FIT ČVUT

MI-LOM

Lineární optimalizace a metody

Metody vnitřního bodu (IPMs)

Evropský sociální fond
Praha & EU: Investujeme do vaší budoucnosti

Michal Černý, 2011

Metody vnitřního bodu (IPMs)

Evropský sociální fond
Praha & EU: Investujeme do vaší budoucnosti

Definice 1. *Polyedr* $\{\boldsymbol{x} : \boldsymbol{A}_1\boldsymbol{x} = \boldsymbol{b}_1,\ \boldsymbol{A}_2\boldsymbol{x} \leqslant \boldsymbol{b}_2\}$ *splňuje* ***podmínku vnitřního bodu*** *(interior point condition, IPC), jestliže existuje a je znám bod* $\boldsymbol{x}_0$ *takový, že* $\boldsymbol{A}_1\boldsymbol{x}_0 = \boldsymbol{b}_1$ *a* $\boldsymbol{A}_2\boldsymbol{x}_0 < \boldsymbol{b}_2$. □

Metody vnitřního bodu lze (obecně) spustit jen na lineární program, jehož polyedr splňuje IPC. Máme-li řešit obecný lineární program, o kterém dopředu nevíme nic (ani nevíme, je-li přípustný, natož pak jestli splňuje IPC), je vždy nejprve třeba jej transformovat na jiný program, který IPC splňuje. Nejprve popíšeme jednu takovou transformační strategii, konstrukci tzv. *autoduálního vnoření*. Pak popíšeme jednu speciální metodu vnitřního bodu, která podstatně využívá vlastností tohoto vnoření.

Značení. Jako obvykle se vektory rozumí sloupcové. $\mathbf{1}$ značí sloupcový vektor ze samých jedniček. $\mathbf{0}$ značí vektor či matici samých nul. Jednotkovou matici značíme $\boldsymbol{I}$. Je-li $\alpha \in \mathbb{R}$, pak $\alpha\boldsymbol{x}$ (příp. $\alpha \cdot \boldsymbol{x}$) je vektor se složkami $\alpha \cdot (\boldsymbol{x})_i$. Symbolem $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ rozumíme skalární součin $\sum_i (\boldsymbol{x})_i \cdot (\boldsymbol{y})_i$. Zápisem $\boldsymbol{xy}$ (příp. $\boldsymbol{x} \cdot \boldsymbol{y}$), $\frac{\boldsymbol{x}}{\boldsymbol{y}}$, $\sqrt{\boldsymbol{x}}$, $\frac{1}{\boldsymbol{x}}$, $\boldsymbol{x}^{-1}$ rozumíme operace po složkách. Například $\mathbf{1}\boldsymbol{x} = \boldsymbol{x}$. Jiný příklad: $\frac{\mathbf{1}}{\boldsymbol{x}} = \frac{1}{\boldsymbol{x}} = \boldsymbol{x}^{-1}$ je vektor se složkami $\frac{1}{(\boldsymbol{x})_i}$. $\sqrt{\boldsymbol{x}}$ je vektor složkami $\sqrt{(\boldsymbol{x})_i}$.

Všimněme si, že platí

$$\mathbf{1}^{\mathrm{T}}(\boldsymbol{xy}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}. \tag{1}$$

1 Autoduální vnoření

Nechť je dán lineární program

$$\min\{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} : \boldsymbol{x} \in \mathcal{F}\}, \quad \text{kde } \mathcal{F} := \{\boldsymbol{x} : \boldsymbol{Ax} \geqslant \boldsymbol{b},\ \boldsymbol{x} \geqslant \mathbf{0}\}; \tag{2}$$

matice $\boldsymbol{A}$ je rozměru $m \times n$. Je jasné, že libovolný lineární program lze representovat v tomto tvaru. Řešme dva úkoly:

(a) (rozhodovací problém) rozhodnout, zdali lineární program (2) má optimum;

(b) (optimalisační problém) má-li lineární program (2) optimum, úkolem je najít některé optimální řešení.

Z přednášky o dualitě víme, že schopnost řešit problém (a) stačí k řešení optimalisační úlohy v plné obecnosti.

Duální program k (2) má tvar (prokažte jako cvičení!)

$$\max\{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} : \boldsymbol{y} \in \mathcal{G}\}, \quad \text{kde } \mathcal{G} := \{\boldsymbol{y} : \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \leqslant \boldsymbol{c},\ \boldsymbol{y} \geqslant \mathbf{0}\}. \tag{3}$$

Ze slabé věty o dualitě víme, že pro libovolné $\boldsymbol{x} \in \mathcal{F}$ a $\boldsymbol{y} \in \mathcal{G}$ platí

$$\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} \leqslant \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}. \tag{4}$$

Z věty o dualitě víme, že stačí hledat libovolné řešení systému

$$\begin{array}{rcll} \boldsymbol{Ax} & \geqslant & \boldsymbol{b}, & \boldsymbol{x} \geqslant \mathbf{0}, \\ -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} & \geqslant & -\boldsymbol{c}, & \boldsymbol{y} \geqslant \mathbf{0}, \\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} & \geqslant & 0. & \end{array} \tag{5}$$

Od systému (5) přejděme k systému

$$\begin{array}{rcll} \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\kappa & \geqslant & \boldsymbol{0}, & \qquad \boldsymbol{x} \geqslant \boldsymbol{0}, \\ -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{c}\kappa & \geqslant & \boldsymbol{0}, & \qquad \boldsymbol{y} \geqslant \boldsymbol{0}, \\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} & \geqslant & 0, & \qquad \kappa \geqslant 0. \end{array} \tag{6}$$

Pro něj očividně platí:

Lemma 2. *Řešení systému (5) si jednoznačně odpovídají s řešeními systému (6), která splňují* $\kappa = 1$. □

Odtud plyne (dokažte jako cvičení):

Důsledek 3. *(i) Systém (6) má řešení splňující $\kappa > 0$, právě když úloha (2) má optimum.*

(ii) Jestliže $(\boldsymbol{x}, \boldsymbol{y}, \kappa)$ je řešení systému (6) splňující $\kappa > 0$, pak $\frac{1}{\kappa}\boldsymbol{x}$ je optimální řešení úlohy (2) a $\frac{1}{\kappa}\boldsymbol{y}$ je optimální řešení úlohy (3). □

Tedy: k řešení rozhodovacího problému stačí rozhodnout, zdali existuje řešení systému (6) splňující $\kappa > 0$. A k řešení optimalizačního problému stačí nějaké takové řešení najít.

Systém (6) je homogenní ($\boldsymbol{0}$ je totiž řešením); proměnné κ se proto říká **homogenisační proměnná**.

Systém (6) můžeme psát v maticovém tvaru

$$\boldsymbol{C}\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \\ \kappa \end{pmatrix} \geqslant \boldsymbol{0}, \quad \boldsymbol{y} \geqslant \boldsymbol{0}, \quad \boldsymbol{x} \geqslant \boldsymbol{0}, \quad \kappa \geqslant \boldsymbol{0},$$

kde

$$\boldsymbol{C} := \begin{pmatrix} \boldsymbol{0} & \boldsymbol{A} & -\boldsymbol{b} \\ -\boldsymbol{A}^{\mathrm{T}} & \boldsymbol{0} & \boldsymbol{c} \\ \boldsymbol{b}^{\mathrm{T}} & -\boldsymbol{c}^{\mathrm{T}} & 0 \end{pmatrix}$$

($\boldsymbol{0}$ značí nulovou matici příslušného rozměru). Všimněme si, že matice $\boldsymbol{C}$ je kososymetrická.

Položme si otázku, zdali systém (6) splňuje IPC. Odpověď je záporná (cvičení). Modifikujme tedy systém (6) do tvaru, ve kterém bude IPC splněna. Položme $N := m + n + 2$,

$$\boldsymbol{r} := \boldsymbol{1} - \boldsymbol{C}\boldsymbol{1}, \quad \boldsymbol{D} := \begin{pmatrix} \boldsymbol{C} & \boldsymbol{r} \\ -\boldsymbol{r}^{\mathrm{T}} & 0 \end{pmatrix}, \quad \boldsymbol{q} := \begin{pmatrix} \boldsymbol{0}_{(N-1)\times 1} \\ N \end{pmatrix}$$

a uvažme systém

$$\boldsymbol{D}\boldsymbol{z} \geqslant -\boldsymbol{q}, \quad \boldsymbol{z} \geqslant \boldsymbol{0}, \quad \text{kde} \quad \boldsymbol{z} := \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \\ \kappa \\ \vartheta \end{pmatrix}. \tag{7}$$

Všimněme si, že matice $\boldsymbol{D}$ je kososymetrická.

Systém (7) má následující pozoruhodné vlastnosti. Než je zformulujeme a prokážeme, potřebujeme snadné lemma, jehož důkaz ponecháváme jako cvičení.

Lemma 4 (ortogonální lemma)**.** *Pro libovolnou kososymetrickou matici $\boldsymbol{\Xi}$ a libovolný vektor $\boldsymbol{v}$ platí $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\Xi}\boldsymbol{v} = 0$.* □

Lemma 5. *(a) Systém (7) splňuje IPC s $\boldsymbol{z} = \boldsymbol{1}$.*

(b) Lineární program

$$\min\{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} : \boldsymbol{D}\boldsymbol{z} \geqslant -\boldsymbol{q}, \ \boldsymbol{z} \geqslant \boldsymbol{0}\} \tag{8}$$

je přípustný, optimální hodnota jeho účelové funkce je 0 a má optimální řešení s $\kappa > 0$, právě když systém (6) má řešení s $\kappa > 0$.

Důkaz. (a)

$$\begin{aligned}
\boldsymbol{D1} &= \begin{pmatrix} \boldsymbol{C} & \boldsymbol{r} \\ -\boldsymbol{r}^{\mathrm{T}} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{1} \\ 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{C1} + \boldsymbol{r} \\ -\boldsymbol{r}^{\mathrm{T}}\boldsymbol{1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{C1} + \boldsymbol{1} - \boldsymbol{C1} \\ -\boldsymbol{1}^{\mathrm{T}}\boldsymbol{r} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{1} \\ -\boldsymbol{1}^{\mathrm{T}}(\boldsymbol{1} - \boldsymbol{C1}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{1} \\ -\boldsymbol{1}^{\mathrm{T}}\boldsymbol{1} + \boldsymbol{1}^{\mathrm{T}}\boldsymbol{C1} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{1} \\ -(N-1) \end{pmatrix} > \begin{pmatrix} \boldsymbol{0} \\ -N \end{pmatrix} = -\boldsymbol{q}.
\end{aligned} \tag{9}$$

Užili jsme toho, že $\boldsymbol{1}^{\mathrm{T}}\boldsymbol{C1} = 0$ z ortogonálního lemmatu (víme, že $\boldsymbol{C}$ je kososymetrická), a $\boldsymbol{1}^{\mathrm{T}}\boldsymbol{1} = N - 1$, protože zde $\boldsymbol{1}$ je vektor $N - 1$ jedniček.

(b). Pro účelovou funkci programu (8) platí $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = N\vartheta$. Protože $\boldsymbol{z} \geqslant \boldsymbol{0}$, platí $\vartheta \geqslant 0$, a tedy hodnota účelové funkce je vždy $\geqslant 0$. Ovšem vektor $\boldsymbol{z} = \boldsymbol{0}$ je přípustný, a tedy optimální hodnota účelové funkce je 0. Optimální prostor lineárního programu (8) je tedy prostor těch $(\boldsymbol{x}, \boldsymbol{y}, \kappa, \vartheta)$, pro která platí $\vartheta = 0$. Důkaz zbytku tvrzení (b) ponecháváme jako cvičení.

□

Pro řešení rozhodovací úlohy tedy stačí rozhodnout, zdali existuje optimální řešení programu (8) splňující $\kappa > 0$. Pro řešení optimalisační úlohy stačí najít libovolné optimum programu (8) splňující $\kappa > 0$. Navíc systém (8) splňuje IPC.

Všimněme si, že lineární program (8) je sám sobě duálem (dokažte jako cvičení); proto se mu říká **autoduální vnoření**. Proměnné ϑ se říká **normalisační proměnná**.

2 Slack

Definice 6. *Vektoru* $\boldsymbol{s}(\boldsymbol{z}) := \boldsymbol{Dz} + \boldsymbol{q}$ *se říká* ***slack****.* □

Název vektoru $\boldsymbol{s}(\boldsymbol{z})$ plyne z toho, že pro libovolné $\boldsymbol{z}$ platí

$$\boldsymbol{Dz} - \boldsymbol{s}(\boldsymbol{z}) = -\boldsymbol{q};$$

tedy, $\boldsymbol{s}(\boldsymbol{z})$ je rozdíl mezi levými a pravými stranami nerovnosti $\boldsymbol{Dz} \geqslant -\boldsymbol{q}$. Speciálně platí, že $\boldsymbol{s}(\boldsymbol{z}) \geqslant \boldsymbol{0}$, právě když $\boldsymbol{Dz} \geqslant -\boldsymbol{q}$. Tedy:

Lemma 7. *Vektor* $\boldsymbol{z}$ *je přípustným vektorem úlohy (8), právě když* $\boldsymbol{z} \geqslant \boldsymbol{0}$ *a* $\boldsymbol{s}(\boldsymbol{z}) \geqslant \boldsymbol{0}$*.* □

Lemma 8. $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = \boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(z)$.

Důkaz. $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = \boldsymbol{z}^{\mathrm{T}}\boldsymbol{q} = \boldsymbol{z}^{\mathrm{T}}(\boldsymbol{s}(\boldsymbol{z}) - \boldsymbol{Dz}) = \boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) - \boldsymbol{z}^{\mathrm{T}}\boldsymbol{Dz} = \boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z})$. (Užili jsme ortogonální lemma.) □

Odtud plyne zajímavé zjištění, které je přímým důsledkem toho, že pro optimální řešení $\boldsymbol{z}$ úlohy (8) platí $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = 0$:

Důsledek 9 (complementary slackness theorem)**.** *Nechť* $\boldsymbol{z} \geqslant \boldsymbol{0}$ *a* $\boldsymbol{s}(\boldsymbol{z}) \geqslant \boldsymbol{0}$*. Pak platí*

(a) $\boldsymbol{z}$ *je optimální řešení úlohy (8), právě když* $\boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) = 0$*.*

(b) $\boldsymbol{z}$ *je optimální řešení úlohy (8), právě když* $\boldsymbol{z} \cdot \boldsymbol{s}(\boldsymbol{z}) = \boldsymbol{0}$*.* □

Slack má i další zajímavé vlastnosti:

Lemma 10. *(a)* $\boldsymbol{s}(\mathbf{1}) = \mathbf{1}$.

(b) $\mathbf{1}^{\mathrm{T}}\boldsymbol{z} + \mathbf{1}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) = N + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}$.

Důkaz. (a) $\boldsymbol{s}(\mathbf{1}) = \boldsymbol{D}\mathbf{1} + \boldsymbol{q} = \left(\begin{smallmatrix}\mathbf{1} \\ -(N-1)\end{smallmatrix}\right) + \left(\begin{smallmatrix}\mathbf{0} \\ N\end{smallmatrix}\right) = \mathbf{1}$ (viz (9)).

