Bakalářská práce

Kvadratické rovnice na slovech

Miroslav Olšák

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

Katedra algebry
Vedoucí práce: doc. Mgr. Štěpán Holub, Ph.D.
Studijní program: ??
Studijní obor: ??
Praha 2013

Draft: 5. 7. 2013

Draft: 5. 7. 2013

Poděkování

Díky všem, kterým je za co děkovat.

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V Praze dne ??. ??. 2013

.....

Přehled

Název práce: Kvadratické rovnice na slovech
Autor: Miroslav Olšák
Katedra algebry
Vedoucí práce: doc. Mgr. Štěpán Holub, Ph.D.
Abstrakt: Kecy kecy blá blá blá, všechy nás to určitě hrozně zajímá.
Klíčová slova: ??

Summary

Title: Quadratic word equations
Author: Miroslav Olšák
Department of algebra (??)
Supervisor: doc. Mgr. Štěpán Holub, Ph.D.
Abstract: Blah, blah.
Keywords: ??

Obsah

Tabulky

Obrázky

[uvod]

Kapitola 1

Úvod

Práce se zabývá řešitelností kvadratických rovnic na slovech. Reprodukuje některé výsledky pánů Robsona a Diekerta [1] a přidává vlastní výzkum v oblasti jednoduše exponenciální meze na velikost řešení. Je totiž známé, že řešitelnost kvadratických rovnic je NP-těžká a byla by NP-úplná za jistého dodatečného předpokladu, o kterém se matematici domnívají jeho platnost nejen pro kvadratické, ale dokonce pro všechny rovnice na slovech – jednoduše exponenciální mez na velikost nejmenšího řešení. Navíc je v mateiálu [1] uvedena silnější hypotéza – polynomiální mez pro kvadratické rovnice na slovech.

V kapitole 2 zavádíme základní pojmy (kvadratických) rovnic na slovech. V kapitolách ??, 4 následují pokusy o pokoření hypotézy o jednoduše exponenciální meze. Sice nejsou zcela úspěšné, ale zužují třídu rovnic, kterým se stačí při dokazování jednoduše exponenciální meze věnovat.

Kapitola 5 pak vysvětluje analogii přístupů z předchozích dvou kapitol pomocí zavedení nového objektu – 2D rovnice.

V následující kapitole 6 jsou sepsány již známé výsledky ohledně otázky výpočetní složitosti. Nakonec kapitola 7 představuje pokus o vyvrácení polynomiální meze – ukazuje rovnici, jejíž nejmenší řešení sice je polynomiálně omezitelné (dokonce lineárně), ale takzvané l-minimální a c-minimální řešení již nikoli. V závěru kapitoly je tato rovnice vylepšena a podána proti-hypotéza.

[pojmy]

Kapitola 2

Základní pojmy a značení

Definice 2.1. *Slovem* σ nad danou abecedou $\mathscr{A}$ (konečnou množinou „písmen“) rozumíme konečnou posloupnost písmen z této abecedy. Symbolem $|\sigma|$ značíme délku slova σ a symbolem $\sigma[i]$ značíme $(i+1)$-té písmeno pro $0 \leq i \leq |\sigma| - 1$[1]. Je-li $|\sigma| = 0$, říkáme, že je toto slovo *prázdné*. Množinu všech slov nad danou abecedou $\mathscr{A}$ značíme $\mathscr{A}^*$.

Definice 2.2. Pro $i_1, i_2 \in \mathbb{Z}$ značíme množinou $\{i_1, \ldots, i_2\}$ množinu $\{i \in \mathbb{Z} \mid i_1 \leq i \leq i_2\}$, tedy specielně prázdnou množinu, je-li $i_1 > i_2$. Pro slovo σ a množinu indexů $M \subseteq \{0, \ldots, |\sigma| - 1\}$ značíme symbolem $\sigma\,|_M$ slovo vzniklé ze σ po odstranění všech písmen na pozicích mimo množinu M.

Definice 2.3. Nechť je dána množina proměnných $\mathscr{V}$ a množina konstant $\mathscr{C}$. *Rovnicí* rozumíme dvojici $\{S_1, S_2\}$, kde $S_1, S_2 \in (\mathscr{C} \cup \mathscr{V})^*$ jsou neprázdná slova. Rovnici zapisujeme $S_1 = S_2$, slova S_1, S_2 nazýváme (levá resp. pravá) strana rovnice. *Soustava* pak je konečná neprázdná množina rovnic. Rovnici současně chápeme jako jednoprvkovou soustavu, tedy soustava je obecnější pojem. Dále předpokládáme, že každá proměnná z $\mathscr{V}$ se v soustavě někde vyskytuje. Počet slov v soustavě $\mathcal{S}$ značíme $\langle\!\langle \mathcal{S} \rangle\!\rangle$.

Značení $\mathscr{V}$ a $\mathscr{C}$ pro množiny proměnných a konstant budeme používat v celém textu. Budeme-li hovořit o různých soustavách s různými příslušnými množinami, budeme množinu proměnných resp. konstant dané soustavy $\mathcal{S}$ značit $\mathscr{V}_\mathcal{S}$ resp. $\mathscr{C}_\mathcal{S}$.

