} \ . \tag{2.4}$$

Důkaz. Nejprve určíme střed f elipsoidu E', pomocí zobrazení R a T:

$$f = T(R(z)) = T\left(\frac{-1}{n+1} \cdot \frac{Qc}{||Qc||}\right) = a - \frac{1}{n+1}\frac{Ac}{\sqrt{c^T Ac}} \ .$$

Před určením C provedeme pomocný výpočet.

$$x - f = x - T(R(z)) = x - a - QRz = QR(R^{-1}Q^{-1}(x-a) - z).$$

Nyní opět využijime vlastnosti T a R, a pomocný výpočet:

$$\begin{aligned} E(f, C) &= \{T(Ry) | (y-z)^T Z^{-1}(y-z) \leq 1\} = \\ &= \{x | (R^{-1}T^{-1}(x) - z)^T Z^{-1}(R^{-1}T^{-1}(x) - z) \leq 1\} = \\ &= \{x | (R^{-1}Q^{-1}(x-a) - z)^T Z^{-1}(R^{-1}Q^{-1}(x-a) - z) \leq 1\} = \\ &= \{x | R^{-1}Q^{-1}(x-f))^T Z^{-1}(R^{-1}Q^{-1}(x-f)) \leq 1\} = \\ &= \{x | (x-f)^T {Q^{-1}}^T {R^{-1}}^T Z^{-1} R^{-1} Q^{-1}(x-f) \leq 1\} = \\ &= \{x | (x-f)^T C^{-1}(x-f) \leq 1\} \ , \end{aligned} \tag{2.5}$$

neboli

$$C = (QRZ^{1/2})(QRZ^{1/2})^T \ .$$

S využitím předešlého výpočtu a znalosti $Z = \text{diag}\left(\frac{n^2}{(n+1)^2}, \frac{n^2}{n^2-1}, \ldots\right)$ vyjádříme matici C pomocí matice A a vektoru a.

$$\begin{aligned} C &= Q \cdot R \cdot Z \cdot R^T \cdot Q^T = \\ &= \frac{n^2}{n^2-1} \cdot Q \cdot R \cdot \text{diag}\left(\frac{n-1}{n+1}, 1, \ldots 1\right) \cdot R^T \cdot Q^T = \\ &= \frac{n^2}{n^2-1} \cdot (QRIR^T Q^T - \frac{2}{n+1} \cdot QRe_1 e_1^T R^T Q^T) = \\ &= \frac{n^2}{n^2-1} \cdot (A - \frac{2}{n+1} \cdot \frac{Ac}{\sqrt{c^T Ac}} \cdot \frac{c^T A^T}{\sqrt{c^T Ac}}) = \\ &= \frac{n^2}{n^2-1} \cdot (A - \frac{2}{n+1}\frac{Acc^T A^T}{c^T Ac}) \ . \end{aligned}$$

Poznamenejme ještě, že C je jistě symetrická a pozitivně definitní matice.

Vlastnost (2.3) i (2.4) plyne ze vztahu (2.1), tedy je přímým důsledkem výpočtů provedených v předešlé kapitolce. □

2.5 Kdy skončit

Lemma 15 *Je-li $P = \{x \mid Cx \leq d\}$ polytop plné dimenze a C a d jsou celočíselné, pak*

$$vol(P) \geq 2^{-(n+1)\langle C\rangle + n^3} \ .$$

Důkaz. Pro jednoduchost zde dokážeme pouze slabší odhad; pro náš hlavní cíl – ukázat, že existuje polynomiální algoritmus pro hledání přípustného řešení – to nehraje žádnou roli.

Protože polytop P je plné dimenze, má $n+1$ afinně nezávislých vrcholů $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_n$ a platí:

$$\mathrm{vol}(\mathrm{conv}(\mathbf{v}_0, \ldots, \mathbf{v}_n)) \leq \mathrm{vol}(P) \ ,$$

kde $\mathrm{conv}(\cdots)$ označuje konvexní obal dané množiny vrcholů. Označme e_i vektor s jedinou nenulovou složkou, totiž i-tou, která je rovna jedné, pro $i = 1, \ldots, n$. Pak

$$\mathrm{conv}(\mathbf{v}_0, \ldots, \mathbf{v}_n) = T(\mathrm{conv}(\mathbf{0}, e_1, \ldots, e_n)) \ ,$$

kde T značí afinní zobrazení $T(x) = \mathbf{v}_0 + (\mathbf{v}_1 - \mathbf{v}_0, \ldots, \mathbf{v}_n - \mathbf{v}_0)\, x$.

V dalším opět využijeme toho, že každý vrchol mnohostěnu P je řešením vhodné soustavy rovnic odvozené z $Cx \leq d$: pro každé $\mathbf{v}_i$ existuje podsystém $C_i x \leq d_i$ systému $Cx \leq d$ takový, že $\mathbf{v}_i$ je jediným řešením soustavy $C_i x = d_i$. Podle Cramerova pravidla $v_{i,j} = \frac{\det C_{i,j}}{\det C_i}$, kde $C_{i,j}$ je matice vzniklá z C_i nahrazením j-tého sloupce vektorem d_i. S užitím Tvrzení 7, rovnosti $\mathrm{vol}(\mathrm{conv}(\mathbf{0}, e_1, \ldots, e_n) = 1/n!$ pro objem simplexu daného vektory $\mathbf{0}, e_1, \ldots, e_n$, a věty o rozvoji determinantu dostáváme, při značení $\mathbf{u}_i = \det C_i \cdot \mathbf{v}_i$, následující odhad:

$$\begin{aligned}
\mathrm{vol}(\mathrm{conv}(\mathbf{v}_0, \ldots, \mathbf{v}_n)) &= \frac{1}{n!}\left|\det\begin{pmatrix} 1 & 1 & \ldots & 1 \\ \mathbf{v}_0 & \mathbf{v}_1 & \ldots & \mathbf{v}_n \end{pmatrix}\right| = \\
&= \frac{1}{n!} \cdot \frac{1}{|\det C_0 \cdot \det C_1 \cdot \ldots \cdot \det C_n|} \cdot \left|\det\begin{pmatrix} \det C_0 & \det C_0 & \ldots & \det C_0 \\ \mathbf{u}_0 & \mathbf{u}_1 & \ldots & \mathbf{u}_n \end{pmatrix}\right| \geq \\
&\geq \frac{1}{n!} \cdot \left(2^{\frac{1}{\langle C\rangle + n\log n}}\right)^{(n+1)} \geq \\
&\geq 2^{-n\log n} \cdot 2^{-(n+1)\cdot\langle C\rangle - n(n+1)\log n} \ .
\end{aligned}$$

U předposlední nerovnosti jsme využili skutečnosti, že se jedná o determinant celočíselné matice, o které z předpokladů (afinní nezávislosti vrcholů $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_n$) víme, že je regulární, a tudíž její determinant je alespoň jedna. □

Věta 16 *Je-li P neprázdný plnodimenzionální polytop, pak algoritmus popsaný v této kapitole nejpozději do $k = 2 \cdot (n+1) \cdot (2(n+1)\langle C\rangle + n \cdot \langle d\rangle - n^3)$ iterací nalezne $x \in P$.*

Důkaz. Začneme jednoduchým odhadem $\mathrm{vol}(E_0) \leq (2r)^n$, kde $r = \sqrt{n} \cdot 2^{\langle C|d\rangle + n\log n}$.