(b) Z ortogonálního lemmatu plyne $(\mathbf{1} - \boldsymbol{z})^{\mathrm{T}}\boldsymbol{D}(\mathbf{1} - \boldsymbol{z}) = 0$; odtud

$$\begin{aligned}
0 &= (\mathbf{1} - \boldsymbol{z})^{\mathrm{T}}\boldsymbol{D}(\mathbf{1} - \boldsymbol{z}) \\
&= (\mathbf{1} - \boldsymbol{z})^{\mathrm{T}}(\boldsymbol{D}\mathbf{1} - \boldsymbol{D}\boldsymbol{z}) \\
&= (\mathbf{1} - \boldsymbol{z})^{\mathrm{T}}(\boldsymbol{s}(\mathbf{1}) - \boldsymbol{q} - (\boldsymbol{s}(\boldsymbol{z}) - \boldsymbol{q})) \\
&= (\mathbf{1} - \boldsymbol{z})^{\mathrm{T}}(\mathbf{1} - \boldsymbol{s}(\boldsymbol{z})) \\
&= \mathbf{1}^{\mathrm{T}}\mathbf{1} - \mathbf{1}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) - \mathbf{1}^{\mathrm{T}}\boldsymbol{z} + \boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) \\
&= N + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} - \mathbf{1}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) - \mathbf{1}^{\mathrm{T}}\boldsymbol{z}. \quad \square
\end{aligned}$$

Důsledek 11. *(a) Je-li* $\boldsymbol{z}$ *optimální řešení úlohy (8), pak* $\mathbf{1}^{\mathrm{T}}\boldsymbol{z} + \mathbf{1}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) = N$.

(b) Optimální prostor úlohy (8) je omezený.

Důkaz. (a) je snadný důsledek tvrzení (b) předchozího lemmatu: v optimu totiž platí $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = 0$.

(b) Protože $\boldsymbol{s}(\boldsymbol{z}) \geqslant \mathbf{0}$, z (a) jest $\sum_{i=1}^{N}(\boldsymbol{z})_i \leqslant N - \mathbf{1}^{\mathrm{T}}\boldsymbol{s}(z) \leqslant N$. Protože $\boldsymbol{z} \geqslant \mathbf{0}$, jest $(\forall i)[0 \leqslant (\boldsymbol{z})_i \leqslant N]$. □

3 Basické a nebasické proměnné

Hovoříme-li nadále o optimálním řešení, máme na mysli optimální řešení úlohy (8).

Definice 12. *(a)* $B := \{i$ *: existuje optimální řešení* $\boldsymbol{z}$ *splňující* $(\boldsymbol{z})_i > 0.\}$*. Indexy v* B *se nazývají* ***basické****.*

(b) $\overline{B} := \{i$ *: existuje optimální řešení* $\boldsymbol{z}$ *splňující* $(\boldsymbol{s}(\boldsymbol{z}))_i > 0.\}$*. Indexy v* $\overline{B}$ *se nazývají* ***nebasické****.* □

K řešení rozhodovacího problému tedy stačí odpovědět otázku, zdali $B \ni N - 1$ (to je ekvivalentní tomu, zdali existuje optimální řešení s $\kappa > 0$).

Vyslovme nyní důležitou — vskutku pozoruhodnou a překvapivou — větu, jejíž důkaz podáme později:

Věta 13. $B \cup \overline{B} = \{1, \ldots, N\}$ a $B \cap \overline{B} = \emptyset$. □

Definice 14. *(a) Dvojice vektorů* $\boldsymbol{x} \geqslant \mathbf{0}$ a $\boldsymbol{y} \geqslant \mathbf{0}$ *se nazývá* ***striktně komplementární****, jestliže* $\boldsymbol{x}\boldsymbol{y} = \mathbf{0}$ a $\boldsymbol{x} + \boldsymbol{y} > \mathbf{0}$.

(b) Řešení $\boldsymbol{z}$ *úlohy (8) se nazývá* ***striktně komplementární****, jestliže vektory* $\boldsymbol{z}$ *a* $\boldsymbol{s}(\boldsymbol{z})$ *jsou striktně komplementární.* □

Je jasné, že každé striktně komplementární řešení je optimální (viz důsledek 9). Položme si otázku, zdali vůbec nějaké striktně komplementární řešení optimum existuje; pokud ano, jeho nalezení poskytne plnou informaci o rozkladu indexové množiny $\{1, \ldots, N\}$ do B a $\overline{B}$, a tedy také poskytne informaci, zdali existuje optimum s $\kappa > 0$.

4 Centrální cesta

Víme, že $\boldsymbol{z} = \mathbf{1}$ je vnitřním bodem úlohy (8) a že platí $\boldsymbol{s}(\mathbf{1}) = \mathbf{1}$ (a tedy $\boldsymbol{z}\boldsymbol{s}(\boldsymbol{z}) = \mathbf{1}$ pro $\boldsymbol{z} = \mathbf{1}$). Víme také, že vektor $\boldsymbol{z}$ je optimální, právě když

$$\boldsymbol{z} \geqslant \mathbf{0}, \quad \boldsymbol{s}(\boldsymbol{z}) \geqslant \mathbf{0}, \quad \boldsymbol{z}\boldsymbol{s}(\boldsymbol{z}) = \mathbf{0}. \tag{10}$$

Tento systém může mít nekonečně mnoho řešení. Uvažme obecnější systém

$$\boldsymbol{z} \geqslant \mathbf{0}, \quad \boldsymbol{s}(\boldsymbol{z}) \geqslant \mathbf{0}, \quad \boldsymbol{z}\boldsymbol{s}(\boldsymbol{z}) = \mu \cdot \mathbf{1}, \tag{11}$$

kde $\mu \geqslant 0$. Systém (10) je speciálním případem s $\mu = 0$. Všimněme si, že $\boldsymbol{z} = \mathbf{1}$ je řešením (11) s $\mu = 1$. Následující věta, kterou dokážeme později, ukazuje pozoruhodnou vlastnost systému (11), omezíme-li se na $\mu > 0$.

Věta 15. *(a) Pro každé $\mu > 0$ má systém (11) právě jedno řešení. (Definice: toto řešení se nazývá **μ-střed** a značí se $\boldsymbol{z}(\mu)$ a jemu příslušný slack se značí $\boldsymbol{s}(\mu)$.)*

*(b) $\{\boldsymbol{z}(\mu) : \mu > 0\}$ je spojitá křivka uvnitř polyedru $\boldsymbol{D}\boldsymbol{z} \geqslant -\boldsymbol{q}$, $\boldsymbol{z} \geqslant \mathbf{0}$; nazývá se **centrální cesta**.*

(c) $\boldsymbol{z}_0 := \lim_{\mu \searrow 0} \boldsymbol{z}(\mu)$ je striktně komplementární optimum úlohy (8). □

Snadno se vidí, jaká je hodnota účelové funkce v μ-středu:

Lemma 16. $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}(\mu) = N \cdot \mu$.

Důkaz. $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}(\mu) \overset{\text{lemma 8}}{=} \boldsymbol{z}(\mu)^{\mathrm{T}}\boldsymbol{s}(\mu) \overset{(1)}{=} \mathbf{1}^{\mathrm{T}}(\boldsymbol{z}(\mu)\boldsymbol{s}(\mu)) = \mathbf{1}^{\mathrm{T}}(\mu\mathbf{1}) = N \cdot \mu.$ □

To znamená, že podél centrální cesty při $\mu \searrow 0$ hodnota účelové funkce klesá úměrně tempu poklesu μ.

5 Aproximace μ-středu

Nechť je dáno $\boldsymbol{z} > \mathbf{0}$ splňující $\boldsymbol{s} := \boldsymbol{s}(\boldsymbol{z}) > \mathbf{0}$ a $\mu > 0$. Úkolem je najít $\Delta\boldsymbol{z}$ takové, že $\boldsymbol{z}^+ = \boldsymbol{z}(\mu)$, kde jsme označili

$$\boldsymbol{z}^+ = \boldsymbol{z} + \Delta\boldsymbol{z}.$$

Najít takové $\Delta\boldsymbol{z}$ se nám sice nepovede (s rozumným výpočetním úsilím) přesně, ale je přibližně. Označme dále

$$\boldsymbol{s}^+ := \boldsymbol{s}(\boldsymbol{z}^+) = \boldsymbol{D}(\boldsymbol{z} + \Delta\boldsymbol{z}) + \boldsymbol{q} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{q} + \boldsymbol{D}\Delta\boldsymbol{z} = \boldsymbol{s} + \boldsymbol{D}\Delta\boldsymbol{z}$$

a

$$\Delta\boldsymbol{s} := \boldsymbol{s}^+ - \boldsymbol{s} = \boldsymbol{D}\Delta\boldsymbol{z}.$$

Snadným užitím ortogonálního lemmatu obdržíme:

Lemma 17. *Vektory $\Delta\boldsymbol{z}$ a $\Delta\boldsymbol{s}$ jsou ortogonální (to jest: $\Delta\boldsymbol{z}^{\mathrm{T}}\Delta\boldsymbol{s} = 0$).* □

Má-li platit $\boldsymbol{z}^+ = \boldsymbol{z}(\mu)$, musí být splněno

$$\boldsymbol{z}^+ \geqslant \mathbf{0}, \quad \boldsymbol{s}^+ \geqslant \mathbf{0}, \quad \boldsymbol{z}^+\boldsymbol{s}^+ = \mu\mathbf{1}.$$

Rovnici $\boldsymbol{z}^+\boldsymbol{s}^+ = \mu\mathbf{1}$ můžeme psát ve tvaru

$$\mu\mathbf{1} = \boldsymbol{z}^+\boldsymbol{s}^+ = (\boldsymbol{z} + \Delta\boldsymbol{z})(\boldsymbol{s} + \Delta\boldsymbol{s}) = \boldsymbol{z}\boldsymbol{s} + \boldsymbol{s}\Delta\boldsymbol{z} + \boldsymbol{z}\Delta\boldsymbol{s} + \Delta\boldsymbol{z}\Delta\boldsymbol{s}.$$

Nyní učiňme heuristický krok: zanedbejme kvadratický člen $\Delta\boldsymbol{z}\Delta\boldsymbol{s}$. Obdržíme pak rovnici $\boldsymbol{z}\boldsymbol{s} + \boldsymbol{s}\Delta\boldsymbol{z} + \boldsymbol{z}\Delta\boldsymbol{s} = \mu\mathbf{1}$ s neznánými $\Delta\boldsymbol{s}$ a $\Delta\boldsymbol{z}$. Víme, že musí platit $\Delta\boldsymbol{s} = \boldsymbol{D}\Delta\boldsymbol{z}$; řešme tedy *lineární* systém

$$\boldsymbol{D}\Delta\boldsymbol{z} - \Delta\boldsymbol{s} = \mathbf{0}, \quad \boldsymbol{z}\Delta\boldsymbol{s} + \boldsymbol{s}\Delta\boldsymbol{z} = \mu\mathbf{1} - \boldsymbol{z}\boldsymbol{s} \tag{12}$$

($\boldsymbol{z}, \boldsymbol{s}$ jsou konstanty a $\Delta\boldsymbol{z}, \Delta\boldsymbol{s}$ jsou neznámé). Snadno se vidí (dokažte jako cvičení), že tento systém má jediné řešení $\Delta\boldsymbol{z}$ a $\Delta\boldsymbol{s}$. Díky zanedbání kvadratického členu $\Delta\boldsymbol{z}\Delta\boldsymbol{s}$ nemůžeme tvrdit, že $\boldsymbol{z} + \Delta\boldsymbol{z} = \boldsymbol{z}(\mu)$; je ale možné, že $\boldsymbol{z}^+$ ($= \boldsymbol{z} + \Delta\boldsymbol{z}$) je v nějakém smyslu blízko $\boldsymbol{z}(\mu)$. Ovšem ani nevíme, zdali $\boldsymbol{z}^+ > \mathbf{0}$ a $\boldsymbol{s}^+ > \mathbf{0}$.

Nicméně obdržíme tuto pozoruhodnou vlastnost: hodnota účelové funkce v bodě $\boldsymbol{z}^+$ je stejná jako hodnota účelové funkce v $\boldsymbol{z}(\mu)$ (viz lemma 16). Můžeme tedy říci, že *krok* $\Delta\boldsymbol{z}$ *nás posune na tu vrstevnici účelové funkce, na níž leží μ-střed.*

Lemma 18. $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}^+ = N \cdot \mu.$

Důkaz.

$$\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}^+ \overset{\text{lemma 8}}{=} (\boldsymbol{z}^+)^{\mathrm{T}}\boldsymbol{s}^+ \overset{(1)}{=} \mathbf{1}^{\mathrm{T}}(\boldsymbol{z}^+\boldsymbol{s}^+) = \mathbf{1}^{\mathrm{T}}[(\boldsymbol{z} + \Delta\boldsymbol{s})(\boldsymbol{s} + \Delta\boldsymbol{s})] = \mathbf{1}^{\mathrm{T}}(\boldsymbol{z}\boldsymbol{s} + \underbrace{\boldsymbol{z}\Delta\boldsymbol{s} + \boldsymbol{s}\Delta\boldsymbol{z}}_{=\mu\mathbf{1}-\boldsymbol{z}\boldsymbol{s}} + \Delta\boldsymbol{z}\Delta\boldsymbol{s})$$

$$= \mathbf{1}^{\mathrm{T}}(\mu\mathbf{1} + \Delta\boldsymbol{z}\Delta\boldsymbol{s}) = \mu\mathbf{1}^{\mathrm{T}}\mathbf{1} + \mathbf{1}^{\mathrm{T}}(\Delta\boldsymbol{s}\Delta\boldsymbol{z}) \overset{(1)}{=} N\mu + \Delta\boldsymbol{z}^{\mathrm{T}}\Delta\boldsymbol{s} \overset{\text{lemma 17}}{=} N\mu. \quad \square$$

Zavedeme důležitou míru vzdálenosti bodu $\boldsymbol{z}$ od μ-středu.

Definice 19. *Vektor* $\boldsymbol{v}_\mu(\boldsymbol{z}) := \sqrt{\frac{1}{\mu}\boldsymbol{z}\boldsymbol{s}(\boldsymbol{z})}$ *se nazývá* ***varianční****.* ***δ-vzdálenost*** *vektoru* $\boldsymbol{z}$ *od* $\boldsymbol{z}(\mu)$ *je definována vztahem* $\delta(\boldsymbol{z}, \mu) := \frac{1}{2}\|\boldsymbol{v}_\mu(\boldsymbol{z}) - \frac{1}{\boldsymbol{v}_\mu(\boldsymbol{z})}\|$*, kde* $\|\cdot\|$ *značí eukleidovskou normu.* $\square$

Snadno se vidí, že $\delta(\boldsymbol{z}, \mu) = 0$, právě když $\boldsymbol{z}$ je μ-středem; protože navíc $\delta(\boldsymbol{z}, \mu) \geqslant 0$, je δ mírou vzdálenosti od μ-středu. Nazývá se také *δ-norma*.

Přichází důležitá věta, jejíž důkaz podáme později.

Věta 20 (o newtonovském kroku)**.** *Nechť je dáno* $\boldsymbol{z}$ *a* $\mu > 0$ *takové, že* $1 > \delta := \delta(\boldsymbol{z}, \mu)$*. Pak*

(a) $\boldsymbol{z}^+ > \mathbf{0}$ a $\boldsymbol{s}^+ > \mathbf{0}$,

(b) $\delta^+ := \delta(\boldsymbol{z}^+, \mu) \leqslant \frac{\delta^2}{\sqrt{2(1-\delta^2)}}$. $\square$

Důsledek 21 (věta o kvadratické konvergenci newtonovské aproximace)**.** *Jestliže* $\delta \leqslant \frac{1}{\sqrt{2}}$*, pak* $\delta^+ \leqslant \delta^2$. $\square$

6 Algoritmus

Připomeňme, že (i) $\boldsymbol{z} = \mathbf{1}$ je 1-střed a (ii) centrální cesta konverguje k striktně komplementárnímu optimu (naším cílem je právě takové optimum najít; to nám totiž dovolí rozhodnout, zdali $B \ni N - 1$). Navíc víme, že newtonovský krok nás (v jistém smyslu) dokáže přiblížit k μ-středu s předepsaným μ. Nabízí se tedy tento nápad: použijme centrální cestu jako referenční cestu, podél níž se (s jistou chybou) pokusíme přiblížit ke kýženému striktně komplementárnímu optimu. Tuto myšlenku shrnuje následující algoritmus.