Definice 2.4. *Délku rovnice* $\mathcal{R}$ definujeme jako aritmetický průměr délek obou stran a *délku soustavy* $\mathcal{S}$ jako součet délek všech jejích rovnic. Tyto délky značíme $|\mathcal{R}|, |\mathcal{S}|$. Jinými slovy je délka soustavy součtem všech délek všech stran podělená dvěma.[2]

Definice 2.5. Mějme soustavu a uvažujme zobrazení r z množiny proměnných do slov nad množinou konstant. Toto zobrazení (jednoznačným způsobem) rozšíříme na monoidový (s operací konkatenace) homomorfismus zachovávající konstanty – pro dané slovo σ nad abecedou $\mathscr{V} \cup \mathscr{C}$ definujeme $r(\sigma)$ tak, že každou proměnnou nahradíme jejím obrazem v zobrazení r. Řekneme, že r je řešením rovnice, pokud $r(S_1) = r(S_2)$. Řešením soustavy pak rozumíme takové r, které je řešením každé rovnice v soustavě. Soustavu $\mathcal{S}$ nazýváme *řesitelná*, pokud má nějaké řešení. Délkou proměnné x v řešení r rozumíme $|r(x)|$ a značíme ji $|x|_r$, stejně definujeme i obecně $|\sigma|_r = |r(\sigma)|$.

Definice 2.6. Vzhledem k tomu, že pro řešení r rovnice $S_1 = S_2$ platí $r(S_1) = r(S_2)$, platí i $|S_1|_r = |S_2|_r$. Tuto délku nazveme *délkou řešení* rovnice. Délkou řešení soustavy $\mathcal{S}$ rozumíme součet délek všech jejích rovnic, značíme ji $|r|_\mathcal{S}$. Nejmenší řešení soustavy je takové, že žádné jiné řešení nemá ostře menší délku. Pro řešitelné rovnice takovém případě definujeme $\min(\mathcal{S})$ jako délku nejmenšího řešení.

Příklad 2.7. Kvadratická rovnice

$$xABCy = yCBAx$$

délky 5, kde x, y jsou neznámé, A, B, C jsou konstanty, má dvě různá nejmenší řešení délky 7:

[1] Tedy indexujeme od nuly jako v programovacím jazyce C.

[2] Dělení dvěma se v tuto chvíli může jevit mírně nelogické, ostatně délka tak nemusí být celé číslo. V kvadratických rovnicích a následných 2-rovnicích, ale dostane tato délka smysl.

- $x = BCB, y = B$,
- $x = B, y = BAB$.

Definice 2.8. Na všech řešeních dané soustavy zavedeme *l-uspořádání* jako následující kvaziuspořádání: Píšeme $r_1 \leq_l r_2$, pokud pro každou proměnnou x soustavy platí $|x|_{r_1} \leq_l |x|_{r_2}$. Pak *l-minimální řešení* je minimální prvek v tomto kvaziuspořádání – tedy r je l-minimální právě když pro každé $r' \leq_l r$ platí $r \leq_l r'$.

Stejně definujeme i l-porovnávání řešení napříč různými soustavami – pro řešení r soustavy $\mathcal{S}$ a konstantu $x \notin \mathcal{V}_\mathcal{S}$ chápeme ve smyslu této definice $|x|_r$ jako 0 a kvantifikujeme přes sjednocení množin konstant obou rovnic.

Pozorování 2.9. Pod každým řešením je nějaké l-minimální a každé nejmenší řešení je l-minimální.

Definice 2.10. *Výskytem* $\boldsymbol{v}$ *proměnné nebo konstanty* $\alpha \in \mathcal{C} \cup \mathcal{V}$ *v soustavě* $\mathcal{S}$ rozumíme některou uspořádanou trojici $\boldsymbol{v} = (R, S, i)$, kde R je některá rovnice soustavy, S je některá strana R, i je přirozené číslo a $S(i) = \alpha$. Obdobně *pozicí* $\boldsymbol{V}$ *konstanty* A *v řešení* r soustavy rozumíme trojici $\boldsymbol{V} = (R, S, i)$, kde R je některá rovnice soustavy, S je některá strana R a $r(S)[i] = A$. Jak pro pozice v řešení, tak pro výskyty v soustavě nazýváme R rovnicí výskytu / pozice, S stranou rovnice výskytu / pozice a i indexem výskytu / pozice. Index výskytu / pozice značíme $\mathrm{Ind}(\boldsymbol{v})$. Symbolem $\overline{\boldsymbol{v}}$ značíme onu konstantu nebo proměnnou, které je $\boldsymbol{v}$ výskyt / pozice. Nakonec symbolem $\mathrm{Freq}_{r,\mathcal{S}}(A)$ značíme počet pozic konstanty A podělený dvěma a nazýváme jej frekvencí konstanty A.

Pozorování 2.11. Pro řešení r soustavy $\mathcal{S}$ platí

$$|r|_\mathcal{S} = \sum_{A \in \mathcal{C}} \mathrm{Freq}_{r,\mathcal{S}}(A)$$

Definice 2.12. Na všech řešeních dané soustavy $\mathcal{S}$ zavedeme *c-uspořádání* jako následující kvaziuspořádání. Píšeme $r_1 \leq_c r_2$, pokud pro každou konstanta A soustavy $\mathcal{S}$ platí nerovnost

$$\mathrm{Freq}_{r_1,\mathcal{S}}(A) \leq \mathrm{Freq}_{r_2,\mathcal{S}}(A)$$

Pak *c-minimální řešení* je minimální prvek v tomto kvaziuspořádání a samotné kvaziuspořádání budeme nazývat c-uspořádání.

Stejně definujeme i c-porovnávání řešení napříč různými soustavami – pro řešení r soustavy $\mathcal{S}$ a konstantu $A \notin \mathcal{C}_\mathcal{S}$ chápeme ve smyslu této definice $\mathrm{Freq}_{r,\mathcal{S}}(A)$ jako 0 a kvantifikujeme přes sjednocení množin konstant obou rovnic.

Pozorování 2.13. Pod každým řešením je nějaké c-minimální a každé nejmenší řešení je c-minimální.

Tvrzení 2.14. Každé l-minimální i každé c-minimální řešení je jednoznačně určené délkami proměnných.