Nyní předpokládejme, že algoritmus provedl k iterací. Pak s použitím uvedeného odhadu a Lemmatu 13 dostáváme $\mathrm{vol}(E_k) < \mathrm{vol}(P)$, což je spor s invariantem algoritmu, že celý mnohostěn P je vždy plně obsažen v elipsoidu E_k. □

Důsledek 17 *Pokud algoritmus do k iterací nenajde $x \in P$, pak je P prázdný.*

2.6 Jak se zbavit zjednodušujících předpokladů

2.6.1 Omezenost

Bez újmy na obecnosti můžeme předpokládat, že matice A v popisu mnohostěnu $P = \{\mathbf{x} \mid Ax \leq b\}$ je plné sloupcové hodnosti; v opačném případě lze snadno přejít do nižší dimenze. Připomeňme, že každá minimální stěna P lze popsat jako množina $\{\mathbf{x} \mid A'x = b'\}$, kde A' je matice téže hodnosti jako A, vzniklá z A výběrem vhodných řádků, a b' je vektor odpovídajících složek b. Má-li A plnou sloupcovou hodnost, je nutně každá minimální stěna mnohostěnu P vrcholem. Tudíž, je-li P neprázdný, obsahuje vrchol.

Dále už víme, že je-li $\mathbf{v}$ libovolný vrchol P, pak pro jeho i-tou souřadnici v_i platí $|v_i| \leq 2^{\langle A,b\rangle + n\log n}$. Pokud tedy P nějaké vrcholy má, jejich konvexní obal se jistě vejde do koule se středem v počátku a průměrem $R = \sqrt{n} \cdot 2^{\langle C,d\rangle + n\log n}$. Algoritmus bude po celou dobu svého běhu zachovávat invariant, že konvexní obal vrcholů P leží uvnitř aktuálního elipsoidu, což zaručuje správné chování algoritmu i pro tento případ.

2.6.2 Plná dimenze

Lemma 18 *Nechť A je celočíselná matice a b celočíselný vektor. Pak systém $Ax \leq b$ má řešení právě když systém $Ax \leq b + \mathbb{1} \cdot \varepsilon$ má řešení pro $\varepsilon = \frac{1}{2m2^{\langle A,b\rangle - n^2}}$.*

Důkaz. Implikace zleva doprava je triviální. Pro opačnou použijeme Farkasovo lemma a ukážame, že y z Farkasova lemmatu pro původní úlohu funguje i pro odvozenou. Podrobnosti za domácí úkol. □

2.6.3 Hledání optimálního řešení

Lemma 19 *Předpokládejme, že úloha „rozhodni, zda mnohostěn P je prázdný či neprázný” je řešitelná v polynomiálním čase. Pak i úloha „nalezni $x \in P$, nebo poznej, že P je prázdné”, je řešitelná v polynomiálním čase.*

Důkaz. Uvědomme si, že řešení úlohy s uvolněnou pravou stranou (viz Lemma 18) nemusí být řešením původní úlohy (což je ta, která nás opravdu zajímá). Předeslé lemma nás ale ujišťuje, že má-li řešení uvolněná soustava, má řešení (ne nutně stejné) i úloha původní. Jde o to, jak tuto znalost využít k tomu, abychom řešení skutečně nalezli.

Idea je následující: snižovat počet nerovností v původní soustavě tak dlouho, až nám zbydou pouze rovnosti. Jak to provést? Nabízí se dva způsoby: ubírání nadbytečných nerovností a přeměna ostatních na rovnice. Systém tvořený pouze rovnicemi vyřešíme např. Gaussovou eliminací.

Popišme si postup trochu přesněji: Předpokládejme, že naše soustava $Ax \leq b$ obsahuje k nerovností (a jedno kolik rovností – dvě opačné nerovnosti počítáme jako jednu rovnost). Vezmeme libovolnou nerovnost, nahradíme ji rovností a otestujeme, zda nová soustava má řešení (pomocí předešlého lemmatu). Má-li nová soustava řešení, ubyla nám jedna nerovnost a soustava má i nadále nějaké řešení. Nemá-li nová soustava řešení, pak vybraná nerovnost neodpovídá žádné stěně mnohostěnu a můžeme ji proto vynechat jako nadbytečnou. □

Lemma 20 *Je-li úloha „nalezni $x \in P$, nebo poznej, že P je prázdné" řešitelná v polynomiálním čase, pak je i optimalizační verze úlohy řešitelná v polynomiálním čase.*

Důkaz. Úlohu $\max\{cx : Ax \leq b\}$ vyřešíme pomocí algoritmu pro úlohy $\exists x : Cx \leq d$.

Ze silné duality víme, že $\max\{cx : Ax \leq b\} = \min\{yb : yA = c, y \geq 0\}$, jsou-li obě množiny neprázdné. Postupovat tedy budeme takto:

1. Otestuj zda $\{x \in \mathbb{R}^n : Ax \leq b\} = \emptyset$. Pokud ano, pak řešení neexistuje.

2. Uvaž $P' = \{(x, y) \in \mathbb{R}^{n+m} : Ax \leq b, y \geq 0, yA \leq c, -yA \leq -c, cx \leq yb, -cx \leq -yb\}$. Pokud $P' = \emptyset$, pak P je neomezený a maximum cílové funkce je libovolně velké. Pokud P' není prázdný, pak existuje $x_0, y_0 \in P'$, přičemž snadno vidíme, že x_0 je přípustné řešení primární úlohy, y_0 je přípustné řešení duální úlohy a silná dualita nás ujišťuje, že obě jsou optimální.

□

2.6.4 Nedohodnocená těžkost

Iracionalita (odmocniny) ve výpočtech …

2.7 Shrnutí

Odhlédneme-li od problémů s odmocninami, dokázali jsme následující větu.

Věta 21 (Chačijan, 1979) *Existuje polynomiální algoritmus pro řešení úlohy lineárního programování.*

Kapitola 3

Metoda vnitřního bodu

3.1 Náčrt algoritmu

V tomto textu budeme uvažovat úlohu lineárního programování v rovnicovém (též nazývaném standardní) tvaru:

$$
\begin{aligned}
\min c^T x & & \qquad (3.1)\\
Ax &= b\\
x &\geq 0,
\end{aligned}
$$

kde A je reálná matice $m \times n$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ a $x \in \mathbb{R}^n$. Zároveň budeme pracovat s duální úlohou ve tvaru

$$
\begin{aligned}
\max y^T b & & \qquad (3.2)\\
A^T y + s &= c\\
s &\geq 0,
\end{aligned}
$$

kde $y \in \mathbb{R}^m$ a $s \in \mathbb{R}^n$; proměnné s_i nazýváme *přídavné.* Bez újmy na obecenosti budeme předpokládat, že matice A má plnou řádkovou hodnost. Všimněme si, že pak řešení (y, s) duální úlohy je možno jednoznačně reprezentovat pomocí vektoru s přídavných proměnných (z něhož lze y dopočítat, např. Gaussovou eliminací). Řekneme, že x je *strikně* přípustné řešení, pokud $x > 0$ a x je přípustné řešení; obdobně pro s. V tomto textu se budeme věnovat problému nalezení optimálního řešení primární úlohy.

