Metoda vnitřního bodu

Petr Kolman

Leden 2011

Abstrakt

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

1 Velmi stručný historický úvod

Ačkoliv soustavám lineárních nerovnic i teorii mnohostěnů včetně jejich praktických aplikací byla pozornost věnována přinejmenším od poloviny 19. století (Fourier, ještě dříve dokonce Lagrange či Bernoulli, dále např. Farkas a Minkowski), za počátek nového oboru nazývaného lineární programování se pokládá až druhá polovina 40. let 20. století. Matematické základy oboru položili především von Neumann a trojice matematiků Gale, Kuhn, Tucker (dualita lineárního programovaní). Jejich bezprostředními předchůdci byli zejména Kantorovič a Hitchcock. Klíčovým momentem pro další vývoj lineárního programování byl Dantzigův simplexový algoritmus, který se na mnoho let stal nezbytnou součástí oboru. 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 doménou především ekonomů.

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). Vzhledem k numerickým slabostem těchto postupů jejich rozkvět nepřerostl výrazně do 70. let.

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 exponeniciální 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, ale žel se záhy ukázalo, že k praktickému použití se metoda nehodí. Hledání dalších alternativ simplexového algoritmu pokračovalo.

Dalším mezníkem v historii lineárního programování byl rok 1984, kdy Karmakar ze společnosti IBM navrhl algoritmus spadající do skupiny metod vnitřních bodů a dokázal, že doba jeho běhu je polynomiální. Jak jsme již zmínili výše, samotná metoda nebyla nová, hlavní Karmakarův 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 vnitřních bodů. 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 balíčcích 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 ke spojení dvou dříve oddělených vědeckých komunit: komunity ekonomů věnujících se lineárnímu programování, a komunity matematiků studujících nelineární optimalizaci.

Na závěr si ještě uvědomme zásadní rozdíl 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 přísně uvnitř mnohostěnu.

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

- Silně polynomiálná algoritmus pro lineární programy v omezené dimenzi - Megiddo [4].
- Silně polynomiální algoritmus pro lineární programy s koeficienty 1, -1, 0 - Tardos [6].
- Zdůvodnění, proč Simplexový algoritmus potřebuje "většinou" pouze polynomiální čas - Spielman a Tang [5].
- Hirschova (polynomiální) hypotéza - do optimálního vrcholu vždy vede "krátká" cesta po hranici mnohostěnu.

2 Náčrt algoritmu

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

$$\begin{aligned} \min c^T x & \qquad (1) \\ Ax &= b \\ x &\geq 0, \end{aligned}$$

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

$$\begin{aligned} \max y^T b & \qquad (2) \\ A^T y + s &= c \\ s &\geq 0, \end{aligned}$$

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 *striktně* 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 striktně 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 striktně 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 $G(x, s) = (n + \sqrt{n}) \ln(x^T s) - \sum_{i=1}^{n} \ln(x_i s_i)$. 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, o kterém jsme dosud nemluvili, 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 zjednodušší práci s gradientem g a jeho projekcí do $Ax = 0$.

Upozornění. Ve stávající verzi textu pomineme hledaní výchozího striktně přípustného řešení i nalezení úplně optimálního řešení z řešení skoro optimálního.

3 Základní definice a pozorování

Lemma 1 *Nechť x je přípustné řešení primární úlohy a s je přípustné řešení duální úlohy. Pak platí*

$$c^T x - y^T b = s^T x \ . \tag{3}$$

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

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

$$G(x,s) = (n+\sqrt{n})\ln(x^T s) - \sum_{i=1}^{n} \ln(x_i s_i) \ . \tag{4}$$

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. Suma $\sum_{i=1}^{n} \ln(x_i s_i)$ v definici potenciálu slouží jako bariéra bránící x a s příliš se některou souřadnicí přiblížit k nule (tj. hranici polyedrů): 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 2 *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*

$$|c^T u - c^T v| > 2^{-2L} \ . \tag{5}$$

Důkaz. Z teorie mnohostěnů víme, že je-li u vrcholem polyedru P, pak u je jediným řešením vhodého podsystému rovnic systému $Ax = b$, $x = 0$. 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 $|\det \hat{A}| < 2^L$ (viz. část 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

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

□

Důsledek 3 *Je-li $x^T s \leq 2^{-2L}$, pro dvojici řešení x a 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 4 (Geometrický a aritmetický průměr) *Pro libovolná $t_j \geq 0,\ j = 1, \ldots, n$, platí:*

$$\left(\prod_{j=1}^{n} t_j \right)^{1/n} \leq \frac{1}{n} \left(\sum_{j=1}^{n} t_j \right) . \tag{6}$$

Lemma 5 (Odhady logaritmů) *Pro $|x| \leq a < 1$ platí*

$$-x - \frac{x^2}{2(1-a)} \leq \ln(1-x) \leq -x \ . \tag{7}$$

Lemma 6 *Je-li A matice plné řádkové hodnosti, pak je matice AA^T regulární.*

4 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 7 *Je-li $G(x, s) \leq -2\sqrt{n}L$, pak $x^T s < e^{-2L}$.*