Důkaz: Uvažujme dvě řešení r_1, r_2 dané soustavy. Předpokládejme, že pro každou proměnnou x je $l_{r_1}(x)) = l_{r_2}(x)$ (tedy totéž platí i pro slova nad $\mathcal{V} \cup \mathcal{C}$), ale $r_1 \neq r_2$. Definujeme, řešení r, které je l-ostře i c-ostře menší než r_1 (tedy i než r_2). Pro $\sigma \in (\mathcal{C} \cup \mathcal{V})^*$ je

$$r(\sigma) = r_1(\sigma)\,|_{\{0 \leq i \leq |\sigma|_{r_1} - 1 | r_1(\sigma)[i] = r_2(\sigma)[i]\}}$$

Nahlédneme, že r je opět monoindový homomorfismus, tedy je řešením. Navíc je řešením menším než r_1, protože délka obrazu r je vždy menší rovna obrazu r_1, navíc na rovnici, kde se r_1 liší od r_2 je ostře menší. ∎

Definice 2.15. Soustavu rovnic nazýváme *kvadratická*, pokud se v ní každá proměnná vyskytuje nejvýše dvakrát.

Konečně uvedeme hlavní problematiku, kterou se tato práce zabývá.

Tvrzení 2.16. Existuje polynom p takový, že pro libovolnou kvadratickou soustavu $\mathcal{S}$ má každé l-minimální řešení délku menší než $2^{2^{p|\mathcal{S}|}}$. Toto omezení na délku nejmenšího řešení nazýváme dvojitě exponenciální.

Důkaz tohoto tvrzení není složitý. Jeho důkaz je uveden v kapitole 4.3.

Tvrzení 2.17. Existuje nekonečný systém řešitelných kvadratických soustav, jejichž nejmenší řešení je kvadratické (v klasickém smyslu) v závislosti na délce.

Důkaz: Pro $n \in \mathbb{N}$ volíme proměnné $x_1, \ldots, x_n$, konstanty $A_1, \ldots, A_n$ a soustavu

$$A_1 A_2 \ldots A_n = x_1, \quad x_1 = x_2, \quad x_2 = x_3, \quad \ldots, \quad x_{n-1} = x_n$$

■

Poznámka 2.18. Příklad dokonce systému rovnic (a ne soustav) je dán v kapitole 7.1, případně jiný je snadno odvoditelný z příkladu v kapitole 3.6.

Lepší odhady známy nejsou, tedy můžeme polemizovat.

[exp-mez]

Hypotéza 2.19. (slabší verze) Existuje polynom p takový, že pro libovolnou řešitelnou kvadratickou soustavu $\mathcal{S}$ má nejmenší řešení délku menší než $2^{p|\mathcal{S}|}$. Toto omezení na délku nejmenšího řešení nazýváme *jednoduše exponenciální*.

Tato skutečnost by stačila k tomu, aby bylo řešení kvadratických rovnic NP-úplné (viz kapitola 6. Nicméně absence protipříkladů umožňuje být ještě odvážnější.

Hypotéza 2.20. (silnější verze) Existuje polynom p takový, že pro libovolnou řešitelnou kvadratickou soustavu $\mathcal{S}$ má nejmenší řešení délku menší než $p|\mathcal{S}|$. Toto omezení na délku nejmenšího řešení nazýváme *polynomiální*.

Kapitola 3

Mikro-operace kvadratickou rovnicí aneb skákání po pozicích

3.1 Pojmy a skoky

3.1.1 Základní pojmy ohledně výskytů a pozic

Definice 3.1. Uvažujme řešení r dané kvadratické soustavy $\mathcal{S}$. Označme $r[\mathcal{S}]$ množinu všech pozic všech konstant v řešení r. Podobně, je li R rovnice této soustavy a S_1 její strana, pak značíme $r[S_1] \subseteq r[R] \subseteq r[\mathcal{S}]$ pozice na příslušné straně resp. pozice v příslušné rovnici.

Definice 3.2. *Délkou výskytu* $\boldsymbol{v}$ *proměnné / konstanty v řešení* r rozumíme délku příslušné proměnné, tedy $|\boldsymbol{v}|_r = |\overline{\boldsymbol{v}}|_r$.

Definice 3.3. Na výskytech a pozicích zavedeme přičítání konstanty jako přičítání konstanty k indexu, tedy pro výskyt nebo pozici $\boldsymbol{v} = (R, S, i)$ a $k \in \mathbb{Z}$ definujeme $\boldsymbol{v} + k = (R, S, i + k)$, je-li $i + k$ v povolených mezích. V opačném případě $\boldsymbol{v} + k$ nedefinujeme.

Podobně definujeme porovnávání výskytů a pozic. Píšeme $\boldsymbol{v} < \boldsymbol{w}$, pokud jsou to oboje výskyty nebo pozice na jedné straně rovnice a $\boldsymbol{w}$ má větší index než $\boldsymbol{v}$.

Definice 3.4. Pro pozici $\boldsymbol{V} = (R, S, i)$ v řešení r definujeme zdroj $\mathrm{Src}(\boldsymbol{V})$ jako výskyt $\boldsymbol{v} = (R, S, j)$, který splňuje $\left|r\left(S\,|_{\{0,\dots,j-1\}}\right)\right| \le i \le \left|r\left(S\,|_{\{0,\dots,j\}}\right)\right|$. Intuitivně se jedná o tu proměnnou nebo konstantu, z které daná pozice v řešení „vzešla“.