Jádrem metody je postup, který z libovolného strikně přípustného řešení primární úlohy spočítá v polynomiálním čase (skoro) optimální řešení. To se na první pohled nezdá příliš uspokojivé: víme, že už samotné nalezení nějakého přípustného řešení úlohy LP je „stejně” obtížné jako nalezení optimálního řešení této úlohy. Trik umožňující začínat rovnou se strikně přípustným řešením, aniž bychom řešili nějaký LP, spočívá v drobné úpravě úlohy LP (přidání dvou proměnných a drobná úprava soustavy); v upravené úloze je možné striktně přípustné řešení získat zadarmo, přičemž optimální řešení upravené úlohy přímo dává optimální řešení úlohy původní.

Podobně jako elipsoidová metoda je i metoda vnitřního bodu *iterační* metoda. U elipsoidové metody jsme konstruovali posloupnost bodů (středy elipsoidů), která začínala typicky někde vně zadaného mnohostěnu P a pokud byl P neprázdný, končila někde uvnitř P. V metodě vnitřního bodu budeme také konstruovat posloupnost bodů, která ovšem začíná někde uvnitř zadaného mnohostěnu P, v mnohostěnu P celou dobu zůstává a končí v (téměř) optimálním řešení; jedná se tedy o posloupnost *přípustných* řešení primární úlohy. V podobě popsané v tomto textu jde o primárně-duální algoritmus, tedy o algoritmus, který pracuje zároveň s řešeními primární i duální úlohy; společně s posloupností řešení primární úlohy budeme konstruovat ještě druhou posloupnost bodů reprezentujících řešení duální úlohy.

Klíčovým nástrojem metody je *potenciál* a zde popsaná varianta algoritmu patří mezi *metody snižování potenciálu.* Pro dvojici (x, s) strikně přípustých řešení primární a duální úlohy definujeme potenciál jako číslo

$$G(x,s) \quad = \quad (n+\sqrt{n})\ln(x^T s) - \sum_{i=1}^{n} \ln(x_i s_i)\ . \tag{3.3}$$

Potenciál nám poslouží dvojím způsobem. Jednak bude užitečný pro kontrolu doby běhu algoritmu (např. podle něj poznáme, kdy už můžeme skončit), a jednak bude řídit volbu dalších bodů v obou konstruovaných posloupnostech. Obdobnou roli hrály v elipsoidové metodě elipsoidy: jejich objem nám sloužil jako počítadlo, hlídající dobu běhu algoritmu, a elipsoidy samotné se používaly ke konstrukci dalších bodů posloupnosti.

Záhy nahlédneme, že pro přibližování se k optimální dvojici řešení potřebujeme potenciál snižovat. Nechť $\bar{x}$ a $\bar{s}$ jsou aktuální body konstruovaných posloupností. Základní myšlenkou metody je spočítat gradient g potenciálu $G(x, s)$ vzhledem k vektoru proměnných x v místě $(\bar{x}, \bar{s})$ a vydat se v posloupnosti řešení primární úlohy směrem $-g$, neboli za další bod posloupnosti vzít $\bar{x} - \alpha g$ pro nějaké vhodné $\alpha > 0$. Takto jednodušše to ovšem nepůjde, protože bychom většinou porušili omezující podmínky $Ax = b$ primární úlohy. Proto nepůjdeme přímo ve směru $-g$, ale ve směru daném projekcí vektoru $-g$ do nadroviny $Ax = 0$. Tím se vyhneme problému s omezujícími podmínkami, ale může se nám stát, že tato projekce bude příliš malá, v důsledku čehož by kroků potřebných k dosažení optima bylo potřeba nepřijatelně mnoho (pokud bychom ho vůbec kdy dosáhli). Z tohoto důvodu provedeme v případě, že projekce g do nadroviny $Ax = 0$ je malá, tzv. *duální krok* a místo změny řešení primární úlohy změníme řešení duální úlohy. Nápovědu ke změně duálního řešení nám opět poskytne potenciál, nebo přesněji gradient potenciálu $G(x, s)$ vzhledem k s v místě $(\bar{x}, \bar{s})$. Ukážeme, že v případě primárního i duálního kroku dojde k výraznému snížení potenciálu, což bude stačit k polynomiálnímu odhadu na čas běhu algoritmu.

Na konci tohoto hrubého náčrtu metody zmiňme ještě jeden klíčový nástroj metody, totiž *afinní transformaci.* Metoda potřebuje, aby všechna řešení, která postupně konstruuje, byla striktně přípustná. To působí jistá omezení v možnostech, kterým směrem a jak daleko se vydat při výpočtu dalšího bodu posloupnosti řešení. Abychom se těmto omezením vyhnuli, použijeme na začátku každé iterace afinní transformaci, která poslední bod primární posloupnosti zobrazí na bod $x' = (1, 1, \ldots, 1)^T$. V tomto transformovaném prostoru spočítáme iterační krok způsobem načrtnutým výše a inverzní transformací se vrátíme do původního prostoru. Uvidíme, že pokles potenciálu v původním prostoru bude stejně velký jako pokles potenciálu v transformovaném prostoru. Použitá afinní transformace nám také podstatným způsobem zjednoduší práci s gradientem g a jeho projekcí do $Ax = 0$.

3.2 Základní definice a pozorování

Lemma 22 *Nechť x je přípustné řešení primární úlohy a s je přípustné řešení duální úlohy. Pak platí*

$$c^T x - y^T b = s^T x \ . \tag{3.4}$$

Důkaz. $c^T x - y^T b = y^T Ax + s^T x - y^T Ax = s^T x$. □

Připomeňme, že *potenciál* dvojice (x, s) primárního a duálního řešení definujeme jako

$$G(x, s) = (n + \sqrt{n}) \ln(x^T s) - \sum_{i=1}^{n} \ln(x_i s_i) \ . \tag{3.5}$$

Všimněme si, že pro dvojici (x, s) blízkou optimální dvojici řešení se podle předešlého lemmatu první člen pravé strany rychle blíží k $-\infty$; abychom se přiblížili k optimálnímu řešení, bude proto třeba potenciál minimalizovat. Pokud se snažíme o minimalizaci potenciálu, pak suma $\sum_{i=1}^{n} \ln(x_i s_i)$ v jeho definici slouží jako bariéra kolem hranic $x_i \geq 0$ a $s_i \geq 0$: pokud se některá složka x nebo s blíží k nule, způsobí odečtení této sumy výrazný nárůst potenciálu. Pro pořádek upozorňujeme, že pro potenciál se v jiných variantách metody vnitřního bodu používají i jiné funkce (např. $(n + \sqrt{n}) \ln(x^T s) - \sum_{i=1}^{n} \ln x_i$). V následujícím textu budeme pro zkrácení zápisu používat značení $q = n + \sqrt{n}$.

Pro velikost zápisu úlohy LP používáme označení L. Následující věta a její důsledek budou důležité při rozhodování o ukončení algoritmu.