Důkaz. Použijeme Lemma 4 pro $t_j = x_j s_j$, $j = 1, \ldots, n$. Po zlogaritmování obou stran nerovnosti (4) a drobné úpravě dostaneme nerovnost

$$n \ln x^T s - \sum_{j=1}^{n} \ln x_j s_j \geq n \ln n \ . \tag{8}$$

Podle předpokladů lemmatu, definice potenciálu (4) a nerovnosti (8) 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 3 spolu s Lemmatem 7 zaručují, že při poklesu potenciálu pod $-2\sqrt{n}L$ už jsme takřka v optimálním řešení. Problému nalezení *přesného* optimálního řešení ze *skoro* optimálního řešení se v tomto úvodním textu věnovat nebudeme.

5 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.

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

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

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

Lemma 8 *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 (1) a (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 (1) a (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 (1) a (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. □

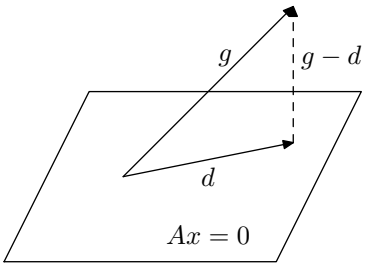
6 Iterační krok

6.1 Snadný případ: $x = e$

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.

$$g = \nabla_x G(x,s)|_{(e,s')} = \left[(n+\sqrt{n})\frac{s}{x^T s} - \begin{pmatrix} \frac{1}{x_1} \\ \vdots \\ \frac{1}{x_n} \end{pmatrix} \right] \Bigg|_{(e,s')} = (n+\sqrt{n})\frac{s'}{e^T s'} - e \ . \tag{9}$$

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

Lemma 9 (Projekce g do nadroviny $\{x \mid Ax = 0\}$)

$$d = (I - A(AA^T)^{-1}A)g \ . \tag{10}$$

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

$$g - d = A^T w \ . \tag{11}$$

Podáří-li se nám tento vetkor w vyjádřit, budeme mít i předpis pro hledaný vektor d. Vynásobením rovnosti (11) 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 (11) dostaneme kýžený vztah (10). □

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

$$\tilde{x} = e - \frac{d}{4\,\|d\|}, \quad \tilde{s} = s' \ . \tag{12}$$

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 10 (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 proto zaručeno $\tilde{x} > 0$. Protože $Ad = 0$, máme zaručenu i přípustnost $\tilde{x}$. □

Lemma 11 (Pokles potenciálu v primárním kroku) *Po provedení primárního kroku platí*

$$G(\tilde{x}, \tilde{s}) - G(e, s') \leq \frac{-1}{120} \ . \tag{13}$$

Důkaz. Pro zkrácení zápisu použijeme v důkazu značení $q = (n + \sqrt{n})$. Po dosazení za $\tilde{x}$ a $\tilde{s}$ podle vzorce (12) dostaneme (indexy nad některými relacemi používáme k pozdějšímu odvolávání se na

ně)

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

Nerovnost (i) vychází z odhadu logaritmu pomocí Lemmatu (5) s parametrem $a = \frac{1}{4}$ (jistě máme $\frac{|d_j|}{4\|d\|} \leq \frac{1}{4}$). Rovnost (ii) používá výpočet gradientu g (9). 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 (9) rovnou nahlédneme, že pro gradient h potenciálu $G(x,s)$ podle s v bodě (e,s') platí

$$h = \nabla_s G(x,s)|_{(e,s')} = (n+\sqrt{n})\frac{e}{e^T s'} - \begin{pmatrix} \frac{1}{s'_1} \\ \vdots \\ \frac{1}{s'_n} \end{pmatrix} . \tag{14}$$

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

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

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

$$A^T y' + s' = c . \tag{15}$$

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

$$A^T \tilde{y} + \tilde{s} = c , \tag{16}$$

máme zaručenu přípustnost $\tilde{s}$. Ze vztahů (15) a (16) plyne pro $\tilde{s}$ omezení

$$\tilde{s} - s' = A^T(y' - \tilde{y}) \ . \tag{17}$$

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

$$\tilde{s} = s' - \frac{e^T s'}{n + \sqrt{n}}(g - d) \ , \quad \tilde{x} = e \ . \tag{18}$$

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

Důkaz. Existence vetkoru $\tilde{y}$ splňujícímu rovnost (16) 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

$$\tilde{s} = s' - \frac{e^T s'}{n + \sqrt{n}}(g - d) = s' - \frac{e^T s'}{n + \sqrt{n}}\left(\frac{n + \sqrt{n}}{e^T s'} s' - e - d\right) = \frac{e^T s'}{n + \sqrt{n}}(d + e) \ . \tag{19}$$

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 13

$$\sum_{j=1}^{n} \ln \tilde{s_j} \quad \geq \quad n \ln \frac{e^T \tilde{s}}{n} - \frac{1}{32} \ . \tag{20}$$

Důkaz.