[1] Vstup: lineární program (8), $\theta \in (0,1)$ a $\varepsilon > 0$.
[2] $\boldsymbol{z} := \mathbf{1}$, $\mu := 1$
[3] while $N\mu \geqslant \varepsilon$ do
[4] $\quad\mu := (1-\theta)\mu$
[5] $\quad\boldsymbol{z} := \boldsymbol{z} + \Delta\boldsymbol{z}$ ($\Delta\boldsymbol{z}$ je řešením (12))
[6] end while.

Připomeňme, že $N\mu$ je hodnota účelové funkce v μ-středu; připomeňme zde též lemma 18.

Otvírá se řada přirozených otázek, které musíme zodpovědět: (1) Jak volit θ, aby v každé iteraci byl vektor $\boldsymbol{z}$ přípustný a přitom nebyl příliš vzdálen od μ-středu? [Číslu θ se říká *barrier update parameter*.] (2) Jak zvolit ε, aby $\boldsymbol{z}$ v poslední iteraci while-cyklu bylo „dostatečně" blízko striktně komplementárnímu optimu? (3) Jak z bodu $\boldsymbol{z}$, který je „dostatečně" blízko striktně komplementárnímu optimu, získat rozklad B a $\overline{B}$? (4) A nedalo by se nějak vyčíst přesné striktně komplementární optimum? (Tím bychom totiž získali přímo hledané hodnoty $\frac{1}{\kappa}\boldsymbol{x}$ a $\frac{1}{\kappa}\boldsymbol{y}$.)

Lemma 22. *Nechť $\boldsymbol{z} > \mathbf{0}$, $\boldsymbol{s}(\boldsymbol{z}) > \mathbf{0}$, $\mu > 0$, $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = N\mu$, $\delta := \delta(\boldsymbol{z},\mu)$ a $\mu^+ := (1-\theta)\mu$. Pak platí*

$$\delta^+ := \delta(\boldsymbol{z},\mu^+) = (1-\theta)\delta^2 + \frac{\theta^2 N}{4(1-\theta)}.$$

Důkaz. Nechť $\boldsymbol{v} := \boldsymbol{v}_\mu(\boldsymbol{z}) = \sqrt{\frac{1}{\mu}\boldsymbol{z}\boldsymbol{s}(\mu)}$ je varianční vektor. Z definice δ-normy platí

$$4(\delta^+)^2 = \left\| \sqrt{1-\theta}\,\boldsymbol{v}^{-1} - \frac{\boldsymbol{v}}{\sqrt{1-\theta}} \right\|^2 = \left\| \sqrt{1-\theta}(\boldsymbol{v}^{-1} - \boldsymbol{v}) - \frac{\theta\boldsymbol{v}}{\sqrt{1-\theta}} \right\|^2.$$

Všimněme si, že platí

$$\begin{aligned}\|\boldsymbol{v}\|^2 &= \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} = \sum_i (\boldsymbol{v})_i^2 = \sum_i \left(\sqrt{\frac{(\boldsymbol{z})_i(\boldsymbol{s}(\boldsymbol{z}))_i}{\mu}}\right)^2 = \frac{1}{\mu}\sum_i (\boldsymbol{z})_i(\boldsymbol{s}(\boldsymbol{z}))_i = \frac{1}{\mu}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z})\\ &= \frac{1}{\mu}\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = \frac{1}{\mu}\mu N = N.\end{aligned}$$

Odtud plyne, že vektory $\boldsymbol{v}$ a $\boldsymbol{v}^{-1} - \boldsymbol{v}$ jsou ortogonální, protože

$$\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{v}^{-1} - \boldsymbol{v}) = \boldsymbol{v}^{\mathrm{T}}(\boldsymbol{v}^{-1}) - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} = N - N = 0.$$

Z Pythagorovy věty tedy plyne[1]

$$\begin{aligned}4(\delta^+)^2 &= \left\| \sqrt{1-\theta}(\boldsymbol{v}^{-1} - \boldsymbol{v}) - \frac{\theta\boldsymbol{v}}{\sqrt{1-\theta}} \right\|^2 = \|\sqrt{1-\theta}(\boldsymbol{v}^{-1} - \boldsymbol{v})\|^2 + \left\| \frac{\theta\boldsymbol{v}}{\sqrt{1-\theta}} \right\|^2\\ &= (1-\theta)\|\boldsymbol{v}^{-1} - \boldsymbol{v}\|^2 + \frac{\theta^2\|\boldsymbol{v}\|^2}{1-\theta} = (1-\theta)\|\boldsymbol{v}^{-1} - \boldsymbol{v}\|^2 + \frac{\theta^2 N}{1-\theta}.\end{aligned}$$

Z definice δ-normy jest $2\delta = \|\boldsymbol{v} - \boldsymbol{v}^{-1}\|$. Tím je důkaz dokončen. □

Následující věta ukazuje, že volba $\theta = \frac{1}{\sqrt{2N}}$ zajistí důležitou vlastnost algoritmu.

[1] Cvičení: dokažte, že pro ortogonální vektory $\boldsymbol{x}$ a $\boldsymbol{y}$ platí $\|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$.

Věta 23. *Nechť* $\theta = \frac{1}{\sqrt{2N}}$. *Pak platí, že na začátku každé iterace while-cyklu jest* $\boldsymbol{z} > \boldsymbol{0}$, $\boldsymbol{s}(\boldsymbol{z}) > \boldsymbol{0}$, $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = N\mu$ *a* $\delta := \delta(\boldsymbol{z}, \mu) \leqslant \frac{1}{2}$.

Důkaz. Že věta platí pro první iteraci, je jasné. Postupujme dále indukcí; nechť víme, že věta platí v jisté iteraci, a prozkoumejme, co se uvnitř iterace stane. Nejprve klademe

$$\mu^+ = (1-\theta)\mu = (1 - \frac{1}{\sqrt{2N}})\mu;$$

odtud (vzhledem k předchozímu lemmatu) platí

$$\begin{aligned}\delta(\boldsymbol{z}, \mu^+)^2 &= (1-\theta)\delta^2 + \frac{\theta^2 N}{4(1-\theta)} \overset{\delta \leqslant \frac{1}{2}}{\leqslant} \frac{1}{4}(1-\theta) + \frac{\theta^2 N}{4(1-\theta)} \leqslant \frac{1}{4}(1-\theta) + \frac{\frac{1}{2N} \cdot N}{4(1-\theta)} \\ &= \frac{1}{4}(1-\theta) + \frac{1}{8(1-\theta)} \overset{(\star)}{\leqslant} \frac{3}{8} < \frac{1}{2}.\end{aligned}$$

Nerovnost $(\star)$ dokažte jako cvičení (návod: vyšetřete průběh funkce $\frac{1}{4}(1-\theta) + \frac{1}{8(1-\theta)}$ a užijte faktu $N \geqslant 2$). Platí tedy $\delta(\boldsymbol{z}, \mu^+) \leqslant \frac{1}{\sqrt{2}}$. Důkaz dokončí věta 20(a) a důsledek 21. □

Nyní tedy víme, že k danému $\varepsilon > 0$ jsme schopni najít přípustné $\boldsymbol{z}$, které je ε-blízko striktně komplementárnímu optimu (měřeno hodnotou účelové funkce) a od příslušného μ-středu je vzdáleno nanejvýš $\frac{1}{2}$ (měřeno δ-normou). Zbývá ukázat, jak volit ε a jak poté z přibližného řešení zjistit rozklad B a $\overline{B}$.

7 Nalezení rozkladu B a $\overline{B}$

Definice 24. *Nechť* $\boldsymbol{z} > \boldsymbol{0}$ *a* $\boldsymbol{s}(\boldsymbol{z}) > \boldsymbol{0}$. *Pak definujeme* $\Delta(\boldsymbol{z}) := \frac{\max_i(\boldsymbol{z}\boldsymbol{s}(\boldsymbol{z}))_i}{\min_i(\boldsymbol{z}\boldsymbol{s}(\boldsymbol{z}))_i}$. □

Všimněme si, že $\Delta(\boldsymbol{z}) = 1$, právě když $\boldsymbol{z}$ leží na centrální cestě; jinak je $\Delta(\boldsymbol{z}) > 1$. Tedy, $\Delta(\boldsymbol{z})$ měří vzdálenost bodu $\boldsymbol{z}$ od centrální cesty.

Lemma 25. *V každé iteraci algoritmu platí* $\Delta(\boldsymbol{z}) \leqslant 4$.

Důkaz. Víme, že v každé iteraci platí $\boldsymbol{z} > \boldsymbol{0}$, $\boldsymbol{s}(\boldsymbol{z}) > \boldsymbol{0}$, $\mu > 0$ a $\delta(\boldsymbol{z}, \mu) \leqslant \frac{1}{2}$. Můžeme psát

$$\Delta(\boldsymbol{z}) = \frac{\max_i(\boldsymbol{z}\boldsymbol{s}(\boldsymbol{z}))_i}{\min_i(\boldsymbol{z}\boldsymbol{s}(\boldsymbol{z}))_i} = \frac{\max_i(\mu \cdot \boldsymbol{v}_\mu(\boldsymbol{z})^2)_i}{\min_i(\mu \cdot \boldsymbol{v}_\mu(\boldsymbol{z})^2)_i} = \frac{\max_i(\boldsymbol{v}_\mu(\boldsymbol{z})^2)_i}{\min_i(\boldsymbol{v}_\mu(\boldsymbol{z})^2)_i}.$$

Bez újmy na obecnosti můžeme předpokládat, že $(\boldsymbol{v}_\mu(\boldsymbol{z}))_1 \geqslant (\boldsymbol{v}_\mu(\boldsymbol{z}))_2 \geqslant \cdots \geqslant (\boldsymbol{v}_\mu(\boldsymbol{z}))_N$. Pak

$$\Delta(\boldsymbol{z}) = \frac{\max_i(\boldsymbol{v}_\mu(\boldsymbol{z})^2)_i}{\min_i(\boldsymbol{v}_\mu(\boldsymbol{z})^2)_i} = \frac{(\boldsymbol{v}_\mu(\boldsymbol{z}))_1^2}{(\boldsymbol{v}_\mu(\boldsymbol{z}))_N^2}. \tag{13}$$

Navíc z $\delta(\boldsymbol{z}, \mu) \leqslant \frac{1}{2}$ plyne $\|\boldsymbol{v}_\mu(\boldsymbol{z}) - \frac{1}{\boldsymbol{v}_\mu(\boldsymbol{z})}\| \leqslant 1$.

Jako cvičení ukažte, že $\max\left\{\frac{(\boldsymbol{v})_1^2}{(\boldsymbol{v})_N^2} : \|\boldsymbol{v} - \frac{1}{\boldsymbol{v}}\| \leqslant 1\right\} = 4$. Zřejmě je $\boldsymbol{v}_\mu(\boldsymbol{z})$ přípustným bodem tohoto optimalisačního problému; odtud a z (13) plyne, že $\Delta(\boldsymbol{z}) \leqslant 4$. □

Z důsledku 11(b) víme, že optimální prostor je omezený. Označme jej

$$\boldsymbol{Q} := \{\boldsymbol{z} : \boldsymbol{z} \geqslant \boldsymbol{0},\ \boldsymbol{s}(\boldsymbol{z}) \geqslant \boldsymbol{0},\ \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = 0\}.$$

Z omezenosti $\boldsymbol{Q}$ plyne, že číslo

$$\Sigma_i := \max_{\boldsymbol{z}\in\boldsymbol{Q}}[(\boldsymbol{z})_i + (\boldsymbol{s}(\boldsymbol{z}))_i]$$

je omezené.

Definice 26. *Číslo* $\Sigma := \min_i \Sigma_i$ *se nazývá* ***condition number****.* □

Věta 27. *Nechť* $\boldsymbol{z} > \boldsymbol{0}$, $\boldsymbol{s}(\boldsymbol{z}) > \boldsymbol{0}$ a $\Delta(\boldsymbol{z}) \leqslant 4$.

(a) *Jestliže* $i \in B$, *pak*

$$(\boldsymbol{z})_i \geqslant \frac{\Sigma}{4N}, \quad (\boldsymbol{s}(\boldsymbol{z}))_i \leqslant \frac{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}}{\Sigma}.$$

(b) *Jestliže* $i \in \overline{B}$, *pak*

$$(\boldsymbol{z})_i \leqslant \frac{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}}{\Sigma}, \quad (\boldsymbol{s}(\boldsymbol{z}))_i \geqslant \frac{\Sigma}{4N}.$$

Důkaz. (b) Nechť $i \in \overline{B}$. Označme $\widehat{\boldsymbol{z}}$ optimální řešení splňující $(\forall \boldsymbol{z} \in \boldsymbol{Q})[(\boldsymbol{s}(\widehat{\boldsymbol{z}}))_i \geqslant (\boldsymbol{s}(\boldsymbol{z}))_i]$ (z kompaktnosti $\boldsymbol{Q}$ jistě $\widehat{\boldsymbol{z}}$ existuje). Jistě platí

$$(\boldsymbol{s}(\widehat{\boldsymbol{z}}))_i \geqslant \Sigma \tag{14}$$

(plyne to z definice Σ a z faktu, že $i \notin B$, pročež $(\boldsymbol{z})_i = 0$ pro všechna $\boldsymbol{z} \in \boldsymbol{Q}$). Z ortogonálního lemmatu plyne

$$\begin{aligned}
0 &= (\boldsymbol{z} - \widehat{\boldsymbol{z}})^{\mathrm{T}}\boldsymbol{D}(\boldsymbol{z} - \widehat{\boldsymbol{z}}) = (\boldsymbol{z} - \widehat{\boldsymbol{z}})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{z} - \boldsymbol{D}\widehat{\boldsymbol{z}}) = (\boldsymbol{z} - \widehat{\boldsymbol{z}})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{z} + \boldsymbol{q} - (\boldsymbol{D}\widehat{\boldsymbol{z}} + \boldsymbol{q})) \\
&= (\boldsymbol{z} - \widehat{\boldsymbol{z}})^{\mathrm{T}}(\boldsymbol{s}(\boldsymbol{z}) - \boldsymbol{s}(\widehat{\boldsymbol{z}})) = \boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) - \boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\widehat{\boldsymbol{z}}) - \widehat{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) + \widehat{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{s}(\widehat{\boldsymbol{z}}) \\
&= \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} - \boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\widehat{\boldsymbol{z}}) - \widehat{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) + \boldsymbol{q}^{\mathrm{T}}\widehat{\boldsymbol{z}} = \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} - \boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\widehat{\boldsymbol{z}}) - \widehat{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}).
\end{aligned}$$

Odtud

$$\boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\widehat{\boldsymbol{z}}) + \widehat{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) = \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z},$$

a tedy

$$(\boldsymbol{z})_i \cdot (\boldsymbol{s}(\widehat{\boldsymbol{z}}))_i \leqslant \boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\widehat{\boldsymbol{z}}) \leqslant \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z},$$

odkud

$$(\boldsymbol{z})_i \leqslant \frac{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}}{(\boldsymbol{s}(\widehat{\boldsymbol{z}}))_i} \overset{(14)}{\leqslant} \frac{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}}{\Sigma}.$$