Věta 23 *Nechť A je celočíselná matice, b, c jsou celočíselné vektory a nechť u a v jsou vrcholy polyedru $P = \{x | Ax = b, x \geq 0\}$. Pak, za předpokladu $c^T u \neq c^T v$, platí*

$$|c^T u - c^T v| > 2^{-2L} \ . \tag{3.6}$$

Důkaz. Z teorie mnohostěnů víme, že je-li u vrcholem polyedru P, pak u je jediným řešením vhodného podsystému rovnic systému $Ax = b$, $x \geq 0$. Tudíž, podle Cramerova pravidla, i-tá souřadnice vrcholu u se dá vyjádřit jako $u_i = \frac{\det \hat{A}_{i \to b}}{\det \hat{A}}$, kde $\hat{A}$ je matice onoho vhodného podsystému (tedy $\hat{A}$ je vhodná podmatice matice $\binom{A}{I}$) a $\hat{A}_{i \to b}$ je matice vzniklá z $\hat{A}$ nahrazením jejího i-tého sloupce odpovídajícími složkami pravé strany původního LP. Z podoby $\hat{A}$ plyne $|\det \hat{A}| \leq 2^{L-n^2} < 2^L$ (viz Pozorování 11 v části o elipsoidové metodě). Obdobná pozorování můžeme udělat i pro vrchol v.

Víme proto, že existují $q_1, q_2 \in \mathbb{N}$ taková, že $q_1, q_2 < 2^L$ a zároveň $u^T q_1$ a $v^T q_2$ jsou celočíselné vektory. Tudíž

$$|c^T u - c^T v| = \left| \frac{q_1 c^T u}{q_1} - \frac{q_2 c^T v}{q_2} \right| = \left| \frac{q_1 q_2 (c^T u - c^T v)}{q_1 q_2} \right| > \frac{1}{2^{2L}} \ ,$$

kde v poslední nerovnosti využíváme skutečnosti, že $q_1 q_2 (c^T u - c^T v)$ je díky volbě q_1 a q_2 celé číslo. □

Důsledek 24 *Je-li $x^T s \leq 2^{-2L}$, pro dvojici řešení x a (y, s) primární a duální úlohy, pak jakýkoli vrchol x' takový, že $c^T x' \leq c^T x$, je optimální.*

Důkaz. Sporem. Předpokládejme, že existuje vrchol x' takový, že $c^T x' \leq c^T x$, ale x' není optimální. Pak podle předešlé věty platí $c^T x^\star < c^T x' - 2^{-2L}$, kde $x^\star$ je nějaký optimalní vrchol. S použitím předpokladů $x^T s \leq 2^{-2L}$ a $c^T x' \leq c^T x$, rovnosti $c^T x = y^T b + x^T s$ a nerovnosti $y^T b \leq c^T x^\star$ plynoucí z optimality $x^\star$, dostáváme kýžený spor $c^T x^\star < c^T x' - 2^{-2L} \leq c^T x - x^T s = y^T b \leq c^T x^\star$. □

Ve zbytku této kapitoly uvádíme tři pomocná technická tvrzení. Jejich důkazy přenecháváme čtenáři jako cvičení z matematické analýzy a lineární algebry.

Lemma 25 (Geometrický a aritmetický průměr) *Pro libovolná $t_j \geq 0$, $j = 1, \dots, n$, platí:*

$$\left(\prod_{j=1}^{n} t_j\right)^{1/n} \leq \frac{1}{n}\sum_{j=1}^{n} t_j. \tag{3.7}$$

Lemma 26 (Odhady logaritmů) *Pro $|x| \leq a < 1$ platí*

$$-x - \frac{x^2}{2(1-a)} \leq \ln(1-x) \leq -x\ . \tag{3.8}$$

Lemma 27 *Je-li A matice plné řádkové hodnosti, pak je matice AA^T regulární.*

3.3 Struktura algoritmu

Na začátku najdeme dvojici striktně přípustných řešení (x_0, s_0) takovou, že $G(x_0, s_0) = O(\sqrt{n}L)$. V každé iteraci pak zajistíme pokles potenciálu o alespoň 1/120 a skončíme ve chvíli, kdy potenciál poprvé klesne pod $-2\sqrt{n}L$. Tím bude zaručen polynomiální počet iterací. Protože čas potřebný pro každou iteraci bude také omezen polynomem ve velikosti zadání, bude i celkový čas algoritmu polynomiální.

Rozhodnutí ukončit algoritmus při poklesu potenciálu pod $-2\sqrt{n}L$ ospravedlňuje následující lemma.

Lemma 28 *Je-li $G(x,s) \leq -2\sqrt{n}L$, pak $x^T s < e^{-2L}$.*

Důkaz. Použijeme Lemma 3.7 pro $t_j = x_j s_j$, $j = 1, \dots, n$. Po zlogaritmování obou stran nerovnosti (3.7) a drobné úpravě dostaneme nerovnost

$$n \ln x^T s - \sum_{j=1}^{n} \ln x_j s_j \geq n \ln n\ . \tag{3.9}$$

Podle předpokladů lemmatu, definice potenciálu (3.5) a nerovnosti (3.9) platí

$$\begin{aligned} -2\sqrt{n}L \quad &\geq \quad G(x,s) = (n+\sqrt{n})\ln(x^T s) - \sum_{i=1}^{n} \ln(x_i s_i) \geq \\ &\sqrt{n}\ln x^T s + n\ln n \geq \sqrt{n}\ln x^T s\ . \end{aligned}$$

Tudíž $-2L \geq \ln x^T s$ a důkaz je hotov. □

Důsledek 24 spolu s Lemmatem 28 zaručují, že při poklesu potenciálu pod $-2\sqrt{n}L$ už jsme takřka v optimálním řešení.

3.4 Jak nalézt strikně přípustná řešení x_0 a s_0 s omezeným potenciálem

Chceme původní dvojici lineárních programů nahradit novou dvojicí tak, aby i) pro nové LP bylo snadné nalézt přípustné řešení s malým potenciálem, a aby ii) z optimálního řešení nového LP bylo možné jednodušše získat optimální řešení původního LP. Poslouží nám níže uvedené lineární programy.

$$
\begin{array}{lrcrcrcl}
\min & c^T x & + & 2^{6L} x_{n+1} & & & & \\
 & Ax & + & (b - 2^{2L}Ae)x_{n+1} & & & = & b \\
 & (2^{4L}e - c)^T x & & & + & 2^{4L}x_{n+2} & = & k \\
 & x & & & & & \geq & 0 \\
 & & & x_{n+1} & & & \geq & 0 \\
 & & & & & x_{n+2} & \geq & 0
\end{array}
$$

kde $k = 2^{6L}(n+1) - 2^{2L}c^T e$.

$$
\begin{array}{lrcrcccccl}
\max & b^T y & + & k y_{m+1} & & & & & & \\
 & A^T y & + & (2^{4L}e - c)y_{m+1} & + & s & & & = & c \\
 & (b - 2^{2L}Ae)^T y & & & & & + & s_{n+1} & = & 2^{6L} \\
 & & & 2^{4L}y_{m+1} & & & & + \; s_{n+2} & = & 0 \\
 & & & & & s & & & \geq & 0 \\
 & & & & & & & s_{n+1} & \geq & 0 \\
 & & & & & & & s_{n+2} & \geq & 0
\end{array}
$$

Lemma 29 *Platí:*

1. *Vektory* $x' = (2^{2L}, 2^{2L}, \ldots, 2^{2L}, 1, 2^{2L})$ *a* $(y'; s') = (0, -1; 2^{4L}, 2^{4L}, \ldots, 2^{4L}, 2^{6L}, 2^{4L})$ *jsou striktně přípustná řešení a navíc* $G(x', s') = O(\sqrt{n}L)$.