Máme-li naopak dán výskyt $\boldsymbol{v}$, definujeme index v řešení r jako

$$\mathrm{Ind}_r(\boldsymbol{v}) = \left|r\left(S\,|_{\{0,\dots,\mathrm{Ind}(\boldsymbol{v})-1\}}\right)\right|$$

a dále je-li $|r(\overline{v})| \ge 1$ definujeme obraz výskytu $\boldsymbol{v}$ jako pozici

$$\mathrm{Img}_r(\boldsymbol{v}) = (R, S, \mathrm{Ind}_r(\boldsymbol{v})).$$

3.1.2 Jednoduché skoky

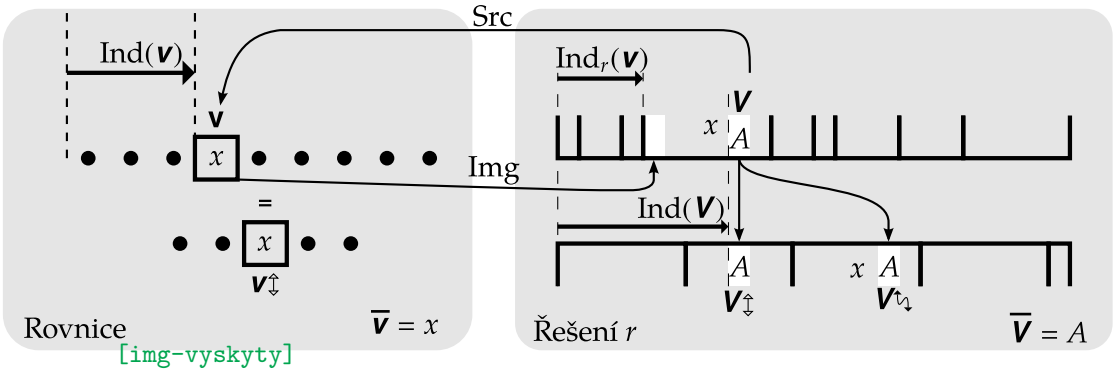
V celé sekci o sekcích budeme předpokládat pevnou soustavu $\mathcal{S}$ a její řešení r.

Definice 3.5. Na množině $r[\mathcal{S}]$ množině definujeme dvě zobrazení. Pro pozici $\boldsymbol{V} \in r[\mathcal{S}]$ definujeme $\boldsymbol{V}\updownarrow$ jako pozici mající stejný index i rovnici, avšak na druhé straně příslušné rovnice.

Dále vezměme $\alpha = \overline{\mathrm{Src}(\boldsymbol{V})}$ a předpokládejme, že α má v soustavě ještě právě jeden další výskyt $v' = (R', S', i)$. Pak definujme pozici

$$V{\scriptstyle\updownarrow} = (R', S', \mathrm{Ind}(\boldsymbol{v}) - \mathrm{Ind}_r(\mathrm{Src}(v)) + \mathrm{Ind}_r(\mathrm{Src}(v'))).$$

Podobně, je-li $\boldsymbol{v}$ výskyt proměnné nebo konstanty, která je v soustavě přesně dvakrát, značíme druhý takový výskyt $\boldsymbol{v}{\scriptstyle\updownarrow}$.



[img-vyskyty]

Obrázek 3.1. Značení okolo výskytů a pozic

Pozorování 3.6. Je-li $\boldsymbol{V}$ pozice konstanty A, pak i $\boldsymbol{V}\updownarrow$ je pozice konstanty A, a je-li definován $\boldsymbol{V}\searrow$, tak i ta je pozicí konstanty A.

Definice 3.7. Buď $\mathcal{S}$ kvadratická soustava, r její řešení a M množina pozic v řešení r s následující vlastností: Zdroj každé pozice $\boldsymbol{V} \in M$ je výskyt proměnné a vyskytuje-li se tato proměnná v soustavě dvakrát, pak i $\boldsymbol{V}\searrow \in M$. Pak definujeme homomorfismus s zachovávající konstanty vzniklý *odstraněním pozic množiny* M následujícím předpisem pro každou proměnnou x a její výskyt $\boldsymbol{v}$.

$$s(x) = r(x)\,|_{\{0 \le i \le |r(x)|-1 | \mathrm{Img}(\boldsymbol{v})+i \notin M\}}$$

minus-vysk-res]

Pozorování 3.8. Pokud odstraníme z řešení r množinu výskytů, bude vzniklý homomorfismus opět řešením právě když pro každou rovnici $S_1 = S_2$ soustavy platí

$$r(S_1)\,|_{\{\mathrm{Ind}(\boldsymbol{V}) | \boldsymbol{V} \in r[S_1] \setminus M\}} = r(S_1)\,|_{\{\mathrm{Ind}(\boldsymbol{V}) | \boldsymbol{V} \in r[S_2] \setminus M\}}$$

Pozorování 3.9. Řešení vzniklé odstraněním neprázdné množiny výskytů je ostře l-menší i c-menší než původní řešení.

3.1.3 Pokročilejší skoky – protějšek

Definice 3.10. Buď $\boldsymbol{V}$ pozice v řešení. Pak uvažujme posloupnost

$$\boldsymbol{V}\updownarrow, \boldsymbol{V}\updownarrow\searrow\updownarrow, \boldsymbol{V}(\updownarrow\searrow)^2\updownarrow, \ldots,$$

První člen této posloupnosti, jehož zdroj je buď konstanta nebo proměnná, která se v soustavě vyskytuje pouze jednou, nazýváme protějškem pozice $\boldsymbol{V}$ a značíme $\boldsymbol{V}\curlyvee$.

Pozorování 3.11. Je-li $\mathrm{Src}(\boldsymbol{V})$ konstanta nebo proměnná, která se v soustavě nevyskytuje dvakrát, pak $\boldsymbol{V}\curlyvee\updownarrow\curlyvee = \boldsymbol{V}$. V opačném případě je $\boldsymbol{V}\curlyvee\updownarrow\curlyvee = \boldsymbol{V}\searrow\curlyvee$.