Tím je dokázána první část tvrzení (a). Dále

$$\begin{aligned}
(\boldsymbol{s}(\boldsymbol{z}))_i &= \frac{(\boldsymbol{z})_i \cdot (\boldsymbol{s}(\boldsymbol{z}))_i}{(\boldsymbol{z})_i} \geqslant \frac{\Sigma}{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}} \cdot (\boldsymbol{s}(\boldsymbol{z}))_i(\boldsymbol{z})_i = \frac{\Sigma}{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z})} \cdot (\boldsymbol{s}(\boldsymbol{z}))_i(\boldsymbol{z})_i \\
&\geqslant \frac{\Sigma}{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z})} \cdot \min_j(\boldsymbol{s}(\boldsymbol{z}))_j(\boldsymbol{z})_j = \frac{\Sigma}{\sum_j (\boldsymbol{z})_j(\boldsymbol{s}(\boldsymbol{z}))_j} \cdot \min_j(\boldsymbol{s}(\boldsymbol{z}))_j(\boldsymbol{z})_j \\
&\geqslant \frac{\Sigma}{N \cdot \max_j(\boldsymbol{z})_j(\boldsymbol{s}(\boldsymbol{z}))_j} \cdot \min_j(\boldsymbol{s}(\boldsymbol{z}))_j(\boldsymbol{z})_j = \frac{\Sigma}{N \cdot \Delta(\boldsymbol{z})} \geqslant \frac{\Sigma}{4N}.
\end{aligned}$$

Tvrzení (a) se dokáže podobně (cvičení). □

Důsledek 28. *Nechť* $\boldsymbol{z} > \boldsymbol{0}$, $\boldsymbol{s}(\boldsymbol{z}) > \boldsymbol{0}$, $\Delta(\boldsymbol{z}) \leqslant 4$ a $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} < \frac{\Sigma^2}{4N}$*. Pak platí*

$$\{i : (\boldsymbol{z})_i > (\boldsymbol{s}(\boldsymbol{z}))_i\} = B, \quad \{i : (\boldsymbol{z})_i < (\boldsymbol{s}(\boldsymbol{z}))_i\} = \overline{B}.$$

Důkaz. Je-li $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}$ tak malé, že platí $\frac{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}}{\Sigma} < \frac{\Sigma}{4N}$, poslední lemma dává jednoznačný rozklad indexové množiny $\{1,\dots,N\}$ do B a $\overline{B}$. A nerovnost $\frac{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}}{\Sigma} < \frac{\Sigma}{4N}$ je splněna, právě když $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} < \frac{\Sigma^2}{4N}$. □

Nyní by se chtělo říci: stačí zvolit ε tak, aby platilo $N\mu = \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} < \frac{\Sigma^2}{4N}$. Ovšem není jasné, jak spočítat číslo Σ. K tomu se hodí následující odhad, jehož důkaz podáme později.

Věta 29. *$\Sigma \geqslant \frac{1}{\Pi_{i=1}^{N}\|\boldsymbol{D}_i\|} =: \frac{1}{\Pi}$, kde $\boldsymbol{D}_i$ je i-tý sloupec $\boldsymbol{D}$.* □

Číslo Π se spočte snadno ze samotné vstupní instance.

Nyní již dokážeme odhadnout počet iterací. Označme L bitovou velikost zápisu matice $\boldsymbol{D}$.

Věta 30. *Rozklad B a $\overline{B}$ dokážeme najít po nanejvýš $O(\sqrt{N}L)$ iteracích.*

Důkaz. Z přednášky o Chačijanově větě známe odhad $\|\boldsymbol{x}\| \leqslant 2^{\sigma(\boldsymbol{x})}$, kde $\sigma(\boldsymbol{x})$ značí velikost zápisu $\boldsymbol{x}$. Odtud

$$\Pi = \prod_{i=1}^{N}\|\boldsymbol{D}_i\| \leqslant \prod_{i=1}^{N} 2^{\sigma(\boldsymbol{D}_i)} = 2^{\sum_{i=1}^{N}\sigma(\boldsymbol{D}_i)} = 2^{\sigma(\boldsymbol{D})} = 2^L,$$

a tedy

$$2^{-2L} \leqslant \frac{1}{\Pi^2}.$$

Nechť algoritmus skončí nanejvýš po $k+1$ iteracích, kde k splňuje

$$N\cdot(1-\theta)^k = \varepsilon$$

(ε zvolíme později); odtud

$$k\cdot\log_2(1-\theta) + \log_2 N = \log_2\varepsilon,$$

a také

$$-k\cdot\theta + \log_2 N \geqslant \log_2\varepsilon,$$

protože $\theta \leqslant -\log_2(1-\theta)$ (dokažte tuto nerovnost jako cvičení; užijte známé nerovnosti $1+x \leqslant e^x$ [tu rovněž dokažte jako cvičení]). Odtud

$$k \leqslant \frac{1}{\theta}\log_2\frac{N}{\varepsilon}.$$

Tedy, algoritmus skončí nanejvýš po

$$1 + \frac{1}{\theta}\log_2\frac{N}{\varepsilon} = 1 + \sqrt{2N}\log_2\frac{N}{\varepsilon}$$

iteracích. Zvolme $\varepsilon = \frac{2^{-2L}}{4N}$; pak

$$\begin{aligned} 1 + \sqrt{2N}\log_2\frac{N}{\varepsilon} &= 1 + \sqrt{2N}\log_2(2^{2L}\cdot 4N^2) = 1 + \sqrt{2N}(2L + 2 + 2\log_2 N) \\ &\leqslant 1 + \sqrt{2N}(2L+2+L) = O(\sqrt{N}L). \end{aligned}$$

Přitom v poslední iteraci platí

$$\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = N\mu < \varepsilon = \frac{2^{-2L}}{4N} \leqslant \frac{1}{\Pi^2\cdot 4N} \leqslant \frac{\Sigma^2}{4N},$$

a tedy můžeme použít důsledek 28. □

Zde naši analýzu ukončeme: prokázali jsme, že za omezený počet iterací se nechá najít rozklad B a $\overline{B}$, pročež jsme schopni rozhodnout, zdali existuje optimum s $\kappa > 0$; jak víme, to stačí k řešení úlohy lineárního programování v plné obecnosti. Nicméně by bylo lze pokračovat — nechá se dokonce ukázat, že hledané řešení $\frac{1}{\kappa}\boldsymbol{x}$ a $\frac{1}{\kappa}\boldsymbol{y}$ se dokonce nechá *najít* (existuje-li optimum s $\kappa > 0$). Stačí k tomu udělat pár iterací navíc, ovšem ne řádově více než $O(\sqrt{N}L)$.

Problému, jak algoritmus implementovat na Turingově stroji tak, aby pracoval v *turingovském* polynomiálním čase (tj. s racionálními čísly, která mají omezenou velikost zápisu), se zde věnovat nebudeme. Zdůrazněme proto, že jsme prokázali rychlost algoritmu *toliko ve výpočetním modelu, kde se nezabýváme délkou zápisů čísel a složitostí aritmetických operací.*

***

8 Důkaz věty 13 — první část

V této části dokážeme, že $B \cap \overline{B} = \emptyset$. Fakt $B \cup \overline{B} = \{1, \ldots, n\}$ je přímým důsledkem existence striktně komplementárního optima. (Jeho existenci dokážeme později.)

Důkaz tvrzení $B \cap \overline{B} = \emptyset$. Nejprve vyslovme tuto pomocnou větu: *jestliže $\boldsymbol{z}_1$ a $\boldsymbol{z}_2$ jsou přípustná, pak platí, že jsou optimální, právě když $\boldsymbol{z}_1\boldsymbol{s}(\boldsymbol{z}_2) = \boldsymbol{0}$ a $\boldsymbol{z}_2\boldsymbol{s}(\boldsymbol{z}_1) = \boldsymbol{0}$.*

S pomocnou větou je důkaz snadný. Nechť existuje $i \in B \cap \overline{B}$. Pak existují dvě optima $\boldsymbol{z}_1 \neq \boldsymbol{z}_2$ taková, že $(\boldsymbol{z}_1)_i > 0$ a $(\boldsymbol{s}(\boldsymbol{z}_2))_i > 0$. Pak ale $(\boldsymbol{z}_1\boldsymbol{s}(\boldsymbol{z}_2))_i > 0$, a to je spor s pomocnou větou.

Důkaz pomocné věty. Z ortogonálního lemmatu plyne

$$
\begin{aligned}
0 &= (\boldsymbol{z}_1 - \boldsymbol{z}_2)^{\mathrm{T}}\boldsymbol{D}(\boldsymbol{z}_1 - \boldsymbol{z}_2) \\
&= (\boldsymbol{z}_1 - \boldsymbol{z}_2)^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{z}_1 - \boldsymbol{D}\boldsymbol{z}_2) \\
&= (\boldsymbol{z}_1 - \boldsymbol{z}_2)^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{z}_1 + \boldsymbol{q} - (\boldsymbol{D}\boldsymbol{z}_2 + \boldsymbol{q})) \\
&= (\boldsymbol{z}_1 - \boldsymbol{z}_2)^{\mathrm{T}}(\boldsymbol{s}(\boldsymbol{z}_1) - \boldsymbol{s}(\boldsymbol{z}_2)) \\
&= \boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_1) - \boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_2) - \boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_1) + \boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_2) \\
&= \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}_1 - \boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_2) - \boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_1) + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}_2 \\
&= -\boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_2) - \boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_1).
\end{aligned}
$$

Tedy platí $\boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_2) + \boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_1) \overset{(\star)}{=} 0$, a protože $\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{s}(\boldsymbol{z}_1), \boldsymbol{s}(\boldsymbol{z}_2) \geqslant \boldsymbol{0}$, rovnost $(\star)$ nastává, právě když $\boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_2) = 0$ a $\boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}_1) = 0$, a to nastává, právě když $\boldsymbol{z}_1\boldsymbol{s}(\boldsymbol{z}_2) = \boldsymbol{0}$ a $\boldsymbol{z}_2\boldsymbol{s}(\boldsymbol{z}_1) = \boldsymbol{0}$. □

9 Konvergence newtonovského kroku

Lemma 31 (tzv. $\boldsymbol{uv}$-lemma)**.** *Nechť $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = 0$. Pak*

$$\|\boldsymbol{uv}\|_\infty \leqslant \frac{1}{4}\|\boldsymbol{u} + \boldsymbol{v}\|^2, \tag{15}$$

$$\|\boldsymbol{uv}\| \leqslant \frac{\sqrt{2}}{4}\|\boldsymbol{u} + \boldsymbol{v}\|^2, \tag{16}$$

kde $\|\cdot\|$ značí L_2-normu ($\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$) a $\|\cdot\|_\infty$ značí L_∞-normu ($\|\boldsymbol{x}\|_\infty = \max\{|(\boldsymbol{x})_1|, \ldots, |(\boldsymbol{x})_n|\}$).

Důkaz. Platí
$$\boldsymbol{uv} = \frac{1}{4}((\boldsymbol{u}+\boldsymbol{v})^2 - (\boldsymbol{u}-\boldsymbol{v})^2) \tag{17}$$
(protože $(\boldsymbol{u}+\boldsymbol{v})^2 - (\boldsymbol{u}-\boldsymbol{v})^2 = \boldsymbol{u}^2 + 2\boldsymbol{uv} + \boldsymbol{v}^2 - (\boldsymbol{u}^2 - 2\boldsymbol{uv} + \boldsymbol{v}^2) = 4\boldsymbol{uv}$), a tedy
$$-\frac{1}{4}(\boldsymbol{u}-\boldsymbol{v})^2 \leqslant \boldsymbol{uv} \leqslant \frac{1}{4}(\boldsymbol{u}+\boldsymbol{v})^2.$$
Odtud plyne
$$-\frac{1}{4}\|\boldsymbol{u}-\boldsymbol{v}\|^2 \cdot \mathbf{1} \leqslant \boldsymbol{uv} \leqslant \frac{1}{4}\|\boldsymbol{u}+\boldsymbol{v}\|^2 \cdot \mathbf{1}.$$
Protože $\boldsymbol{u}$ a $\boldsymbol{v}$ jsou ortogonální, platí $\|\boldsymbol{u}+\boldsymbol{v}\| = \|\boldsymbol{u}-\boldsymbol{v}\|$ (dokažte jako cvičení); odtud
$$|\boldsymbol{uv}| \leqslant \frac{1}{4}\|\boldsymbol{u}+\boldsymbol{v}\|^2 \cdot \mathbf{1},$$
a tedy $\|\boldsymbol{uv}\|_\infty \leqslant \frac{1}{4}\|\boldsymbol{u}+\boldsymbol{v}\|^2$. Tím je dokázána nerovnost (15).

Z (17) plyne
$$\begin{aligned}\|\boldsymbol{uv}\|^2 = \mathbf{1}^{\mathrm{T}}(\boldsymbol{uv})^2 = \frac{1}{16}\mathbf{1}^{\mathrm{T}}((\boldsymbol{u}+\boldsymbol{v})^2-(\boldsymbol{u}-\boldsymbol{v})^2)^2 = \frac{1}{16}\mathbf{1}^{\mathrm{T}}((\boldsymbol{u}+\boldsymbol{v})^4-2(\boldsymbol{u}+\boldsymbol{v})^2(\boldsymbol{u}-\boldsymbol{v})^2+(\boldsymbol{u}-\boldsymbol{v})^4)\\ \leqslant \frac{1}{16}\mathbf{1}^{\mathrm{T}}((\boldsymbol{u}+\boldsymbol{v})^4 + (\boldsymbol{u}-\boldsymbol{v})^4) \overset{(\star)}{\leqslant} \frac{1}{16}(\|\boldsymbol{u}+\boldsymbol{v}\|^4 + \|\boldsymbol{u}-\boldsymbol{v}\|^4) = \frac{2}{16}(\|\boldsymbol{u}+\boldsymbol{v}\|^4);\end{aligned}$$
čímž je dokázána nerovnost (16). Opět jsme využili toho, že z ortogonality $\boldsymbol{u}$ a $\boldsymbol{v}$ plyne $\|\boldsymbol{u}+\boldsymbol{v}\| = \|\boldsymbol{u}-\boldsymbol{v}\|$. Nerovnost $(\star)$ plyne z toho, že pro libovolné $\boldsymbol{x}$ platí $\mathbf{1}^{\mathrm{T}}\boldsymbol{x}^4 \leqslant \|\boldsymbol{x}\|^4$ (dokažte jako cvičení). □

Lemma 32. *Nechť $\mu > 0$, $\boldsymbol{z} > \mathbf{0}$ a $\boldsymbol{s} := \boldsymbol{s}(\boldsymbol{z}) > \mathbf{0}$. Položme $\delta := \delta(\boldsymbol{z}, \mu)$. Nechť $\Delta\boldsymbol{z}$ a $\Delta\boldsymbol{s}$ označují newtonovský krok. Pak $\|\Delta\boldsymbol{z}\Delta\boldsymbol{s}\|_\infty \leqslant \mu\delta^2$ a $\|\Delta\boldsymbol{z}\Delta\boldsymbol{s}\| \leqslant \sqrt{2}\mu\delta^2$.*

Důkaz. Z newtonovského kroku víme, že platí
$$\boldsymbol{z}\Delta\boldsymbol{s} + \boldsymbol{s}\Delta\boldsymbol{z} = \mu\mathbf{1} - \boldsymbol{zs}.$$
Pak
$$\frac{1}{\sqrt{\boldsymbol{zs}}} \cdot \boldsymbol{z}\Delta\boldsymbol{s} + \frac{1}{\sqrt{\boldsymbol{zs}}} \cdot \boldsymbol{s}\Delta\boldsymbol{z} = \frac{1}{\sqrt{\boldsymbol{zs}}} \cdot (\mu\mathbf{1} - \boldsymbol{zs}),$$
což je ekvivalentní s
$$\sqrt{\frac{\boldsymbol{z}}{\boldsymbol{s}}}\Delta\boldsymbol{s} + \sqrt{\frac{\boldsymbol{s}}{\boldsymbol{z}}}\Delta\boldsymbol{z} = \frac{1}{\sqrt{\boldsymbol{zs}}}(\mu\mathbf{1} - \boldsymbol{zs}),$$
a také s
$$\sqrt{\frac{\boldsymbol{z}}{\boldsymbol{s}}}\Delta\boldsymbol{s} + \sqrt{\frac{\boldsymbol{s}}{\boldsymbol{z}}}\Delta\boldsymbol{z} = \sqrt{\mu}\left(\frac{1}{\boldsymbol{v}} - \boldsymbol{v}\right),$$
kde $\boldsymbol{v} := \sqrt{\frac{\boldsymbol{zs}}{\mu}}$ označuje variančnı́ vektor (užili jsme faktu $\sqrt{\mu}\boldsymbol{v} = \sqrt{\boldsymbol{zs}}$).