2. *Velikost nového LP je* $O(L)$.

3. *Jsou-li vektory* $x^\star$ *a* $y^\star$ *optimálním řešením LP (3.1) a (3.2), pak vektory* $x'' = (x^\star, 0, (k - (2^{4L}e - c)^T x^\star)/2^{4L})$ *a* $(y'', s'') = (y^\star, 0, s^\star, 2^{6L} - (b - 2^{2L}Ae)^T y^\star, 0)$ *jsou optimálním řešením nových LP.*

4. *Mají-li oba LP (3.1) a (3.2) přípustná řešení a jsou-li vektory* x'' *and* (y'', s'') *optimálním řešením nových LP, pak vektory* x *a* (y, s)*, kde* x *a* s *jsou restrikcí* x'' *a* s'' *na prvních* n *složek a* y *je restrikcí* y'' *na prvních* m *složek, jsou optimálním řešením LP (3.1) a (3.2).*

Důkaz. Ověření prvního, druhého a třetího tvrzení lemmatu je pouze technickou záležitostí. Poslední část vyžaduje o trochu více práce a její důkaz přeskočíme; pilný čtenář se o důkaz může s pomocí vět (podmínek) o komplementaritě primárního a duálního řešení pokusit sám. □

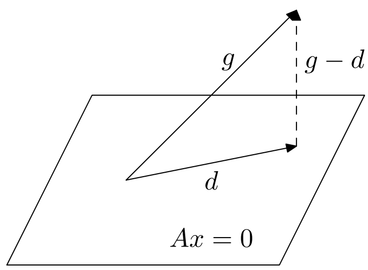
3.5 Iterační krok

3.5.1 Snadný případ

V této podkapitole předpokládáme, že aktuální řešení primární úlohy je $e = (1, \ldots, 1)^T$ a aktuální řešení duální úlohy je $s' > 0$. Jak jsme již popsali v úvodu, základní myšlenkou je změnit řešení primární úlohy ve směru opačném ke gradientu potenciálu $G(x, s)$ podle x v bodě (e, s'); začneme proto výpočtem tohoto gradientu g.

$$g = \nabla_x G(x,s)|_{(e,s')} = \left[(n + \sqrt{n}) \frac{s}{x^T s} - \begin{pmatrix} \frac{1}{x_1} \\ \vdots \\ \frac{1}{x_n} \end{pmatrix} \right] \Bigg|_{(e,s')} = (n + \sqrt{n}) \frac{s'}{e^T s'} - e \ . \quad (3.10)$$

Přímo ve směru $-g$ ovšem jít nemůžeme, protože bychom mohli porušit podmínky $Ax = 0$. Z toho důvodu spočítáme projekci d vektoru g do nadroviny $\{x \mid Ax = 0\}$, a pokud bude dostatečně velká, přejdeme od aktuálního řešení e primární úlohy k řešení $e - \alpha d$, pro vhodné $\alpha > 0$, jehož velikost upřesníme později.



Obrázek 3.1: Projekce vektoru g do nadroviny $\{x \mid Ax = 0\}$.

Lemma 30 (Projekce g do nadroviny $\{x \mid Ax = 0\}$)

$$d = (I - A(AA^T)^{-1}A)g \ . \quad (3.11)$$

Důkaz. Nejprve si všimneme, že vektor $g - d$ je z ortogonálního doplňku vektorového prostoru $\{x \mid Ax = 0\}$ a tudíž je lineární kombinací řádků matice A. Jinými slovy, existuje vektor $w \in \mathbb{R}^m$ takový, že

$$g - d = A^T w \ . \quad (3.12)$$

Podáří-li se nám tento vektor w vyjádřit, budeme mít i předpis pro hledaný vektor d. Vynásobením rovnosti (3.12) zleva maticí A dostaneme (s použitím rovnosti $Ad = 0$ vyjadřující, že d je projekce do požadované nadroviny)

$$Ag = AA^T w \ .$$

Díky plné řádkové hodnosti matice A je matice AA^T regulární, proto platí

$$w = (AA^T)^{-1} Ag \ .$$

Dosazením za w do rovnosti (3.12) dostaneme kýžený vztah (3.11). □

Primární krok Pokud bude $\|d\| \geq 0,22$ provedeme primární krok a další dvojici přípustných řešení určíme podle předpisu

$$\tilde{x} = e - \frac{d}{4\|d\|}, \quad \tilde{s} = s' \; . \tag{3.13}$$

V opačném případě provedeme krok duální (vysvětlen dále). Pro úspěšnost primárního kroku je třeba ověřit dvě věci:

- opět jsme získali striktně přípustné řešení primární úlohy,
- potenciál podstatně klesl.

Ověření těchto podmínek je předmětem následujících dvou lemmat.

Lemma 31 (Striktní přípustnost $\tilde{x}$) *Vektor $\tilde{x}$ je strikně přípustné řešení primární úlohy.*

Důkaz. Protože každá složka vektoru $\frac{d}{\|d\|}$ je nejvýše 1, máme zaručeno $\tilde{x} > 0$. Protože $Ad = 0$, máme zaručenou i přípustnost $\tilde{x}$. □

Lemma 32 (Pokles potenciálu v primárním kroku) *Po provedení primárního kroku platí*

$$G(\tilde{x}, \tilde{s}) - G(e, s') \leq \frac{-1}{120} \; . \tag{3.14}$$

Důkaz. Pro zkrácení zápisu použijeme v důkazu značení $q = (n + \sqrt{n})$. Po dosazení za $\tilde{x}$ a $\tilde{s}$ podle vzorce (3.13) dostaneme (indexy nad některými relacemi používáme k pozdějšímu odvolávání se na ně)

$$\begin{aligned}
G(\tilde{x}, \tilde{s}) - G(e, s') &= q \ln(e^T s' - \frac{d^T s'}{4\|d\|}) - \sum_{j=1}^{n} \ln(1 - \frac{d_j}{4\|d\|}) - \sum_{j=1}^{n} \ln(s'_j) - q\ln(e^T s') + \sum_{j=1}^{n} \ln(s'_j) = \\
&= q \ln(1 - \frac{d^T s'}{4\|d\| e^T s'}) - \sum_{j=1}^{n} \ln(1 - \frac{d_j}{4\|d\|}) \leq \\
&\overset{i}{\leq} -q \frac{d^T s'}{4\|d\| e^T s'} + \sum_{j=1}^{n} \frac{d_j}{4\|d\|} + \sum_{j=1}^{n} \frac{d_j^2}{16\|d\|^2 2(1 - \frac{1}{4})} = \\
&= -q \frac{d^T s'}{4\|d\| e^T s'} + \frac{d^T e}{4\|d\|} + \frac{1}{24} = \\
&= \frac{d^T}{4\|d\|} \left(e - q \frac{s'}{e^T s'} \right) + \frac{1}{24} = \\
&\overset{ii}{=} -\frac{d^T g}{4\|d\|} + \frac{1}{24} = \\
&\overset{iii}{=} -\frac{\|d\|^2}{4\|d\|} + \frac{1}{24} = -\frac{\|d\|}{4} + \frac{1}{24} \leq -\frac{0{,}2}{4} + \frac{1}{24} = -\frac{1}{120}
\end{aligned}$$