Za všech okolností ale platí $\boldsymbol{V}\curlyvee\curlyvee = \boldsymbol{V}\updownarrow\curlyvee$.

Definice 3.12. Mějme k-tici po sobě jdoucích pozic $z = (\boldsymbol{V}, \boldsymbol{V}+1, \ldots, \boldsymbol{V}+(k-1))$. Pak definujeme po složkách

$$z\updownarrow = (\boldsymbol{V}\updownarrow, (V+1)\updownarrow, \ldots, (\boldsymbol{V}+(k-1))\updownarrow), \quad z\searrow = (\boldsymbol{V}\searrow, (V+1)\searrow, \ldots, (\boldsymbol{V}+(k-1))\searrow).$$

Nakonec definujeme protějšek k-tice z značený $z\curlyvee$ jako první člen posloupnosti

$$z\updownarrow, z\updownarrow\searrow\updownarrow, z\updownarrow\searrow\updownarrow\searrow\updownarrow, \ldots,$$

kde se buď liší zdroje některých prvků k-tice nebo společný zdroj $\boldsymbol{v}$ všech prvků je konstanta[1] nebo nemá definovaný $\boldsymbol{v}\searrow$.

[1] Pro $k > 1$ tento případ ani nemůže nastat.

Definice 3.13. Mějme k-tici po sobě jdoucích pozic $z = (\boldsymbol{v}, \boldsymbol{v}+1, \ldots, \boldsymbol{v}+(k-1))$ a dále danou množinu P pozic. Pak definujeme protějšek k-tice z v množině P jako první člen posloupnosti

$$z\updownarrow, z\updownarrow\downarrow\updownarrow, z\updownarrow\downarrow\updownarrow\downarrow\updownarrow, \ldots,$$

který buď splňuje podmínky na protějšek k-tice z (resp. na protějšek prvku, je-li $k = 1$) nebo pro některý $\boldsymbol{v} \in z$ je $\boldsymbol{v}\downarrow \in P$.

Protějšek v množině
Prostota v množině
zlom

Definice 3.14. *Zlomem* v soustavě rozumíme dvojici výskytů v soustavě $(\boldsymbol{v}, \boldsymbol{v}+1)$. Představa zlomu je „ta mezera mezi písmeny“.

Pozorování 3.15. V soustavě délky n o r rovnících je přesně $2(n-r)$ zlomů.

Protějšek zlomu
Prostota na zlomech

3.2 Operace se soustavou a jejím řešením

3.2.1 Smazání prázdných proměnných

3.2.2 Vepsání konstanty

3.2.3 Slepené konstanty

3.2.4 Slití slepených konstant

3.3 Na které soustavy se stačí zaměřit

Definice 3.16. (2-rovnice, 2-soustava)

Pozorování 3.17. Pro 2-soustavu $\mathcal{R}$ platí $\mathcal{R} = \mathscr{V}_{\mathcal{R}} + \mathscr{C}_{\mathcal{R}}$.

Tvrzení 3.18. Je-li dána řešitelná kvadratická soustava $\mathcal{S}$, pak existuje 2-soustava $\mathcal{S}'$, pro kterou platí

- $|\mathcal{S}| \geq |\mathcal{R}'|/2$,
- $\min(\mathcal{S}) \leq \min(\mathcal{S}')$.

Důsledek 3.19. Kdybychom měli polynomiální resp. jednoduše exponenciální mez na 2-soustavy, měli bychom polynomiální resp. jednoduše exponenciální mez na všechny kvadratické soustavy.

3.4 Dvojitě exponenciální mez

3.5 Co by stačilo pro jednoduše exponenciální mez

[omezeni-konstant]

Tvrzení 3.20. Je-li dána řešitelná 2-soustava $\mathcal{S}$, pak existuje 2-soustava $\mathcal{S}'$, pro kterou platí

1) $|\mathscr{C}_{\mathcal{S}'}| \leq 3|\mathscr{V}_{\mathcal{S}'}| + \langle\!\langle \mathcal{S} \rangle\!\rangle$,

2) $|\mathcal{V}_{\mathcal{S}'}| = |\mathcal{V}_{\mathcal{S}}|$,
3) $\langle\!\langle S' \rangle\!\rangle = \langle\!\langle \mathcal{S} \rangle\!\rangle$,
4) $|\mathcal{S}| \geq |\mathcal{S}'|$,
5) $\min(\mathcal{S}) \leq \min(\mathcal{S}') \cdot |\mathcal{S}|$.

Důkaz: Zdrcnutím konstant vytvoříme soustavu $\mathcal{S}'$. Každá konstanta soustavy $\mathcal{S}'$ odpovídá nejvýše $|\mathcal{S}|$ konstantám původní soustavy, tedy máme-li libovolné řešení r' soustavy S', je možné z něj vytvořit řešení r soustavy $\mathcal{S}$ splňující $|r| \leq |r'| \cdot |\mathcal{S}|$. Podmínky (2), (3), (4) jsou splněny zřejmě, zbývá ukázat omezení na počet konstant. Víme, že $\mathcal{R}'$ je řešitelná, uvažujme tedy řešení r.

Víme, že v $\mathcal{S}'$ již nejsou žádné dvě konstanty spojené. Proto kdykoli vezmeme zlom mezi dvěmi konstantami, musí být jeho protějškem zlom u proměnné. Protějšek je na zlomech bijekce, tedy označíme-li c počet zlomů konstant a d počet zlomů u proměnných, máme nerovnost $c \leq d$. Současně $d \leq 4\mathcal{V}_{\mathcal{S}'}$, jelikož je v soustavě každá proměnná dvakrát a má nejvýše dva zlomy.