Snadno se vidí, že vektory $\sqrt{\frac{\boldsymbol{z}}{\boldsymbol{s}}}\Delta\boldsymbol{s}$ a $\sqrt{\frac{\boldsymbol{s}}{\boldsymbol{z}}}\Delta\boldsymbol{z}$ jsou ortogonální (dokažte jako cvičení; užijte lemma 17). Platí
$$\sqrt{\frac{\boldsymbol{z}}{\boldsymbol{s}}}\Delta\boldsymbol{s} \cdot \sqrt{\frac{\boldsymbol{s}}{\boldsymbol{z}}}\Delta\boldsymbol{z} = \Delta\boldsymbol{z}\Delta\boldsymbol{s};$$

podle $\boldsymbol{uv}$-lemmatu tedy

$$\begin{aligned}
\|\Delta \boldsymbol{z} \Delta \boldsymbol{s}\| &= \left\| \sqrt{\frac{\boldsymbol{z}}{\boldsymbol{s}}} \Delta \boldsymbol{s} \cdot \sqrt{\frac{\boldsymbol{s}}{\boldsymbol{z}}} \Delta \boldsymbol{z} \right\| \\
&\leqslant \frac{\sqrt{2}}{4} \left\| \sqrt{\frac{\boldsymbol{z}}{\boldsymbol{s}}} \Delta \boldsymbol{s} + \sqrt{\frac{\boldsymbol{s}}{\boldsymbol{z}}} \Delta \boldsymbol{z} \right\|^2 \\
&= \frac{\sqrt{2}}{4} \left\| \sqrt{\mu} \left(\frac{1}{\boldsymbol{v}} - \boldsymbol{v} \right) \right\|^2 \\
&= \frac{\sqrt{2}}{4} \cdot \mu \cdot \left\| \frac{1}{\boldsymbol{v}} - \boldsymbol{v} \right\|^2 \\
&= \frac{\sqrt{2}}{4} \cdot \mu \cdot (2\delta)^2 \\
&= \sqrt{2} \mu \delta^2.
\end{aligned}$$

Podobně se ukáže $\|\Delta \boldsymbol{z} \Delta \boldsymbol{s}\|_\infty \leqslant \mu\delta^2$. (Užili jsme toho, že $\|\frac{1}{\boldsymbol{v}} - \boldsymbol{v}\| = 2\delta$ z definice δ-normy.) □

Důkaz věty 20. Označme $\boldsymbol{z}_\alpha := \boldsymbol{z} + \alpha \Delta \boldsymbol{z}$ a $\boldsymbol{s}_\alpha := \boldsymbol{s} + \alpha \Delta \boldsymbol{s}$, kde $\alpha \in [0,1]$ je libovolné. Prokažme tuto pomocnou větu: *pro každé* $\alpha \in [0,1]$ *jest* $\boldsymbol{z}_\alpha \boldsymbol{s}_\alpha > \boldsymbol{0}$. Na základě předchozího lemmatu platí nerovnost $(\star)$ v

$$\begin{aligned}
\boldsymbol{z}_\alpha \boldsymbol{s}_\alpha &= (\boldsymbol{z} + \alpha \Delta \boldsymbol{z})(\boldsymbol{s} + \alpha \Delta \boldsymbol{s}) = \boldsymbol{z}\boldsymbol{s} + \alpha \boldsymbol{z} \Delta \boldsymbol{s} + \alpha \boldsymbol{s} \Delta \boldsymbol{z} + \alpha^2 \Delta \boldsymbol{z} \Delta \boldsymbol{s} = \boldsymbol{z}\boldsymbol{s} + \alpha(\mu \boldsymbol{1} - \boldsymbol{z}\boldsymbol{s}) + \alpha^2 \Delta \boldsymbol{z} \Delta \boldsymbol{s} \\
&= (1-\alpha)\boldsymbol{z}\boldsymbol{s} + \alpha\mu\boldsymbol{1} + \alpha^2 \Delta \boldsymbol{z} \Delta \boldsymbol{s} = (1-\alpha)\boldsymbol{z}\boldsymbol{s} + \alpha(\mu \boldsymbol{1} + \alpha \Delta \boldsymbol{z} \Delta \boldsymbol{s}) \\
&\geqslant (1-\alpha)\boldsymbol{z}\boldsymbol{s} + \alpha(\mu\boldsymbol{1} - \alpha \|\Delta \boldsymbol{z} \Delta \boldsymbol{s}\|_\infty \boldsymbol{1}) = (1-\alpha)\boldsymbol{z}\boldsymbol{s} + \alpha(\mu - \alpha\|\Delta \boldsymbol{z} \Delta \boldsymbol{s}\|_\infty)\boldsymbol{1} \\
&\overset{(\star)}{\geqslant} (1-\alpha)\boldsymbol{z}\boldsymbol{s} + \alpha(\mu - \alpha\mu\delta^2)\boldsymbol{1} = (1-\alpha)\boldsymbol{z}\boldsymbol{s} + \alpha\mu \underbrace{(1-\alpha\delta^2)}_{>0} \boldsymbol{1} > \boldsymbol{0};
\end{aligned}$$

poslední nerovnost plyne z toho, že $\boldsymbol{z}\boldsymbol{s} > \boldsymbol{0}$, $\alpha \leqslant 1$ a $\delta < 1$. Tím je pomocná věta prokázána.

Speciálně pro $\alpha = 1$ tedy víme, že $\boldsymbol{z}_\alpha \boldsymbol{s}_\alpha = \boldsymbol{z}^+ \boldsymbol{s}^+ > \boldsymbol{0}$. Odtud plyne, že pro každé i platí buďto $(\boldsymbol{z}^+)_i > 0$ a $(\boldsymbol{s}^+)_i > 0$, anebo $(\boldsymbol{z}^+)_i < 0$ a $(\boldsymbol{s}^+)_i < 0$. Zbývá ukázat, že druhá možnost nemůže nastat.

Pro spor nechť existuje i tak, že $(\boldsymbol{z})_i < 0$ a $(\boldsymbol{s}_i) < 0$. Víme, že $(\boldsymbol{z})_i = (\boldsymbol{z}_{\alpha=0})_i > 0$ a $(\boldsymbol{z}^+)_i = (\boldsymbol{z}_{\alpha=1})_i < 0$. Všimněme si, že funkce $\alpha \mapsto (\boldsymbol{z}_\alpha)_i$ je na $\alpha \in [0,1]$ spojitá. Tedy, z Bolzanovy věty existuje $\alpha_0 \in (0,1)$ takové, že $(\boldsymbol{z}_{\alpha_0})_i = 0$. To je ovšem spor s tím, že $(\boldsymbol{z}_{\alpha_0})_i (\boldsymbol{s}_{\alpha_0})_i > 0$ (spor s pomocnou větou při $\alpha = \alpha_0$).

Dokázali jsme, že $\boldsymbol{z}^+ > \boldsymbol{0}$ a $\boldsymbol{s}^+ > \boldsymbol{0}$.

Zbývá ukázat tvrzení (b). Označme $\delta^+ := \delta(\boldsymbol{z}^+, \mu)$ a $\boldsymbol{v}^+ := \sqrt{\frac{\boldsymbol{v}^+ \boldsymbol{s}^+}{\mu}}$. Z definice δ-normy jest

$$2\delta^+ = \left\| \frac{1}{\boldsymbol{v}^+} - \boldsymbol{v}^+ \right\| = \left\| \frac{1}{\boldsymbol{v}^+} (\boldsymbol{1} - (\boldsymbol{v}^+)^2) \right\|.$$

Protože

$$\boldsymbol{z}^+ \boldsymbol{s}^+ = (\boldsymbol{z} + \Delta \boldsymbol{z})(\boldsymbol{s} + \Delta \boldsymbol{s}) = \boldsymbol{z}\boldsymbol{s} + \underbrace{\boldsymbol{z}\Delta \boldsymbol{s} + \boldsymbol{s}\Delta \boldsymbol{z}}_{=\mu\boldsymbol{1} - \boldsymbol{z}\boldsymbol{s}} + \Delta \boldsymbol{z} \Delta \boldsymbol{s} = \mu \boldsymbol{1} + \Delta \boldsymbol{z} \Delta \boldsymbol{s},$$

platí

$$(\boldsymbol{v}^+)^2 = \frac{\boldsymbol{z}^+ \boldsymbol{s}^+}{\mu} = \frac{1}{\mu}(\mu \boldsymbol{1} + \Delta \boldsymbol{z} \Delta \boldsymbol{s}) = \boldsymbol{1} + \frac{\Delta \boldsymbol{z} \Delta \boldsymbol{s}}{\mu},$$

a tedy (s užitím lemmatu 32)

$$\begin{aligned}
2\delta^+ &= \left\| \frac{1}{\boldsymbol{v}^+}(\mathbf{1} - (\boldsymbol{v}^+)^2) \right\| = \left\| \frac{1}{\sqrt{\mathbf{1} + \frac{\Delta \boldsymbol{z} \Delta \boldsymbol{s}}{\mu}}} \left(\mathbf{1} - \left(\mathbf{1} + \frac{\Delta \boldsymbol{z} \Delta \boldsymbol{s}}{\mu} \right) \right) \right\| \\
&= \left\| \frac{1}{\sqrt{\mathbf{1} + \frac{\Delta \boldsymbol{z} \Delta \boldsymbol{s}}{\mu}}} \left(- \frac{\Delta \boldsymbol{z} \Delta \boldsymbol{s}}{\mu} \right) \right\| = \sqrt{\sum_{i=1}^N \frac{1}{1 + \frac{(\Delta \boldsymbol{z})_i (\Delta \boldsymbol{s})_i}{\mu}} \cdot \frac{(\Delta \boldsymbol{z})_i^2 (\Delta \boldsymbol{s})_i^2}{\mu^2}} \\
&\leqslant \sqrt{\sum_{i=1}^N \frac{1}{1 - \left\| \frac{\Delta \boldsymbol{z} \Delta \boldsymbol{s}}{\mu} \right\|_\infty} \cdot \frac{(\Delta \boldsymbol{z})_i^2 (\Delta \boldsymbol{s})_i^2}{\mu^2}} = \sqrt{\frac{1}{1 - \left\| \frac{\Delta \boldsymbol{z} \Delta \boldsymbol{s}}{\mu} \right\|_\infty}} \cdot \sqrt{\sum_{i=1}^N \frac{(\Delta \boldsymbol{z})_i^2 (\Delta \boldsymbol{s})_i^2}{\mu^2}} \\
&= \sqrt{\frac{1}{1 - \left\| \frac{\Delta \boldsymbol{z} \Delta \boldsymbol{s}}{\mu} \right\|_\infty}} \cdot \left\| \frac{\Delta \boldsymbol{z} \Delta \boldsymbol{s}}{\mu} \right\| = \sqrt{\frac{1}{1 - \frac{1}{\mu} \|\Delta \boldsymbol{z} \Delta \boldsymbol{s}\|_\infty}} \cdot \frac{1}{\mu} \|\Delta \boldsymbol{z} \Delta \boldsymbol{s}\| \\
&\leqslant \sqrt{\frac{1}{1 - \frac{1}{\mu} \|\Delta \boldsymbol{z} \Delta \boldsymbol{s}\|_\infty}} \cdot \frac{1}{\mu} \cdot \sqrt{2} \mu \delta^2 \leqslant \sqrt{\frac{1}{1 - \frac{1}{\mu} \cdot \mu \delta^2}} \cdot \sqrt{2} \delta^2 = \frac{\sqrt{2} \delta^2}{\sqrt{1 - \delta^2}}.
\end{aligned}$$

Odtud $\delta^+ \leqslant \frac{\delta^2}{\sqrt{2(1-\delta^2)}}$. Všimněme si, že odhad je korektní, protože $1 - \left\| \frac{\Delta \boldsymbol{z} \Delta \boldsymbol{s}}{\mu} \right\|_\infty > 0$ — z lemmatu 32 totiž platí

$$1 - \left\| \frac{\Delta \boldsymbol{z} \Delta \boldsymbol{s}}{\mu} \right\|_\infty = 1 - \frac{1}{\mu} \|\Delta \boldsymbol{z} \Delta \boldsymbol{s}\|_\infty \geqslant 1 - \frac{1}{\mu} \cdot \mu \delta^2 = 1 - \delta^2 > 0,$$

ježto dle předpokladu jest $\delta < 1$. □

10 Existence centrální cesty

Věta 33. *Pro každé $\mu > 0$ existuje nanejvýš jedno $\boldsymbol{z}$ splňující $\boldsymbol{z}\boldsymbol{s}(\boldsymbol{z}) = \mu \mathbf{1}$, $\boldsymbol{z} \geqslant \mathbf{0}$ a $\boldsymbol{s}(\boldsymbol{z}) \geqslant \mathbf{0}$.*

Důkaz. Nechť existují dvě různá řešení $\boldsymbol{z}_1$ a $\boldsymbol{z}_2$. Protože $\mu > 0$, platí

$$\boldsymbol{z}_1 > \mathbf{0}, \quad \boldsymbol{z}_2 > \mathbf{0}, \quad \boldsymbol{s}(\boldsymbol{z}_1) > \mathbf{0}, \quad \boldsymbol{s}(\boldsymbol{z}_2) > \mathbf{0}.$$

Označme $\Delta \boldsymbol{z} := \boldsymbol{z}_2 - \boldsymbol{z}_1$ a $\Delta \boldsymbol{s} := \boldsymbol{s}(\boldsymbol{z}_2) - \boldsymbol{s}(\boldsymbol{z}_1)$. Platí

$$\boldsymbol{D} \Delta \boldsymbol{z} = \Delta \boldsymbol{s}, \tag{18}$$

protože

$$\boldsymbol{D} \Delta \boldsymbol{z} = \boldsymbol{D} \boldsymbol{z}_2 - \boldsymbol{D} \boldsymbol{z}_1 = \boldsymbol{D} \boldsymbol{z}_2 + \boldsymbol{q} - (\boldsymbol{D} \boldsymbol{z}_1 + \boldsymbol{q}) = \boldsymbol{s}(\boldsymbol{z}_2) - \boldsymbol{s}(\boldsymbol{z}_1) = \Delta \boldsymbol{s}.$$