Nerovnost (i) vychází z odhadu logaritmu pomocí Lemmatu (26) s parametrem $a = \frac{1}{4}$ (jistě máme $\frac{|d_j|}{4\|d\|} \leq \frac{1}{4}$). Rovnost (ii) používá výpočet gradientu g (3.10). Rovnost (iii) využívá vztah $g^T d = d^T d$ plynoucí z Pythagorovy věty ($g^T g = d^T d + (g - d)^T (g - d)$). Poslední nerovnost využívá dolní mez na $\|d\|$ pro primární krok. □

Duální krok Duální krok provedeme, pokud $\|d\| < 0{,}22$. Základní myšlenka duálního kroku je stejná jako v kroku primárním: spočítat gradient h potenciálu $G(x,s)$ podle s v bodě (e,s') a posunout řešení duální úlohy ve směru určeném vektorem $-h$. Podobně jako v kroku primárním bude ale třeba ohlídat, abychom neporušili žádnou omezující podmínku.

Začněme výpočtem gradientu h. Protože role proměnných x a s ve funkci $G(x,s)$ jsou symetrické, s přihlédnutím ke vztahu (3.10) rovnou nahlédneme, že pro gradient h potenciálu $G(x,s)$ podle s v bodě (e,s') platí

$$h = \nabla_s G(x,s)|_{(e,s')} = (n+\sqrt{n})\frac{e}{e^T s'} - \begin{pmatrix} \frac{1}{s'_1} \\ \vdots \\ \frac{1}{s'_n} \end{pmatrix} . \tag{3.15}$$

Porovnejme nyní jednotlivé složky vektorů g a h. Platí jednoduchý vztah, pro $j = 1,\dots,n$,

$$h_j = \frac{g_j}{s'_j} .$$

Protože všechny složky vektoru s' jsou kladné, ukazují vektory g a h podobným směrem, presněji řečeno, do stejného ortantu. Vzhledem k této podobnosti g a h, a vhledem k tomu, že o vektoru g už lecos víme, budeme v duálním kroku pracovat s vektorem g a nikoli s vektorem h.

Označme si y' druhou část složek přípustného řešení duální úlohy odpovídající vektoru s'; protože se jedná o přípustné řešení, vektory y' a s' splňují

$$A^T y' + s' = c . \tag{3.16}$$

Naším dílčím cílem je změnit aktuální duální přípustné řešení s' na nové duální přípustné řešení $\tilde{s}$ (a při tom zajistit pokles potenciálu). Zajistíme-li při volbě $\tilde{s}$, že existuje vektor $\tilde{y}$ splňující rovnici

$$A^T \tilde{y} + \tilde{s} = c , \tag{3.17}$$

máme zaručenu přípustnost $\tilde{s}$. Ze vztahů (3.16) a (3.17) plyne pro $\tilde{s}$ omezení

$$\tilde{s} - s' = A^T(y' - \tilde{y}) . \tag{3.18}$$

Vyjádřeno slovy, $\tilde{s} - s'$ musí být lineární kombinací řádků matice A, neboli $\tilde{s} - s'$ musí být kolmé na nadrovinu $\{x \mid Ax = 0\}$. S jedním vektorem kolmým na nadrovinu $\{x \mid Ax = 0\}$ už jsme pracovali, totiž s vektorem $g - d$. Proto se při rozhodování, jak zlepšovat duální řešení s', nabízí vydat se směrem $-(g-d)$. Tuto myšlenku provedeme a v duálním kroku určíme další dvojici přípustných řešení podle předpisu

$$\tilde{s} = s' - \frac{e^T s'}{n+\sqrt{n}}(g-d) , \quad \tilde{x} = e . \tag{3.19}$$

Podobně jako v primárním kroku je třeba ověřit dvě věci:

- získali jsme striktně přípustné řešení duální úlohy,

- potenciál podstatně klesl.

Ověření těchto podmínek je předmětem následujících dvou lemmat.

Lemma 33 (Striktní přípustnost $\tilde{s}$) *Po provedení duálního kroku je $\tilde{s}$ striktně přípustné řešení duální úlohy.*

Důkaz. Existence vektoru $\tilde{y}$ splňujícímu rovnost (3.17) plyne z definice $\tilde{s}$ (volíme $\tilde{s}$, aby $\tilde{s} - s'$ bylo lineární kombinací řádků A^T), proto je $\tilde{s}$ přípustné řešení. Ještě ověříme jeho *striktní* přípustnost.

Podle definice $\tilde{s}$ a definice g, pro vektor $\tilde{s}$ platí

$$\tilde{s} = s' - \frac{e^T s'}{n+\sqrt{n}}(g-d) = s' - \frac{e^T s'}{n+\sqrt{n}}(\frac{n+\sqrt{n}}{e^T s'}s' - e - d) = \frac{e^T s'}{n+\sqrt{n}}(d+e)\ . \quad (3.20)$$

Protože duální krok provádíme v případě $\|d\| < 0{,}22$, je z posledního vyjádření vidět $\tilde{s} > 0$. Tím je důkaz hotov. □

Dříve než odhadneme pokles potenciálu při provedení duálního kroku, dokážeme pomocné lemma.

Lemma 34

$$\sum_{j=1}^{n} \ln \tilde{s}_j \ \ \geq \ \ n \ln \frac{e^T \tilde{s}}{n} - \frac{1}{32}\ . \quad (3.21)$$

Důkaz.

$$\sum_{j=1}^{n} \ln \tilde{s}_j - n \ln \frac{e^T \tilde{s}}{n} = \sum_{j=1}^{n} \ln(\frac{e^T s'}{q}(1+d_j)) - n\ln(\frac{e^T s'}{q}(1+\frac{e^T d}{n})) =$$
$$= \sum_{j=1}^{n} \ln(1+d_j) - n\ln(1+\frac{e^T d}{n}) \geq \sum_{j=1}^{n}(d_j - \frac{d_j^2}{2(1-0{,}22)}) - n\frac{e^T d}{n} =$$
$$= -\frac{\|d\|^2}{2(1-0{,}22)} \geq -\frac{1}{32}\ .$$

První rovnost plyne ze vztahu (3.20) odvozeného v důkazu předešlého lemmatu. Při dalších úpravách jsou použity základní vlastnosti funkce logaritmus, její odhady z Lemmatu 26 a horní odhad na velikost $\|d\|$ v duálním kroku. □