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

První rovnost plyne ze vztahu (19) odvozeného v důkazu předešlého lemmatu. Při dalších úpravách jsou použity základní vlastnosti funkce logaritmus, její odhady z Lemmatu 5 a horní odhad na velikost $\|d\|$ v duálním kroku. □

Lemma 14 (Pokles potenciálu v duálním kroku) *Po provedení duálního kroku platí*

$$G(\tilde{x},\tilde{s}) - G(e,s') \quad \leq \quad -\frac{1}{3} \ . \tag{21}$$

Důkaz. Začneme pomocným výpočtem, kde v první rovnosti používáme identitu (19) a pro první nerovnost vztah $(\sum_{i=1}^n d_i)^2 \leq n \sum_{i=1}^n d_i^2$ (dokažte si):

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

Dostáváme

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

Pokles potenciálu pak odhadneme takto:

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

kde nerovnost (i) plyne z odhadu (20) (Lemma 13) a ze vztahu $\sum_{j=1}^n \ln s'_j \leq n\ln\frac{e^Ts'}{n}$ vycházejícího z Lemmatu (4) (aritmetický a geometrický průměr). Nerovnost (ii) využívá odhad (22) a Lemma 5 o odhadu logaritmu. V poslední nerovnosti odhadujeme shora q pomocí $2n$. □

6.2 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

$$\bar{X} = \begin{pmatrix} \bar{x}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \bar{x}_n \end{pmatrix} \ . \tag{23}$$

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

$$\bar{X}^{-1} = \begin{pmatrix} \bar{x}_1^{-1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \bar{x}_n^{-1} \end{pmatrix} \ . \tag{24}$$

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:

$$\begin{aligned} \min \bar{c}^T x' & \\ \bar{A}x' &= b \\ x' &\geq 0 \end{aligned} \tag{25}$$

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 (25) pak vypadá takto:

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

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

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

Lemma 15 (Zachování potenciálu) *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 $x'_i = x_i/x_i$ a $s'_i = s_i x_i$. □

7 Jak ze skoro optimálního řešení najít úplně optimální

Pro n-složkový vektor $u \in \mathbb{R}^n$ zavedeme následující značení: $M(u) = \{j \mid u_j < 2^{-2L}\}$, $V(u) = \{j \mid u_j \geq 2^{-2L}\}$. 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 zaokrouhlení.

Lemma 16 *Je-li $\bar{x}$ přípustné řešení (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 (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 17 *Jsou-li $\bar{x}$ a $(\bar{y}, \bar{s})$ přípustné řešení (1) a (2) taková, že $\bar{x}^T\bar{s} < 2^{-4L}$ pak existuje vrchol v mnohostěnu $P = \{x \mid Ax = b,\ x \geq 0\}$ splňující $v_j = 0$ pro každé $j \in M(\bar{x})$.*

Důkaz. Víme, že je-li v vrchol mnohostěnu P, pak každá jeho nenulová složka je vetší než $1/2^L$. Pro zjednodušení pro zatím předpokládejme, že mnohostěn P je omezený. Pak dané přípustné řešení $\bar{x}$ je 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 ($\sum_{i=1}^{p} \lambda_i = 1$). Tudíž jistě alespoň jeden koeficient 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 $\bar{x}_j$:

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

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

Pro daná přípustná řešení $\bar{x}$ a $\bar{s}$ najdeme nejprve podle Lemmatu 16 přípustné řešení $\hat{x}$. Podle Lemmatu 17 existuje vrchol v se všemi souřadnicemi z $M(\hat{x})$ 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 $i \in M(\hat{x})$. Položme $x^\star = v$.

Protože matice $(A_{V(\hat{x})})^T$ má lineárně nezávislé řádky, je soustava $(A_{V(\hat{x})})^T y = c_{V(\hat{x})}$ řešitelná. Nechť $y^\star$ označuje libovolné řešení, a nechť $s^\star = c - A^T y^\star$. Pak pro $i \in M(\hat{x})$ platí $x_i^\star = 0$, a pro $i \in V(\hat{x})$ naopak $s_j^\star = 0$. Řešení $x^\star$ a $s^\star$ splňují podmínky komplementarity a jedná se tedy o optimální řešení.

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

8 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 18 (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 [7] (viz. též Freund [1] a Goemansův učební text [2]). Karmarkarův algoritmus byl ale prvním ze skupiny metod vnitřních bodů, pro které byl dokázán polynomiální odhad na dobu běhu.

Poděkování Dušanovi Knopovi za elektronickou podobu jeho poznámek z běhu přednášky ve školním roce 2008/2009.

Literatura

[1] 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.

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

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

[4] N. Megiddo. Linear programming in linear time when the dimension is fixed. *Journal of Association for Computing Machinery*, 31(1):114–127, Jan. 1984.

[5] 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.

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

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