Dále platí

$$\boldsymbol{z}_1 \Delta \boldsymbol{s} + \boldsymbol{s}(\boldsymbol{z}_1) \Delta \boldsymbol{z} + \Delta \boldsymbol{z} \Delta \boldsymbol{s} = \mathbf{0}, \tag{19}$$

protože

$$\begin{aligned}
&\boldsymbol{z}_1 \Delta \boldsymbol{s} + \boldsymbol{s}(\boldsymbol{z}_1) \Delta \boldsymbol{z} + \Delta \boldsymbol{z} \Delta \boldsymbol{s} \\
&\quad = \boldsymbol{z}_1 (\boldsymbol{s}(\boldsymbol{z}_2) - \boldsymbol{s}(\boldsymbol{z}_1)) + \boldsymbol{s}(\boldsymbol{z}_1)(\boldsymbol{z}_2 - \boldsymbol{z}_1) + (\boldsymbol{z}_2 - \boldsymbol{z}_1)(\boldsymbol{s}(\boldsymbol{z}_2) - \boldsymbol{s}(\boldsymbol{z}_1)) \\
&\quad = \boldsymbol{z}_1 \boldsymbol{s}(\boldsymbol{z}_2) - \boldsymbol{z}_1 \boldsymbol{s}(\boldsymbol{z}_1) + \boldsymbol{s}(\boldsymbol{z}_1) \boldsymbol{z}_2 - \boldsymbol{s}(\boldsymbol{z}_1) \boldsymbol{z}_1 + \boldsymbol{z}_2 \boldsymbol{s}(\boldsymbol{z}_2) - \boldsymbol{z}_2 \boldsymbol{s}(\boldsymbol{z}_1) - \boldsymbol{z}_1 \boldsymbol{s}(\boldsymbol{z}_2) + \boldsymbol{z}_1 \boldsymbol{s}(\boldsymbol{z}_1) \\
&\quad = \boldsymbol{s}(\boldsymbol{z}_2) \boldsymbol{z}_2 - \boldsymbol{z}_1 \boldsymbol{s}(\boldsymbol{z}_1) = \mu \mathbf{1} - \mu \mathbf{1} = \mathbf{0}.
\end{aligned}$$

Snadno se vidí, že

$$\Delta \boldsymbol{z}^{\mathrm{T}} \Delta \boldsymbol{s} = 0, \tag{20}$$

protože (s užitím ortogonálního lemmatu)

$$\Delta \boldsymbol{z}^{\mathrm{T}} \Delta \boldsymbol{s} \overset{(18)}{=} \Delta \boldsymbol{z}^{\mathrm{T}} \boldsymbol{D} \Delta \boldsymbol{z} = 0.$$

Z (19) plyne

$$\boldsymbol{0} = \boldsymbol{z}_1 \Delta \boldsymbol{s} + \boldsymbol{s}(\boldsymbol{z}_1) \Delta \boldsymbol{z} + \Delta \boldsymbol{z} \Delta \boldsymbol{s} = (\boldsymbol{z}_1 + \Delta \boldsymbol{z}) \Delta \boldsymbol{s} + \boldsymbol{s}(\boldsymbol{z}_1) \Delta \boldsymbol{z} = \underbrace{\boldsymbol{z}_2}_{>\boldsymbol{0}} \Delta \boldsymbol{s} + \underbrace{\boldsymbol{s}(\boldsymbol{z}_1)}_{>\boldsymbol{0}} \Delta \boldsymbol{z};$$

odtud plyne, že pro každé i musí být buďto $(\Delta \boldsymbol{z})_i \geqslant 0 \;\;\&\;\; (\Delta \boldsymbol{s})_i \leqslant 0$ anebo $(\Delta \boldsymbol{z})_i \leqslant 0 \;\;\&\;\; (\Delta \boldsymbol{s})_i \geqslant 0$. Tedy

$$\Delta \boldsymbol{z} \Delta \boldsymbol{s} \leqslant \boldsymbol{0}.$$

Ovšem (18) je ekvivalentní s $\boldsymbol{1}^{\mathrm{T}}(\Delta \boldsymbol{z} \Delta \boldsymbol{s}) = 0$, a tedy $\Delta \boldsymbol{z} \Delta \boldsymbol{s} = \boldsymbol{0}$. Proto pro každé i platí

$$(\Delta \boldsymbol{z})_i = 0 \vee (\Delta \boldsymbol{s})_i = 0.$$

Z (19) dále plyne

$$\underbrace{\boldsymbol{z}_1}_{>\boldsymbol{0}} \Delta \boldsymbol{s} + \underbrace{\boldsymbol{s}(\boldsymbol{z}_1)}_{>\boldsymbol{0}} \Delta \boldsymbol{z} = \boldsymbol{0},$$

a tedy $\Delta \boldsymbol{z} = \boldsymbol{0}$ a $\Delta \boldsymbol{s} = \boldsymbol{0}$, odkud $\boldsymbol{z}_1 = \boldsymbol{z}_2$. □

Nyní víme, že jestliže centrální cesta existuje, pak je jednoznačná. Zbývá ukázat existenci. Ukážeme, že bod na centrální cestě se nechá získat jako limita posloupnosti newtonovských kroků.

Připomeňme dvě základní věty z matematické analýzy:

- množina $\Omega \subseteq \mathbb{R}^n$ je kompaktní, právě když Ω je omezená a uzavřená (pro účely tohoto textu to můžeme vzít za *definici* kompaktnosti);
- jestliže pro posloupnost $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots$ platí $(\forall i)[\boldsymbol{x}_i \in \Omega]$, kde Ω je kompaktní, pak z posloupnosti $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots$ lze vybrat konvergentní podposloupnost. To jest: existuje posloupnost přirozených čísel $i_0 < i_1 < \cdots$ taková, že posloupnost $\boldsymbol{x}_{i_0}, \boldsymbol{x}_{i_1}, \ldots$ má limitu. Navíc platí $\boldsymbol{x}^* := \lim_{k \longrightarrow \infty} \boldsymbol{x}_{i_k} \in \Omega$. Bod $\boldsymbol{x}^*$ se nazývá *hromadný bod* posloupnosti $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots$.

Víme, že μ-střed existuje pro $\mu = 1$ (je jím $\boldsymbol{z}(1) = \boldsymbol{1}$). Položme $\mu_0 := 1$. Snadno se ukáže (dokažte jako cvičení), že pro $\mu > 0$ platí

$$\delta(\boldsymbol{z}(\mu_0), \mu) \leqslant \frac{1}{\sqrt{2}},$$

právě když[2]

$$\frac{1}{\gamma} \leqslant \frac{\mu}{\mu_0} \leqslant \gamma,$$

kde $\gamma := 1 + \frac{1}{N} + \sqrt{\frac{1}{N^2} + \frac{2}{N}}$. Podle důsledku 21 pro posloupnost newtonovských kroků

$$\boldsymbol{z}^0 := \boldsymbol{z}(\mu_0), \quad \boldsymbol{z}^1 := \boldsymbol{z}_0^+, \quad \boldsymbol{z}^2 := \boldsymbol{z}_1^+, \quad \ldots$$

[2] Pro varianční vektor platí $\boldsymbol{v} = \sqrt{\frac{\boldsymbol{z}(\mu_0)\boldsymbol{s}(\mu_0)}{\mu}} = \sqrt{\frac{\mu_0 \boldsymbol{1}}{\mu}} = \sqrt{\frac{\mu_0}{\mu}}\boldsymbol{1}$ (z definice μ-středu), a tedy $\delta(\boldsymbol{z}(\mu_0), \mu) \leqslant \frac{1}{\sqrt{2}}$, právě když $\frac{1}{2}\left|\sqrt{\frac{\mu_0}{\mu}} - \sqrt{\frac{\mu}{\mu_0}}\right| \cdot \|\boldsymbol{1}\| \leqslant \frac{1}{\sqrt{2}}$, a odtud se tvrzení dokáže snadno.

s $\mu \in [\frac{1}{\gamma}, \gamma]$ platí

$$\lim_{k\longrightarrow\infty} \delta(\boldsymbol{z}^k, \mu) = 0.$$

Tato posloupnost má alespoň jeden hromadný bod (podle lemmatu 18 totiž platí

$$\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}^1 = \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}^2 = \cdots = N\mu$$

a podle lemmatu 10(b) leží posloupnost $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots$ v kompaktní množině

$$\{\boldsymbol{z}: \ \mathbf{1}^{\mathrm{T}}\boldsymbol{z} + \mathbf{1}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) = N + N\mu, \ \boldsymbol{z} \geqslant \mathbf{0}, \ \boldsymbol{s}(\boldsymbol{z}) \geqslant \mathbf{0}\}$$

[jako cvičení dokažte, že jde skutečně o kompaktní množinu]). Tento hromadný bod označme $\boldsymbol{z}^*$. Protože $\delta(\boldsymbol{z}^*, \mu) = 0$, jest $\boldsymbol{z}^*$ bod na centrální cestě — je to tedy μ-střed. Tím jsme prokázali jeho existenci. Z předchozí věty víme, že μ-střed je jediný.

Dokázali jsme existenci μ-středů pro $\mu \in [\frac{1}{\gamma}, \gamma]$. Položíme-li nyní $\mu_0 := \frac{1}{\gamma}$ a $\mu_0 = \gamma$, stejným způsobem prokážeme existenci μ-středů pro $\mu \in [\frac{1}{\gamma^2}, \gamma^2]$. Protože $\gamma > 1$, jest $\lim_{k\longrightarrow\infty} \gamma^k = \infty$ a $\lim_{k\longrightarrow\infty} \frac{1}{\gamma^k} = 0$. Indukcí prokážeme existenci μ-středů pro všechna $\mu > 0$. Dokázali jsme větu 15(a).

Fakt, že centrální cesta je spojitá křivka (věta 15(b)), pro analýzu algoritmů nepotřebujeme; důkaz proto vynecháme. Dokonce se nechá ukázat, že centrální cesta je hladká.

11 Limita centrální cesty při $\mu \searrow 0$; analytický střed

Ukážeme, že centrální cesta má limitu (pro $\mu_k \searrow 0$) a touto limitou je striktně komplementární optimum. Tím dokážeme větu 15(a) a zároveň dokončíme důkaz věty 13 (připomeňme, že zbývá ukázat $B \cup \overline{B} = \{1, \ldots, N\}$).

Nejprve dokážeme, že lze vybrat posloupnost $\boldsymbol{z}(\mu_0), \boldsymbol{z}(\mu_1), \ldots$, jež konverguje ke striktně komplementárnímu optimu (krok 1). Pak ukážeme, že libovolná posloupnost μ-středů konverguje (při $\mu_k \searrow 0$) k jedinému striktně komplementárnímu optimu (krok 2); tento bod se prostě nazývá *limita centrální cesty* nebo také *analytický střed* optimálního prostoru.

Krok 1. Nechť $\mu_k \searrow 0$. Označme $\boldsymbol{z}_k := \boldsymbol{z}(\mu_k)$ a $\boldsymbol{s}_k := \boldsymbol{s}(\boldsymbol{z}(\mu_k))$. Posloupnost $\boldsymbol{z}_k$ leží v kompaktní množině $\{\boldsymbol{z} : \mathbf{1}^{\mathrm{T}}\boldsymbol{z} + \mathbf{1}^{\mathrm{T}}\boldsymbol{s}(\boldsymbol{z}) \leqslant N + \mu_0 N, \boldsymbol{z} \geqslant \mathbf{0}, \boldsymbol{s}(\boldsymbol{z}) \geqslant \mathbf{0}\}$, a má tedy alespoň jeden hromadný bod; označme jej $\boldsymbol{z}^*$. Označme dále $\boldsymbol{s}^* := \boldsymbol{s}(\boldsymbol{z}^*)$. Jistě je $\boldsymbol{z}^*$ optimální, protože

$$\lim_{\mu_k \searrow 0} \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}(\mu_k) \overset{\text{lemma 16}}{=} \lim_{\mu_k \searrow 0} \mu_k N = 0,$$

odkud $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}^* = 0$.

Prokažme, že $\boldsymbol{z}^*$ je striktně komplementární (tj. že platí $\boldsymbol{z}^* + \boldsymbol{s}^* > \mathbf{0}$). Bude se hodit jeden pojem.

Definice 34. *Symbolem* supp$(\boldsymbol{x}) := \{i : (\boldsymbol{x})_i > 0\}$ *označujeme* ***nosič (support)*** *nezáporného vektoru* $\boldsymbol{x}$. $\square$

Z ortogonálního lemmatu víme, že

$$\begin{aligned}
0 &= (\boldsymbol{z}_k - \boldsymbol{z}^*)^{\mathrm{T}} \boldsymbol{D}(\boldsymbol{z}_k - \boldsymbol{z}^*) \\
&= (\boldsymbol{z}_k - \boldsymbol{z}^*)^{\mathrm{T}} (\boldsymbol{D}\boldsymbol{z}_k - \boldsymbol{D}\boldsymbol{z}^*) \\
&= (\boldsymbol{z}_k - \boldsymbol{z}^*)^{\mathrm{T}} (\boldsymbol{D}\boldsymbol{z}_k + \boldsymbol{q} - (\boldsymbol{D}\boldsymbol{z}^* - \boldsymbol{q})) \\
&= (\boldsymbol{z}_k - \boldsymbol{z}^*)^{\mathrm{T}} (\boldsymbol{s}_k - \boldsymbol{s}^*) \\
&= \underbrace{\boldsymbol{z}_k^{\mathrm{T}} \boldsymbol{s}_k}_{=\boldsymbol{q}^{\mathrm{T}} \boldsymbol{z}_k = \mu_k N} - \boldsymbol{z}_k^{\mathrm{T}} \boldsymbol{s}^* - (\boldsymbol{z}^*)^{\mathrm{T}} \boldsymbol{s}_k + \underbrace{(\boldsymbol{z}^*)^{\mathrm{T}} \boldsymbol{s}^*}_{=\boldsymbol{q}^{\mathrm{T}} \boldsymbol{z}^* = 0} \\
&= \mu_k N - (\boldsymbol{z}_k^{\mathrm{T}} \boldsymbol{s}^* + (\boldsymbol{z}^*)^{\mathrm{T}} \boldsymbol{s}_k) \\
&= \mu_k N - \left(\sum_{i=1}^{N} (\boldsymbol{z}_k)_i \cdot (\boldsymbol{s}^*)_i + \sum_{i=1}^{N} (\boldsymbol{z}^*)_i \cdot (\boldsymbol{s}_k)_i \right) \\
&= \mu_k N - \left(\sum_{i \in \mathit{supp}(\boldsymbol{s}^*)} (\boldsymbol{z}_k)_i \cdot (\boldsymbol{s}^*)_i + \sum_{i \in \mathit{supp}(\boldsymbol{z}^*)} (\boldsymbol{z}^*)_i \cdot (\boldsymbol{s}_k)_i \right).
\end{aligned}$$

Odtud

$$\sum_{i \in \mathit{supp}(\boldsymbol{s}^*)} (\boldsymbol{z}_k)_i \cdot (\boldsymbol{s}^*)_i + \sum_{i \in \mathit{supp}(\boldsymbol{z}^*)} (\boldsymbol{z}^*)_i \cdot (\boldsymbol{s}_k)_i = \mu_k N,$$

a také (vydělením obou stran μ_k)

$$\sum_{i \in \mathit{supp}(\boldsymbol{s}^*)} \frac{(\boldsymbol{z}_k)_i}{\mu_k} \cdot (\boldsymbol{s}^*)_i + \sum_{i \in \mathit{supp}(\boldsymbol{z}^*)} (\boldsymbol{z}^*)_i \cdot \frac{(\boldsymbol{s}_k)_i}{\mu_k} = N.$$