Lemma 35 (Pokles potenciálu v duálním kroku) *Po provedení duálního kroku platí*

$$G(\tilde{x}, \tilde{s}) - G(e, s') \ \ \leq \ \ -\frac{1}{3}\ . \quad (3.22)$$

Důkaz. Začneme pomocným výpočtem, kde v první rovnosti používáme identitu (3.20) a pro první nerovnost vztah $(\sum_{i=1}^{n} d_i)^2 \ \leq \ n\sum_{i=1}^{n} d_i^2$ (dokažte si s použitím Cauchy-Schwarzovy nerovnosti):

$$e^T \tilde{s} = \frac{e^T s'}{q}(e^T d + n) \leq \frac{e^T s'}{q}(\sqrt{n}\,\|d\| + n) \leq \frac{e^T s'}{q}(n + 0{,}22\sqrt{n})\ .$$

Dostáváme

$$\frac{e^T\tilde{s}}{e^Ts'} \leq \frac{n+0{,}22\sqrt{n}}{n+\sqrt{n}} = 1 - \frac{\sqrt{n}-0{,}22\sqrt{n}}{q} \ . \tag{3.23}$$

Pokles potenciálu pak odhadneme takto:

$$\begin{aligned}
G(e,\tilde{s}) - G(e,s') &= q\ln(e^T\tilde{s}) - \sum_{j=1}^{n} \ln \tilde{s}_j - q\ln(e^Ts') + \sum_{j=1}^{n} \ln {s'}_j = \\
&= q\ln\frac{e^T\tilde{s}}{e^Ts'} - \sum_{j=1}^{n} \ln \tilde{s}_j + \sum_{j=1}^{n} \ln {s'}_j \leq \\
&\stackrel{i}{\leq} q\ln\frac{e^T\tilde{s}}{e^Ts'} - n\ln\frac{e^T\tilde{s}}{n} + \frac{1}{32} + n\ln\frac{e^Ts'}{n} = \\
&= (q-n)\ln\frac{e^T\tilde{s}}{e^Ts'} + \frac{1}{32} \leq \\
&\stackrel{ii}{\leq} -(\sqrt{n})\frac{\sqrt{n}-0{,}22\sqrt{n}}{q} + \frac{1}{32} \ \leq \ -\frac{1}{3}
\end{aligned}$$

kde nerovnost (i) plyne z odhadu (3.21) (Lemma 34) a ze vztahu $\sum_{j=1}^{n} \ln s'_j \leq n\ln\frac{e^Ts'}{n}$ vycházejícího z Lemmatu (3.7) (aritmetický a geometrický průměr). Nerovnost (ii) využívá odhad (3.23) a Lemma 26 o odhadu logaritmu. V poslední nerovnosti odhadujeme shora q pomocí $2n$. □

3.5.2 Obecný případ

Pomocí vhodné afinní transformace převedeme obecný případ na snadný (tj. $x = e$), pro který iterační krok provést již umíme. Předpokládejme, že $\bar{x}$ je poslední sestrojené primární řešení a $\bar{s}$ poslední sestrojené duální řešení. Chceme affinní zobrazení, které zobrazí $\bar{x}$ na e. Položme

$$\bar{X} = \begin{pmatrix} \bar{x}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \bar{x}_n \end{pmatrix} . \tag{3.24}$$

Protože $\bar{x}$ je strikně přípustné řešení, je matice $\bar{X}$ regulární a existuje k ní inverzní matice $\bar{X}^{-1}$ a platí

$$\bar{X}^{-1} = \begin{pmatrix} \bar{x}_1^{-1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \bar{x}_n^{-1} \end{pmatrix} . \tag{3.25}$$

Matice $\bar{X}^{-1}$ je matice hledaného afiního zobrazení:

$$\bar{X}^{-1}\bar{x} = e$$

Označme $x' = \bar{X}^{-1}x$ vektor nových proměnných odvozených od x pomoci matice $\bar{X}^{-1}$, a vyjádřeme primární úlohu v řeči těchto nových proměnných.

$$\begin{aligned} \min c^T \bar{X} x' & & \\ A\bar{X}x' &=& b \\ x' &\geq& 0 \end{aligned}$$

Při označení $\bar{A} = A\bar{X}$ a $\bar{c} = c^T\bar{X}$ lze úlohu přepsat následovně:

$$\begin{aligned} \min \bar{c}^T x' & & \\ \bar{A}x' &=& b \\ x' &\geq& 0 \end{aligned} \tag{3.26}$$

Uvědomme si, že x je přípustné řešení pro původní primární úlohu právě tehdy, když $x' = \bar{X}^{-1}x$ je přípustné řešení pro posledně uvedenou úlohu LP.

Duální úloha k (3.26) pak vypadá takto:

$$\begin{aligned} \max b^T y & & \\ (A\bar{X})^T y + s' &=& \bar{X}\bar{c} \\ s' &\geq& 0 \end{aligned}$$

kde $s' = \bar{X}s$.

Následující jednoduché lemma vystihuje klíčovou vlastnost užité afinní transformace: potenciál dvojice přípustných řešení se transformací nezmění.

Lemma 36 (Zachování potenciálu) *Pro regulární diagonální matici $\bar{X}$ s hlavní diagonálou $\bar{x}_1, \ldots, \bar{x}_n$, a pro libovolná přípustná řešení (x, s) původní primární a duální úlohy a pro $x' = \bar{X}^{-1}x$ a $s' = \bar{X}s$ platí*

$$G(x', s') = G(x, s) .$$

Důkaz. Lemma plyne okamžitě z definice potenciálu a předpisů pro x' a s', podle kterých $x'_i = x_i/\bar{x}_i$ a $s'_i = s_i\bar{x}_i$, a tedy $x_i s_i = x'_i s'_i$. □

3.6 Jak ze skoro optimálního řešení najít úplně optimální

Předpokládejme, že máme k dispozici řešení $\bar{x}$ a $(\bar{y}, \bar{s})$ úloh (3.1) a (3.2) taková, že $\bar{x}^T\bar{s} < 2^{-2L}$. Pak pro každé i jistě $\bar{x}_i < 2^{-L}$ nebo $\bar{s}_i < 2^{-L}$. Budeme se snažit každou malou složku nahradit nulou, a ostatní změnit tak, abychom stále měli přípustná řešení. Pokud se nám to podaří, bude výsledná dvojice řešení jistě splňovat podmínky komplementarity (tj. $x^T s = 0$) a půjde tedy o optimální řešení. Podobně jako v předešlé části budeme předpokládat, že matice A a vektory b a c jsou celočíselné.

Pro n-složkový vektor $u \in \mathbb{R}^n$ zavedeme následující značení: $M(u) = \{j \mid u_j < 2^{-L}\}$, $V(u) = \{j \mid u_j \geq 2^{-L}\}$. Pro matici A a množinu indexů J označme A_J podmatici matice A tvořenou sloupci s indexy z J. Začneme dvěma pomocnými lemmaty, s jejichž pomocí pak provedeme naznačený plán na zaokrouhlení.