Kdykoli vezmeme výskyt $\boldsymbol{v}$ nějaké konstanty, tak tento výskyt může být poslední v některé rovnici (nejvýše ve $2\langle\!\langle \mathcal{S} \rangle\!\rangle$ případech) nebo je $\boldsymbol{v}+1$ proměnná (nejvýše v $2|\mathcal{V}_{\mathcal{S}'}|$ případech), ve všech ostatních případech je $(\boldsymbol{v}, \boldsymbol{v}+1)$ zlom mezi konstantami. To dává odhad

$$2\mathcal{C}_{\mathcal{S}'} - 2\langle\!\langle \mathcal{S} \rangle\!\rangle - 2\mathcal{V}_{\mathcal{S}'} \leq c \leq d \leq 4\mathcal{V}_{\mathcal{S}'},$$

ekvivalentně

$$|\mathcal{C}_{\mathcal{S}'}| \leq 3|\mathcal{V}_{\mathcal{S}'}| + \langle\!\langle \mathcal{S} \rangle\!\rangle .$$

■

Důsledek 3.21. Kdybychom měli polynomiální resp. jednoduše exponenciální mez pro 2-soustavy splňující $|\mathcal{V}| \geq 3|\mathcal{C}|+1$, měli bychom polynomiální resp. jednoduše exponenciální mez pro všechny kvadratické soustavy.

[omez-min]
Tvrzení 3.22. Předpokládejme, že existuje rostoucí polynom p s vlastností: Pro každou řešitelnou 2-soustavu resp. 2-rovnici s alespoň jednou proměnnou existuje řešení r a některá proměnná této soustavy x, že platí $|r(x)| \leq 2^{p|\mathcal{R}|}$. Jinými slovy předpokládáme, že máme jednoduše exponenciální omezení na nejmenší možnou délku nejmenší proměnné. Tvrdíme, že pak už bychom měli jednoduše exponenciální mez pro délku celého řešení pro kvadratické soustavy resp. rovnice.

Důkaz: Stačí ukázat mez na délku nejmenšího řešení pro 2-soustavu $\mathcal{S}$ splňující $|\mathcal{V}_{\mathcal{S}}| \geq 3|\mathcal{C}_{\mathcal{S}}| + \langle\!\langle \mathcal{S} \rangle\!\rangle$, označme $k = |\mathcal{V}_{\mathcal{S}}|$. Označme $\mathcal{S}_k = \mathcal{S}$. Dále postupně vždy vytvoříme ze soustavy $\mathcal{S}_i$ s i proměnnými soustavu $\mathcal{T}_{i-1}$ s $(i-1)$ proměnnými a z té pak soustavu $\mathcal{S}_{i-1}$ s $(i-1)$ proměnnými. Přitom počet rovnic zachováváme. To provádíme až do $\mathcal{S}_0$ následujícím postupem:

(i) Máme-li $\mathcal{S}_i$, tak na základě předpokladu existuje řešení r a proměnná x pro které $|r(x)| \leq 2^{p|\mathcal{S}_i|}$. Rovnici T_{i-1} sestrojíme tak, že tuto proměnnou v rovnici přímo nahradíme jejím řešením a následně rozrůzníme konstanty. Platí $|\mathcal{T}_{i-1}| \leq |\mathcal{S}_i| + 2^{p|\mathcal{S}_i|}$ a $\min(\mathcal{S}_i) \leq \min(\mathcal{T}_{i-1})$.
(ii) Máme-li $\mathcal{T}_i$, sestrojíme S_i podle tvrzení [omezeni-konstant] jako takovou rovnici, která splňuje

1) $|\mathcal{C}_{\mathcal{S}_i}| \leq 3i + \langle\!\langle \mathcal{R} \rangle\!\rangle \leq 3k + \langle\!\langle \mathcal{R} \rangle\!\rangle$,
2) $|\mathcal{Q}_i| \geq |\mathcal{R}_i|$,
3) $\min(\mathcal{Q}_i) \leq \min(\mathcal{R}_i) \cdot |\mathcal{Q}|$.

Pak pro všechna $i = k, \ldots, 0$ platí

- $|\mathcal{S}_i| \leq 4k + \langle\!\langle \mathcal{S} \rangle\!\rangle \leq 5|\mathcal{S}|$,

- $|\mathcal{T}_{i-1}| \leq |\mathcal{S}_i| + 2^{p|\mathcal{S}_i|} \leq 2^{p_2|\mathcal{S}|}$ pro nějaký rostoucí polynom p_2 nezávislý na $\mathcal{S}$.
- $\min(\mathcal{S}_i) \leq \min(\mathcal{T}_{i-1}) \leq \min(\mathcal{S}_{i-1}) \cdot 2^{p_2|\mathcal{S}|}$.

Tedy celkem

$$\min(\mathcal{S}_k) \leq \min(\mathcal{S}_{k-1}) \cdot 2^{p_2|\mathcal{S}|} \leq \cdots \leq \min(\mathcal{S}_0) \cdot (2^{p_2|\mathcal{S}|})^k.$$

Vzhledem k tomu, že $|\mathcal{R}_0| \leq \langle\!\langle \mathcal{S} \rangle\!\rangle$, je i $\min(\mathcal{R}_0) \leq \langle\!\langle \mathcal{S} \rangle\!\rangle$. Tedy

$$\min(\mathcal{R}) \leq \langle\!\langle \mathcal{S} \rangle\!\rangle \cdot (2^{p_2(k)})^k \leq |\mathcal{S}| \cdot 2^{p_2|\mathcal{S}| \cdot k}.$$

Pokud bychom uvažovali pouze jednu rovnici, důkaz projde stejně, vzhledem k tomu, že při postupném odvozování T_i a S_i počet rovnic v soustavě neroste. ∎

[vysk-to-eq]
Tvrzení 3.23. Nechť je dána řešitelná 2-soustava $\mathcal{S}$ a její proměnná A. Pak existuje řešitelná 2-soustava $\mathcal{S}'$ řešení r soustavy $\mathcal{S}$ splňující

- $\mathcal{V}_{\mathcal{R}} \subseteq \mathcal{V}_{\mathcal{S}}$,
- $|\mathcal{S}'| \leq |\mathcal{S}|$,
- $\mathrm{Freq}_{r,\mathcal{S}}(A) = \min \mathcal{S}'$,
- V soustavě $\mathcal{S}$ existuje nejvýše jeden zlom mezi konstantami.