Víme, že $\boldsymbol{z}_k \boldsymbol{s}_k = \mu_k \mathbf{1}$ (protože $\boldsymbol{z}_k$ je μ_k-střed); odtud $(\boldsymbol{z}_k)_i (\boldsymbol{s}_k)_i = \mu_k$, a tedy $\frac{(\boldsymbol{z}_k)_i}{\mu_k} = \frac{1}{(\boldsymbol{s}_k)_i}$ a $\frac{(\boldsymbol{s}_k)_i}{\mu_k} = \frac{1}{(\boldsymbol{z}_k)_i}$. Můžeme tedy psát

$$\sum_{i \in \mathit{supp}(\boldsymbol{s}^*)} \frac{(\boldsymbol{s}^*)_i}{(\boldsymbol{s}_k)_i} + \sum_{i \in \mathit{supp}(\boldsymbol{z}^*)} \frac{(\boldsymbol{z}^*)_i}{(\boldsymbol{z}_k)_i} = N.$$

V limitě $k \longrightarrow \infty$ jest $\frac{(\boldsymbol{s}^*)_i}{(\boldsymbol{s}_k)_i} \longrightarrow 1$ a $\frac{(\boldsymbol{z}^*)_i}{(\boldsymbol{z}_k)_i} \longrightarrow 1$; tedy

$$\sum_{i \in \mathit{supp}(\boldsymbol{s}^*)} \frac{(\boldsymbol{s}^*)_i}{(\boldsymbol{s}_k)_i} \longrightarrow |\mathit{supp}(\boldsymbol{s}^*)|, \qquad \sum_{i \in \mathit{supp}(\boldsymbol{z}^*)} \frac{(\boldsymbol{z}^*)_i}{(\boldsymbol{z}_k)_i} \longrightarrow |\mathit{supp}(\boldsymbol{z}^*)|.$$

Zjistili jsme, že $|\mathit{supp}(\boldsymbol{s}^*)| + |\mathit{supp}(\boldsymbol{z}^*)| = N$. To znamená, že $\boldsymbol{s}^*$ a $\boldsymbol{z}^*$ jsou striktně komplementární vektory (víme totiž, že pro každé i jest buďto $(\boldsymbol{z}^*)_i = 0$ anebo $(\boldsymbol{s}^*)_i = 0$ [ježto $\boldsymbol{z}^* \boldsymbol{s}^* = \mathbf{0}$]; pro každé i tedy platí, že jedno z čísel $(\boldsymbol{z}^*)_i$ a $(\boldsymbol{s}^*)_i$ musí být kladné, pročež platí $\boldsymbol{z}^* + \boldsymbol{s}^* > \mathbf{0}$).

Definice 35. *Nechť polyedr $\mathcal{F} := \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geqslant \mathbf{0}\}$ je neprázdný a omezený.*

(a) ***Nosič (support)*** *polyedru $\mathcal{F}$ je indexová množina $\mathit{supp}(\mathcal{F}) = \{i : (\exists \boldsymbol{x} \in \mathcal{F})(\boldsymbol{x})_i > 0\}$.*

(b) $\Phi(\boldsymbol{x}) := \prod_{i \in \mathit{supp}(\mathcal{F})} (\boldsymbol{x})_i$.

(c) ***Relativní vnitřek*** *polyedru $\mathcal{F}$ je množina $\mathit{int}(\mathcal{F}) := \{\boldsymbol{x} \in \mathcal{F} : (\forall i \in \mathit{supp}(\mathcal{F}))(\boldsymbol{x})_i > 0\}$.*

(d) ***Analytický střed*** *polyedru $\mathcal{F}$ je bod*

$$AC(\mathcal{F}) := \begin{cases} \mathbf{0}, & \text{je-li } supp(\mathcal{F}) = \emptyset, \\ \operatorname{argmax}_{\boldsymbol{x}\in\mathcal{F}} \Phi(\boldsymbol{x}), & \text{je-li } supp(\mathcal{F}) \neq \emptyset. \end{cases} \quad \square$$

Nás bude zajímat analytický střed polyedru

$$\mathcal{F} := \{(\boldsymbol{z},\boldsymbol{s}) : \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = 0,\ \boldsymbol{D}\boldsymbol{z} - \boldsymbol{s} = -\boldsymbol{q}, \boldsymbol{z} \geqslant \mathbf{0}, \boldsymbol{s} \geqslant \mathbf{0}\};$$

všimněme si, že $\mathcal{F}$ je právě optimální prostor naší úlohy. Z lemmatu 10(b) víme, že

$$\mathcal{F} \subseteq \{(\boldsymbol{z},\boldsymbol{s}) : \mathbf{1}^{\mathrm{T}}\boldsymbol{z} + \mathbf{1}^{\mathrm{T}}\boldsymbol{s} = N,\ \boldsymbol{z} \geqslant \mathbf{0},\ \boldsymbol{s} \geqslant \mathbf{0}\};$$

to znamená, že polyedr $\mathcal{F}$ je omezený. Jistě je neprázdný, protože $\mathbf{0} \in \mathcal{F}$. Tedy, jeho analytický střed je dobře definován. Všimněme si:

(i) z konvexity $\mathcal{F}$ plyne, že existuje $\boldsymbol{x}_0 \in \mathcal{F}$ takové, že $(\forall i \in supp(\mathcal{F}))(\boldsymbol{x}_0)_i > 0$ (jako cvičení podejte podrobný důkaz);

(ii) na množině $int(\mathcal{F})$ je funkce $\Phi(\boldsymbol{x})$ ryze konkávní (dokažte jako cvičení; návod: zkoumejte funkci $\ln \Phi(\boldsymbol{x})$).

Z bodu (ii) plyne, že analytický střed je jednoznačný.

Ukážeme, že limitou centrální cesty je právě analytický střed optimálního prostoru $\mathcal{F}$. Nebo jinak: pro libovolnou posloupnost $\mu_k \searrow 0$ platí, že $\lim \boldsymbol{z}(\mu_k) = AC(\mathcal{F})$. Z toho také plyne jednoznačnost limity; hovoříme pak prostě o *limitě centrální cesty.*

Krok 2. Nechť $\mu'_k \searrow 0$ je libovolná a $\boldsymbol{z}^*$ je libovolný hromadný bod posloupnosti $\boldsymbol{z}(\mu'_k)$. Označme $\boldsymbol{s}^* := \boldsymbol{s}(\boldsymbol{z}^*)$. Z kroku 1 víme, že $\boldsymbol{z}^*$ je striktně komplementární optimum; proto $\boldsymbol{z}^*\boldsymbol{s}^* = \mathbf{0}$ a $\boldsymbol{z}^* + \boldsymbol{s}^* > \mathbf{0}$. Nechť $\mu_k \searrow 0$ je posloupnost, pro kterou platí $\boldsymbol{z}(\mu_k) \longrightarrow \boldsymbol{z}^*$; označme $\boldsymbol{s}_k := \boldsymbol{s}(\boldsymbol{z}_k)$. Ukážeme, že $\boldsymbol{z}^*$ je analytickým středem optimálního prostoru $\mathcal{F}$. Protože analytický střed je jednoznačný, je i limita centrální cesty jednoznačná (tedy: má jen jediný hromadný bod).

Nechť $\widehat{\boldsymbol{z}}$ je libovolné optimum. Označme $\widehat{\boldsymbol{s}} := \boldsymbol{s}(\widehat{\boldsymbol{z}})$. Z ortogonálního lemmatu plyne

$$\begin{aligned}
0 &= (\boldsymbol{z}_k - \widehat{\boldsymbol{z}})^{\mathrm{T}}\boldsymbol{D}(\boldsymbol{z}_k - \widehat{\boldsymbol{z}}) \\
&= (\boldsymbol{z}_k - \widehat{\boldsymbol{z}})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{z}_k - \boldsymbol{D}\widehat{\boldsymbol{z}}) \\
&= (\boldsymbol{z}_k - \widehat{\boldsymbol{z}})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{z}_k + \boldsymbol{q} - (\boldsymbol{D}\widehat{\boldsymbol{z}} + \boldsymbol{q})) \\
&= (\boldsymbol{z}_k - \widehat{\boldsymbol{z}})^{\mathrm{T}}(\boldsymbol{s}_k - \widehat{\boldsymbol{s}}) \\
&= \underbrace{\boldsymbol{z}_k^{\mathrm{T}}\boldsymbol{s}_k}_{=\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}_k=\mu_k N} - \boldsymbol{z}_k^{\mathrm{T}}\widehat{\boldsymbol{s}} - \widehat{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{s}_k + \underbrace{\widehat{\boldsymbol{z}}^{\mathrm{T}}\widehat{\boldsymbol{s}}}_{=\boldsymbol{q}^{\mathrm{T}}\widehat{\boldsymbol{z}}=0} \\
&= \mu_k N - (\boldsymbol{z}_k^{\mathrm{T}}\widehat{\boldsymbol{s}} + \widehat{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{s}_k) \\
&= \mu_k N - \left(\sum_{i=1}^{N}(\boldsymbol{z}_k)_i(\widehat{\boldsymbol{s}})_i + \sum_{i=1}^{N}(\widehat{\boldsymbol{z}})_i(\boldsymbol{s}_k)_i\right) \\
&= \mu_k N - \left(\sum_{i\in supp(\widehat{\boldsymbol{s}})}(\boldsymbol{z}_k)_i(\widehat{\boldsymbol{s}})_i + \sum_{i\in supp(\widehat{\boldsymbol{z}})}(\widehat{\boldsymbol{z}})_i(\boldsymbol{s}_k)_i\right) \\
&= \mu_k N - \left(\sum_{i\in supp(\boldsymbol{s}^*)}(\boldsymbol{z}_k)_i(\widehat{\boldsymbol{s}})_i + \sum_{i\in supp(\boldsymbol{z}^*)}(\widehat{\boldsymbol{z}})_i(\boldsymbol{s}_k)_i\right);
\end{aligned}$$

poslední rovnost plyne z toho, že $\boldsymbol{z}^*$ je striktně komplementární optimum (a proto $supp(\widehat{\boldsymbol{z}}) \subseteq supp(\boldsymbol{z}^*)$ a $supp(\widehat{\boldsymbol{s}}) \subseteq supp(\boldsymbol{s}^*)$). Můžeme tedy psát

$$\sum_{i \in supp(\boldsymbol{s}^*)} (\boldsymbol{z}_k)_i (\widehat{\boldsymbol{s}})_i + \sum_{i \in supp(\boldsymbol{z}^*)} (\widehat{\boldsymbol{z}})_i (\boldsymbol{s}_k)_i = \mu_k N,$$

a také

$$\sum_{i \in supp(\boldsymbol{s}^*)} \frac{(\boldsymbol{z}_k)_i}{\mu_k} (\widehat{\boldsymbol{s}})_i + \sum_{i \in supp(\boldsymbol{z}^*)} (\widehat{\boldsymbol{z}})_i \frac{(\boldsymbol{s}_k)_i}{\mu_k} = N.$$

Protože $\boldsymbol{z}_k \boldsymbol{s}_k = \mu_k \mathbf{1}$, víme, že $\frac{(\boldsymbol{z}_k)_i}{\mu_k} = \frac{1}{(\boldsymbol{s}_k)_i}$ a $\frac{(\boldsymbol{s}_k)_i}{\mu_k} = \frac{1}{(\boldsymbol{z}_k)_i}$. Proto

$$\sum_{i \in supp(\boldsymbol{s}^*)} \frac{(\widehat{\boldsymbol{s}})_i}{(\boldsymbol{s}_k)_i} + \sum_{i \in supp(\boldsymbol{z}^*)} \frac{(\widehat{\boldsymbol{z}})_i}{(\boldsymbol{z}_k)_i} = N.$$

V limitě $k \longrightarrow \infty$ obdržíme

$$\sum_{i \in supp(\boldsymbol{s}^*)} \frac{(\widehat{\boldsymbol{s}})_i}{(\boldsymbol{s}^*)_i} + \sum_{i \in supp(\boldsymbol{z}^*)} \frac{(\widehat{\boldsymbol{z}})_i}{(\boldsymbol{z}^*)_i} = N.$$

Jako cvičení dokažte známou nerovnost „*geometrický průměr* $\leqslant$ *aritmetický průměr*" (tedy: pro libovolná nezáporná čísla $x_1, \ldots, x_t$ platí $\left(\prod_{i=1}^t x_i\right)^{1/t} \leqslant \frac{1}{t} \sum_{i=1}^t x_i$). Tato nerovnost říká

$$\left(\prod_{i \in supp(\boldsymbol{s}^*)} \frac{(\widehat{\boldsymbol{s}})_i}{(\boldsymbol{s}^*)_i} \cdot \prod_{i \in supp(\boldsymbol{z}^*)} \frac{(\widehat{\boldsymbol{z}})_i}{(\boldsymbol{z}^*)_i} \right)^{1/N} \leqslant \underbrace{\frac{1}{N} \left(\sum_{i \in supp(\boldsymbol{s}^*)} \frac{(\widehat{\boldsymbol{s}})_i}{(\boldsymbol{s}^*)_i} + \sum_{i \in supp(\boldsymbol{z}^*)} \frac{(\widehat{\boldsymbol{z}})_i}{(\boldsymbol{z}^*)_i} \right)}_{=1};$$

tedy

$$\left(\prod_{i \in supp(\boldsymbol{s}^*)} \frac{(\widehat{\boldsymbol{s}})_i}{(\boldsymbol{s}^*)_i} \cdot \prod_{i \in supp(\boldsymbol{z}^*)} \frac{(\widehat{\boldsymbol{z}})_i}{(\boldsymbol{z}^*)_i} \right)^{1/N} \leqslant 1,$$

a také

$$\prod_{i \in supp(\boldsymbol{s}^*)} \frac{(\widehat{\boldsymbol{s}})_i}{(\boldsymbol{s}^*)_i} \cdot \prod_{i \in supp(\boldsymbol{z}^*)} \frac{(\widehat{\boldsymbol{z}})_i}{(\boldsymbol{z}^*)_i} \leqslant 1,$$

odkud

$$\prod_{i \in supp(\boldsymbol{s}^*)} (\widehat{\boldsymbol{s}})_i \cdot \prod_{i \in supp(\boldsymbol{z}^*)} (\widehat{\boldsymbol{z}})_i \leqslant \prod_{i \in supp(\boldsymbol{s}^*)} (\boldsymbol{s}^*)_i \cdot \prod_{i \in supp(\boldsymbol{z}^*)} (\boldsymbol{z}^*)_i,$$

nebo jinak řečeno

$$\Phi((\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{s}})) \leqslant \Phi((\boldsymbol{z}^*, \boldsymbol{s}^*))$$

(jako cvičení dokažte snadné pozorování, že $supp(\mathcal{F}) = supp(\boldsymbol{z}^*) \cup supp(\boldsymbol{s}^*)$). Protože $(\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{s}})$ je libovolné optimum, ukázali jsme, že $(\boldsymbol{z}^*, \boldsymbol{s}^*)$ maximalisuje funkci $\Phi((\boldsymbol{z}, \boldsymbol{s}))$; je to tedy analytický střed. Tím je důkaz dokončen.

12 Odhad na condition number

V této části dokážeme větu 29.