Lemma 37 *Je-li $\bar{x}$ přípustné řešení úlohy (3.1), pak v polynomiálním čase lze najít přípustné řešení $\hat{x}$ takové, že*

- $\hat{x}_j = \bar{x}_j$ *pro* $j \in M(\bar{x})$,
- *sloupce matice* $A_{V(\hat{x})}$ *jsou lineárně nezávislé.*

Důkaz. Předpokládejme, že sloupce matice $A_{V(\bar{x})}$ jsou lineárně závislé (v opačném případě položíme $\hat{x} = \bar{x}$ a jsme hotovi). Pak jistě homogenní soustava $A_{V(\bar{x})}v = 0$ má netriviální řešení; označme v' nějaké takové řešení doplněné nulami na souřadnicích $M(\bar{x})$ na vektor délky n. Vhodnou volbou skaláru $\lambda \neq 0$ je vektor $\bar{x}' = \bar{x} - \lambda v$ přípustným řešením soustavy (3.1) a navíc $V(\bar{x}') \subsetneq V(\bar{x})$. Opakujeme-li tento postup, dostaneme po nejvýšše n iteracích vektor $\hat{x}$ s kýženými vlastnostmi. □

Lemma 38 *Je-li $\bar{x}$ přípustné řešení úlohy (3.1), pak existuje vrchol v mnohostěnu $P = \{x \mid Ax = b,\ x \geq 0\}$ splňující $v_j = 0$ pro každé $j \in M(\bar{x})$. Obdobně, je-li $(\bar{y}, \bar{s})$ přípustné řešení úlohy (3.2), pak existuje vrchol (y', s') mnohostěnu $Q = \{(y, s) \mid s \geq 0 \text{ and } A^T y + s = c\}$ splňující $A_j^T y' = c_j$ pro každé $j \in M(\bar{s})$.*

Důkaz. V důkazu Věty 23 jsme nahlédli, že každá nenulová složka libovolného vrcholu v mnohostěnu P je vetší než $1/2^{L-n^2}$, kde L je velikost zápisu lineárního programu (3.1). Pro zjednodušení pro zatím předpokládejme, že mnohostěn P je omezený. Pak dané přípustné řešení $\bar{x}$ je, podle Carathéodoryho věty o konvexním obalu, konvexní kombinací nějakých $p \leq n + 1$ jeho vrcholů $v^1, \ldots, v^p$, neboli $\bar{x} = \sum_{i=1}^{p} \lambda_i v^i$ pro nějaké nezáporné koeficienty $\lambda_1, \ldots, \lambda_p$ se součtem jedna. Tudíž jistě alespoň jeden koeficient, označme ho i, splňuje $\lambda_i \geq 1/(n + 1)$. Ukážeme, že $v_j^i = 0$ pro každé $j \in M(\bar{x})$, neboli $v = v^i$ je hledaný vrchol.

Předpokládejme pro spor, že $v_j^i > 0$ pro nějaké $j \in M(\bar{x})$. Pak dokonce $v_j^i > 1/2^L$, protože jde o vrchol P. Použijme tento fakt k dolnímu odhadu velikosti $\bar{x}_j$:

$$\bar{x}_j = \sum_{k=1}^{p} \lambda_k v_j^k \geq \lambda_i v_j^i > \frac{1}{n+1} \cdot \frac{1}{2^{L-n^2}} > \frac{1}{2^L}$$

což je ovšem spor se skutečností $j \in M(\bar{x})$.

Druhá část lemmatu se dokáže obdobně. Podle Carathéodorovy věty existuje $q = m + 1$ vrcholů (y^i, s^i) mnohostěnu Q takových, že $(\bar{y}, \bar{s})$ je jejich konvexní kombinací, neboli $(\bar{y}, \bar{s}) = \sum_{i=1}^{q} \mu_i (y^i, s^i)$ pro vhodné nezáporné koeficienty μ_i se součtem jedna. Nechť i je index splňující $\mu_i \geq 1/(m + 1)$. Ukážeme, že vrchol (y^i, s^i) má požadované vlastnosti. Kdyby pro nějaké $j \in M(\bar{s})$ neplatilo $A_j^T y^i = c_j$, pak podobně jako v předešlém případě dokonce platí $s_j^i > 1/2^L$. Protože $\bar{s}$ je konvexní kombinací $s^1, \ldots, s^q$ s koeficienty $\mu_1, \ldots, \mu_q$, dostaneme opět spor. □

Pro daná přípustná řešení $\bar{x}$ a $\bar{s}$ najdeme nejprve podle Lemmatu 37 přípustné řešení $\hat{x}$ takové, že matice $A_{V(\hat{x})}$ má lineárně nezávislé sloupce. Podle Lemmatu 38 existuje přípustné řešení v primární úlohy se všemi souřadnicemi z $M(\hat{x})$ nulovými, a také přípustné řešení (y', s') duální úlohy se všemi souřadnicemi $V(\hat{x}) \subseteq M(\hat{s})$ nulovými. Protože sloupce matice $A_{V(\hat{x})}$ jsou lineárně nezávislé, je vrchol v určen jednoznačně a lze získat řešením soustavy lineárních rovnic $Ax = b$ při nastavení $x_i = 0$ pro každé $i \in M(\hat{x})$. Protože $v^T s' = 0$, je v hledané optimální řešení.

Pro úplnost ještě dodáváme, že existuje několik dalších postupů pro přechod od skoro optimálního řešení k optimálnímu.

3.7 Shrnutí

Přehlédneme-li drobné implementační obtíže s odmocninami, dopracovali jsme se k následující větě. Připomeňme si, že počet iterací algoritmu je $O(\sqrt{n}L)$ a nejsložitější operací v iteraci je maticové násobení (při Gaussově eliminaci $O(n^3)$ aritmetických operací).

Věta 39 (Karmarkar, 1984) *Existuje algoritmus pro řešení úlohy lineárního programování s použitím $O(n^{3,5}L)$ aritmetických operací.*

Poznámka. Původní Karmarkarův algoritmus z roku 1984 se od zde popsaného algoritmu v mnohém liší (jiný potenciál, jiná transformace, jiný počet iterací, …, nicméně složitost $O(n^{3,5}L)$); tento text vychází především z algoritmu navrženého Ye [9] (viz též Freund [2] a Goemansův učební text [3]). Karmarkarův algoritmus byl ale prvním ze skupiny metod vnitřních bodů, pro které byl dokázán polynomiální odhad na dobu běhu.

Literatura

[1] W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver. *Combinatorial Optimization*. John Wiley, New York, 1997.

[2] R. M. Freund. Polynomial algorithms for linear programming based only on primal scaling and projected gradients of a potential function. *Mathematical Programming*, 51:203–222, 1991.

[3] M. Goemans. Advanced algorithms: Linear programming. Technical report, MIT, Boston, 1994.

[4] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

[5] M. Grötschel and L. Lovász and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer, Berlin, 1988.

[6] R. Seidel. Skript zur Vorlesung Optimierung. Technical report, Universität des Saarlandes, Saarbrücken, 1996.

[7] D. A. Spielman and S.-H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51(3):385–463, 2004.

[8] E. Tardos. A strongly polynomial algorithm to solve combinatorial linear programs. *Oper. Res.*, 34:250–256, 1986.

[9] Y. Ye. An $O(n^3L)$ potential reduction algorithm for linear programming. *Mathematical Programming*, 50:239–258, 1991.