Důkaz: ... ∎

Poznámka 3.24. Jen poznamenejme, že kvadratické soustavy, ve kterých se nevyskytují dvě konstanty za sebou nejsou o moc výjimečnější či méně obecné než obecné kvadratické soustavy. Z libovolné kvadratické soustavy totiž můžeme udělat „ekvivalentní" soustavu splňující požadavek na řídké konstanty takto: Pro každý výskyt $\boldsymbol{v}$ konstanty založíme 3 nové proměnné $x_{\boldsymbol{v}}, a_{\boldsymbol{v}}, y_{vsv}$. Označme původní konstantu $A = \boldsymbol{v}$ a nahradíme konstantu $A_{\boldsymbol{v}}$ v tomto výskytu proměnnou $a_{\boldsymbol{v}}$. Navíc přidáme do soustavy rovnici

$$x_{\boldsymbol{v}} A_{\boldsymbol{v}} y_{\boldsymbol{v}} = x_{\boldsymbol{v}} a_{\boldsymbol{v}} y_{\boldsymbol{v}}.$$

[periodicita]
3.6 Periodicita

3.6.1 Opakující se konstanta

Tvrzení 3.25. Uvažujme kvadratickou soustavu $\mathcal{S}$, její řešení r a její konstantu A. Číslem n označme počet výskytů konstanty A a předpokládejme, že v řešení existuje alespoň $(n+1)$ po sobě jdoucích pozic konstanty A. Pak existuje řešení r' soustavy $\mathcal{S}$, které je l-menší i c-menší než r. Je-li navíc $\mathcal{S}$ orientovaná soustava, stačí n volit jako počet výskytů konstanty A na jedné straně.

Důkaz: Označme $(\boldsymbol{V}, \boldsymbol{V}+1, \ldots, \boldsymbol{V}+n)$ po sobě jdoucí pozice konstanty A. Protějšek každé takové pozice je obraz některého výskytu konstanty A. Pak tedy existují dvě různé takové pozice $\boldsymbol{V}+i$, $\boldsymbol{V}+j$ (kde $0 \leq i, j \leq n$), které mají stejný protějšek. Pokud je navíc soustava orientovaná, budou všechny protějšky na opačné straně než zmiňované výskyty, proto pro existenci dvou výskytů se stejným protějškem stačí volit n jako počet těchto výskytů.

Uvažme tedy pozici $\boldsymbol{W}$ jako společný protějšek pozic $\boldsymbol{V}+i$ a $\boldsymbol{V}+j$. V posloupnosti

$$\boldsymbol{W}\updownarrow, \boldsymbol{W}\updownarrow\downarrow\updownarrow, \boldsymbol{W}\updownarrow\downarrow\updownarrow\downarrow\updownarrow, \ldots$$

se musí někde objevit pozice $\boldsymbol{V}+i$ i pozice $\boldsymbol{V}+j$. BÚNO se $\boldsymbol{V}+i$ v této posloupnosti objeví dříve. Pak je $\boldsymbol{V}+j = (\boldsymbol{V}+i)(\downarrow\updownarrow)^k$ pro nějaké k a $\mathrm{Src}(\boldsymbol{V}+i)$ je výskyt proměnné.

Označme množinu M pozic jako

$$\{(\boldsymbol{V}+i), (\boldsymbol{V}+i)\searrow, (\boldsymbol{V}+i)(\searrow\updownarrow)^1, (\boldsymbol{V}+i)(\searrow\updownarrow)^1\searrow, \ldots, (\boldsymbol{V}+i)(\searrow\updownarrow)^{k-1}, (\boldsymbol{V}+i)(\searrow\updownarrow)^{k-1}\searrow\}.$$

Homomorfiskus r' pak volíme odstraněním pozic množiny M. Zbývá ověřit, že tento homomorfismus je řešením, tedy že $r'(S_1) = r'(S_2)$ pro každou rovnici $S_1 = S_2$ soustavy $\mathcal{S}$. Pro rovnice R různé od rovnice pozice $\boldsymbol{V}$ je to zřejmé, jelikož pro takové rovnice každé $\boldsymbol{U} \in r[R]$ splňuje $\boldsymbol{U} \in M$ právě když $\boldsymbol{U}\updownarrow \in M$.

Nakonec ukážeme rovnost pro rovnici pozice $\boldsymbol{V}$. Označme S_1 stranu rovnice pozice $\boldsymbol{V}$ a S_2 druhou stranu téže rovnice. Pak

$$r'(S_1) = r(S_1)\,|_{\{\mathrm{Ind}_r(\boldsymbol{U})|\boldsymbol{U}\in r[S_1]\setminus M\}},$$

zatímco

$$r'(S_2) = r(S_2)\,|_{\{\mathrm{Ind}_r(\boldsymbol{U})|\boldsymbol{U}\in r[S_1]\setminus M\}} =$$
$$= r(S_1)\,|_{\{\mathrm{Ind}_r(\boldsymbol{U})|U\in r[S_1]\setminus M\}\cup\{\boldsymbol{V}+j\}\setminus\{\boldsymbol{V}+i\}},$$

a to je totéž, vzhledem k tomu, že $\boldsymbol{V}+i$ a $\boldsymbol{V}+j$ leží ve stejném souvislém bloku A-ček. ■

[const-strid]

Důsledek 3.26. Pro libovolné řešení r' orientované 2-soustavy existuje řešení r, které je l-menší i c-menší a současně v něm nesou žádné dvě po sobě jdoucí pozice jedné konstanty. Specielně tedy v libovolném l-minimálním i c-minimálním řešení orientované 2-soustavy nejdou nikde dvě stejné konstanty po sobě.