Lemma 36 (Hadamardova nerovnost)**.** *Nechť čtvercová matice $\boldsymbol{A}$ má sloupce $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. Pak platí $|\det \boldsymbol{A}| \leqslant \|\boldsymbol{a}_1\| \cdot \|\boldsymbol{a}_2\| \cdots \|\boldsymbol{a}_n\|$.*

Důkaz. $|\det \boldsymbol{A}|$ je objem rovnoběžnostěnu určeného hranami $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ (to je rovnoběžnostěn s vrcholy $\{\sum_{i=1}^n \alpha_i \boldsymbol{a}_i : (\forall i)\alpha_i \in \{0,1\}\}$). Pravá strana nerovnosti je objem kvádru se stranami $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. (Snadno se dokonce vidí, že Hadamardova nerovnost je splněna jako rovnost, právě když $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ jsou ortogonální.) $\square$

Lemma 37. *Nechť celočíselná matice $\boldsymbol{A}$ rozměru $m \times n$ má sloupce $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. Nechť existuje $\boldsymbol{x}_0 \in \mathcal{F} := \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geqslant \boldsymbol{0}\}$ splňující $\boldsymbol{x}_0 > \boldsymbol{0}$. Nechť polyedr $\mathcal{F}$ je omezený. Pak pro každé $i \in \{1, \ldots, m\}$ platí $\max\{(\boldsymbol{x})_i : \boldsymbol{x} \in \mathcal{F}\} \geqslant \frac{1}{\|\boldsymbol{a}_1\| \cdot \|\boldsymbol{a}_2\| \cdots \|\boldsymbol{a}_n\|}$.*

Důkaz. Jistě $\boldsymbol{a}_j \neq \boldsymbol{0}$ pro každé $1 \leqslant j \leqslant n$ (jinak by polyedr $\mathcal{F}$ nebyl omezený).

Nechť i je pevné. Označme $\mathcal{G} := \{\boldsymbol{x} \in \mathcal{F} : (\forall \boldsymbol{y} \in \mathcal{F})(\boldsymbol{y})_i \leqslant (\boldsymbol{x})_i\}$ a zvolme $\boldsymbol{x}^* \in \mathcal{G}$, které má co nejvíce nulových složek. Díky existenci $\boldsymbol{x}_0 \in \mathcal{F}$ splňujícího $\boldsymbol{x}_0 > \boldsymbol{0}$ víme, že $(\boldsymbol{x}^*)_i > 0$. Bez újmy na obecnosti můžeme předpokládat (můžeme totiž permutovat proměnné), že $\boldsymbol{x}^* = \binom{\boldsymbol{x}'}{\boldsymbol{0}}$, kde $(x'_1, \ldots, x'_J)^{\mathrm{T}} := \boldsymbol{x}' > \boldsymbol{0}$; pak $i \leqslant J$. Nechť $\boldsymbol{A}_J$ je matice obsahující prvních J sloupců matice $\boldsymbol{A}$. Z matice $\boldsymbol{A}_J$ můžeme vybrat regulární podmatici $\boldsymbol{A}'_J$ a z vektoru $\boldsymbol{b}$ můžeme vybrat podvektor $\boldsymbol{b}'$ takové, že systém rovnic

$$\boldsymbol{A}'_J \boldsymbol{x} = \boldsymbol{b}'$$

má jediné řešení, a tím je $\boldsymbol{x}'$. Nechť $\boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_J$ značí sloupce matice $\boldsymbol{A}'_J$. Z Cramerova pravidla jest

$$(\boldsymbol{x}^*)_i = x'_i = \frac{\det \boldsymbol{A}''_J}{\det \boldsymbol{A}'_J} \overset{(\star)}{\geqslant} \frac{1}{|\det \boldsymbol{A}'_J|} \overset{(\dagger)}{\geqslant} \frac{1}{\|\boldsymbol{a}'_1\| \cdot \|\boldsymbol{a}'_2\| \cdots \|\boldsymbol{a}'_J\|} \overset{(\ddagger)}{\geqslant} \frac{1}{\|\boldsymbol{a}_1\| \cdot \|\boldsymbol{a}_2\| \cdots \|\boldsymbol{a}_J\|} \overset{(\S)}{\geqslant} \frac{1}{\|\boldsymbol{a}_1\| \cdot \|\boldsymbol{a}_2\| \cdots \|\boldsymbol{a}_n\|},$$

kde matice $\det \boldsymbol{A}''_J$ vznikla z matice $\det \boldsymbol{A}'_J$ nahrazením i-tého sloupce vektorem $\boldsymbol{b}'$. Nerovnost $(\star)$ plyne z toho, že $(\boldsymbol{x}^*)_i > 0$; v čitateli je determinant celočíselné matice, a ten nemůže být menší než 1 (v absolutní hodnotě). Nerovnost $(\dagger)$ je nerovnost Hadamardova. Nerovnost $(\ddagger)$ plyne z toho, že vektor $\boldsymbol{a}'_j$ je podvektorem vektoru $\boldsymbol{a}_j$ (pro libovolné $1 \leqslant j \leqslant J$). Nerovnost $(\S)$ plyne z toho, že $\boldsymbol{a}_j$ (pro každé j) je nenulový celočíselný vektor (pročež má normu $\geqslant 1$). $\square$

Důkaz věty 29. Bez újmy na obecnosti řekněme, že rozklad indexové množiny $\{1, \ldots, N\}$ na B a $\overline{B}$ je tvaru $B = \{1, \ldots, k\}$ a $\overline{B} = \{k+1, \ldots, N\}$. Pak můžeme psát

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{z}_B \\ \boldsymbol{z}_{\overline{B}} \end{pmatrix}, \quad \boldsymbol{s} = \begin{pmatrix} \boldsymbol{s}_B \\ \boldsymbol{s}_{\overline{B}} \end{pmatrix}, \quad \boldsymbol{q} = \begin{pmatrix} \boldsymbol{q}_B \\ \boldsymbol{q}_{\overline{B}} \end{pmatrix},$$

kde jsme každý z uvedených vektorů rozdělili na složky indexované $1, \ldots, k$ a složky indexované $k+1, \ldots, N$. Analogicky můžeme rozčlenit i matici $\boldsymbol{D}$ na čtyři podmatice $\boldsymbol{D}_{BB}, \boldsymbol{D}_{B\overline{B}}, \boldsymbol{D}_{\overline{B}B}, \boldsymbol{D}_{\overline{BB}}$ rozměru postupně $k \times k$, $(N-k) \times k$, $k \times (N-k)$, $(N-k) \times (N-k)$:

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_{BB} & \boldsymbol{D}_{B\overline{B}} \\ \boldsymbol{D}_{\overline{B}B} & \boldsymbol{D}_{\overline{BB}} \end{pmatrix}.$$

Pak definiční rovnost slacku $\boldsymbol{s} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{q}$ můžeme psát ve tvaru

$$\begin{pmatrix} \boldsymbol{s}_B \\ \boldsymbol{s}_{\overline{B}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_{BB} & \boldsymbol{D}_{B\overline{B}} \\ \boldsymbol{D}_{\overline{B}B} & \boldsymbol{D}_{\overline{B}\overline{B}} \end{pmatrix} \begin{pmatrix} \boldsymbol{z}_B \\ \boldsymbol{z}_{\overline{B}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{q}_B \\ \boldsymbol{q}_{\overline{B}} \end{pmatrix}.$$

Víme, že řešení je optimální, právě když $\boldsymbol{z} \geqslant \boldsymbol{0}$, $\boldsymbol{s} \geqslant \boldsymbol{0}$ a $\boldsymbol{z}_{\overline{B}} = \boldsymbol{0}$ a $\boldsymbol{s}_B = 0$. Dosaďme:

$$\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{s}_{\overline{B}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_{BB} & \boldsymbol{D}_{B\overline{B}} \\ \boldsymbol{D}_{\overline{B}B} & \boldsymbol{D}_{\overline{B}\overline{B}} \end{pmatrix} \begin{pmatrix} \boldsymbol{z}_B \\ \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{q}_B \\ \boldsymbol{q}_{\overline{B}} \end{pmatrix}.$$

Navíc víme, že optimální hodnota účelové funkce je 0; odtud

$$0 = \boldsymbol{q}^{\mathrm{T}}\boldsymbol{z} = (\boldsymbol{q}_B^{\mathrm{T}}\ \boldsymbol{q}_{\overline{B}})^{\mathrm{T}} \begin{pmatrix} \boldsymbol{z}_B \\ \boldsymbol{z}_{\overline{B}} \end{pmatrix} = (\boldsymbol{q}_B^{\mathrm{T}}\ \boldsymbol{q}_{\overline{B}})^{\mathrm{T}} \begin{pmatrix} \boldsymbol{z}_B \\ \boldsymbol{0} \end{pmatrix} = \boldsymbol{q}_B^{\mathrm{T}}\boldsymbol{z}_B.$$

To platí pro každé optimální, a tedy i pro striktně komplementární optimální řešení. Takové má ovšem $\boldsymbol{z}_B > \boldsymbol{0}$; odtud je jasné, že z podmínky $\boldsymbol{q}_B^{\mathrm{T}}\boldsymbol{z}_B = 0$ plyne $\boldsymbol{q}_B = \boldsymbol{0}$. Můžeme tedy opět dosadit:

$$\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{s}_{\overline{B}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_{BB} & \boldsymbol{D}_{B\overline{B}} \\ \boldsymbol{D}_{\overline{B}B} & \boldsymbol{D}_{\overline{B}\overline{B}} \end{pmatrix} \begin{pmatrix} \boldsymbol{z}_B \\ \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{q}_{\overline{B}} \end{pmatrix}.$$

Obrželi jsme systém

$$\begin{aligned} \boldsymbol{D}_{BB}\boldsymbol{z}_B &= \boldsymbol{0}, \\ \boldsymbol{D}_{\overline{B}B}\boldsymbol{z}_B - \boldsymbol{s}_{\overline{B}} &= -\boldsymbol{q}_{\overline{B}}, \end{aligned}$$

který můžeme psát ve tvaru

$$\underbrace{\begin{pmatrix} \boldsymbol{D}_{BB} & \boldsymbol{0} \\ \boldsymbol{D}_{\overline{B}B} & -\boldsymbol{I} \end{pmatrix}}_{=:\boldsymbol{D}'} \begin{pmatrix} \boldsymbol{z}_B \\ \boldsymbol{s}_{\overline{B}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ -\boldsymbol{q}_{\overline{B}} \end{pmatrix}.$$

Tedy, optimální prostor tvoří právě řešení systému

$$\boldsymbol{D}' \begin{pmatrix} \boldsymbol{z}_B \\ \boldsymbol{s}_{\overline{B}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ -\boldsymbol{q}_{\overline{B}} \end{pmatrix}, \quad \boldsymbol{z}_B \geqslant \boldsymbol{0}, \quad \boldsymbol{s}_{\overline{B}} \geqslant \boldsymbol{0}, \tag{21}$$

kde stačí — pro úplnost — doplnit $\boldsymbol{z}_{\overline{B}} = \boldsymbol{0}$ a $\boldsymbol{s}_B = \boldsymbol{0}$.

Víme, že existuje striktně komplementární optimum; pro něj platí $\boldsymbol{z}_B > \boldsymbol{0}$ a $\boldsymbol{s}_{\overline{B}} > \boldsymbol{0}$, a tedy systém (21) má positivní bod. A také víme, že optimální prostor je omezený. Můžeme tedy na (21) použít lemma 37. Z něj plyne, že pro každé $1 \leqslant i \leqslant k$ platí

$$\begin{aligned} \max\{(\boldsymbol{z}_B)_i : (\boldsymbol{z}_B, \boldsymbol{s}_{\overline{B}}) \text{ je řešením (21)}\} &\geqslant \frac{1}{\|\boldsymbol{d}'_1\| \cdot \|\boldsymbol{d}'_2\| \cdots \|\boldsymbol{d}'_N\|} \\ &= \frac{1}{\|\boldsymbol{d}'_1\| \cdot \|\boldsymbol{d}'_2\| \cdots \|\boldsymbol{d}'_k\|} \geqslant \frac{1}{\|\boldsymbol{d}_1\| \cdot \|\boldsymbol{d}_2\| \cdots \|\boldsymbol{d}_N\|} \end{aligned}$$

a pro každé $1 \leqslant i \leqslant N - k$ platí

$$\begin{aligned} \max\{(\boldsymbol{s}_{\overline{B}})_i : (\boldsymbol{z}_B, \boldsymbol{s}_{\overline{B}}) \text{ je řešením (21)}\} &\geqslant \frac{1}{\|\boldsymbol{d}'_1\| \cdot \|\boldsymbol{d}'_2\| \cdots \|\boldsymbol{d}'_N\|} \\ &= \frac{1}{\|\boldsymbol{d}'_1\| \cdot \|\boldsymbol{d}'_2\| \cdots \|\boldsymbol{d}'_k\|} \geqslant \frac{1}{\|\boldsymbol{d}_1\| \cdot \|\boldsymbol{d}_2\| \cdots \|\boldsymbol{d}_N\|}, \end{aligned}$$

kde $\boldsymbol{d}'_1, \ldots, \boldsymbol{d}'_N$ jsme označili sloupce matice $\boldsymbol{D}'$ a $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_N$ jsme označili sloupce matice $\boldsymbol{D}$. Využili jsme toho, že

- norma $(k+1),\ldots,N$-tého sloupce $\boldsymbol{D}'$ je 1;
- $\boldsymbol{D}$ má celočíselné nenulové sloupce, pročež každý její sloupec má normu $\geqslant 1$.

Připomeňme definici condition number: $\Sigma = \min_i \Sigma_i$, kde

$$\Sigma_i = \max_{\boldsymbol{z}\in\boldsymbol{Q}} [(\boldsymbol{z})_i + (\boldsymbol{s}(\boldsymbol{z}))_i]$$

$$= \max_{(\boldsymbol{z}_B,\boldsymbol{s}_{\overline{B}}) \text{ splňuje (21)}} [(\boldsymbol{z}_B\ \boldsymbol{0})_i + (\boldsymbol{0}\ \boldsymbol{s}_{\overline{B}})_i] = \begin{cases} \max_{(\boldsymbol{z}_B,\boldsymbol{s}_{\overline{B}}) \text{ splňuje (21)}} (\boldsymbol{z}_B)_i, & \text{je-li } i \leqslant k, \\ \max_{(\boldsymbol{z}_B,\boldsymbol{s}_{\overline{B}}) \text{ splňuje (21)}} (\boldsymbol{s}_{\overline{B}})_{i-k}, & \text{je-li } i \geqslant k+1. \end{cases}$$

Pro oba případy máme odhad $\geqslant \frac{1}{\|\boldsymbol{d}_1\|\cdot\|\boldsymbol{d}_2\|\cdots\|\boldsymbol{d}_N\|}$. (Připomeňme ještě, že symbolem $\boldsymbol{Q}$ jsme v definici condition number označovali optimální prostor.) Jestliže pro každé i platí $\Sigma_i \geqslant \frac{1}{\|\boldsymbol{d}_1\|\cdot\|\boldsymbol{d}_2\|\cdots\|\boldsymbol{d}_N\|}$, pak také $\min_i \Sigma_i \geqslant \frac{1}{\|\boldsymbol{d}_1\|\cdot\|\boldsymbol{d}_2\|\cdots\|\boldsymbol{d}_N\|}$; odtud

$$\Sigma \geqslant \frac{1}{\|\boldsymbol{d}_1\|\cdot\|\boldsymbol{d}_2\|\cdots\|\boldsymbol{d}_N\|}. \quad \square$$