3.6.2 Lineární mez na exponent periodicity

Ve zdroji [1] je dokázaná lineární mez na exponent periodicity. Ukážeme zde tentýž výsledek pomocí protějšku množiny. Navíc pak ukážeme, že je možné libovolnou řešitelnou soustavu „převést“ na soustavu s pouze jedničkovým exponentem periodicity.

Definice 3.27. Buď $\mathcal{S}$ soustava na slovech a r její řešení. Pak označme *exponent periodicity řešení* r jako největší přirozené číslo n takové, že existuje proměnná x a slova $\alpha, \beta, \psi \in \mathscr{C}_{\mathcal{S}}$ splňující

$$r(x) = \alpha\psi^n\beta.$$

Exponentem periodicity řešitelné soustavy $\mathcal{S}$ rozumíme největší exponent periodicity přes všecha možná nejmenší řešení.

Tvrzení 3.28. Uvažujme kvadratickou soustavu $\mathcal{S}$, její řešení r, pozici $\boldsymbol{V}$ a číslo $k > 1$. Dále volme z jako počet zlomů soustavy. Předpokládejme, že pro všechna celá čísla i splňující $0 \le i \le k\cdot(z+2)-1$ je

$$\overline{\boldsymbol{V}+i} = \overline{\boldsymbol{V}+i+k}.$$

Pak existuje řešení soustavy $\mathcal{S}$, které je ostře l-menší i c-menší než r.

Důkaz: Volme k je nejmenší možné, pak bude platit, že pro každé $i \in \{0, \ldots, k-1\}$ budou uspořádané k-tice konstant

$$(\overline{(\boldsymbol{V}+i)}, \overline{(\boldsymbol{V}+i)+1}, \ldots \overline{(\boldsymbol{V}+i)+(k-1)}),$$

navzájem různé.

Uvažme nyní $z+3$ následujících k-tic po sobě jdoucích pozic

$$(\boldsymbol{V}, \boldsymbol{V}+1, \ldots \boldsymbol{V}+(k-1)),$$
$$((\boldsymbol{V}+k), (\boldsymbol{V}+k)+1, \ldots (\boldsymbol{V}+k)+(k-1)),$$
$$\vdots$$
$$((\boldsymbol{V}+k\cdot(z+2)), (\boldsymbol{V}+k\cdot(z+2)+1), \ldots, (\boldsymbol{V}+k\cdot(z+2)+(k-1)))$$

Všem těmto k-ticím najděme protějšek v množině $M = \{\boldsymbol{V}, \ldots, \boldsymbol{V} + k \cdot (z+3) - 1\}$. Všechny tyto protěšky jsou disjunktní. Nejvýše z těchto protějšků může obsahovat zlom, pro všechny ostatní protějšky p musí splňovat, že $p\updownarrow$ protíná množinu M. Navíc vzhledem k tomu, že M je množina po sobě jdoucích pozic, mohou existovat nejvýše dva protějšky p, pro které $p\updownarrow$ množinu M protíná, ale není v ní celé obsaženo. Proto zde je protějšek p takový, že $p\updownarrow$ je podmnožinou M a proto (z minimality k) je $p\updownarrow$ opět jednou z původních k-tic.

Označme k-tici splňující $t\curvearrowright\updownarrow \subseteq M$ jako t, přitom $t\curvearrowright = t(\updownarrow\updownarrow)^n\updownarrow$ pro vhodné n. Pak můžeme z řešení r odstranit množinu výskytů

$$t\updownarrow \cup t\updownarrow\updownarrow \cup \ldots \cup t(\updownarrow\updownarrow)^{n+1}.$$

■

Důsledek 3.29. Exponent periodicity kvadratických rovnic je možné shora lineárně omezit vzhledem k její délce.

Poznámka 3.30. Lineární mez nejde zlepšit, a to ani pro 2-soustavy. Uvažme následující rovnici, kde A, B jsou konstanty, jednotlivá a_i, b_i pak proměnné

$$a_1 B a_2 b_1 a_3 b_2 \ldots a_n b_{n-1} A b_n = A b_1 a_1 b_2 a_2 \ldots b_n a_n B.$$

V této rovnici nemůže mít v řešení žádná proměnná nulovou délku, taková skutečnost by totiž znamenala rozpadnutí rovnice na dvě neřešitelné 2-rovnice. Jediné nejmenší řešení r tedy je, když pro všechna i je $r(a_i) = A$ a $r(b_i) = B$.

Pokud zavedeme novou proměnnou x a sestrojíme soustavu

$$a_1 B a_2 b_1 a_3 b_2 \ldots a_n b_{n-1} A b_n = x, \quad x = A b_1 a_1 b_2 a_2 \ldots b_n a_n B,$$

bude nejmenší řešení stále stejné, avšak exponent periodicity bude $n + 1$.

3.6.3 Převeditelnost na jednotkový exponent periodicity

Tvrzení 3.31. Pro každou kvadratickou soustavu $\mathcal{S}$ existuje kvadratická soustava $\mathcal{T}$, s následujícími vlastnostmi.

- Exponent periodicity soustavy $\mathcal{T}$ je roven jedné,
- $\min(\mathcal{T}) = \min(\mathcal{S})$,
- $|\mathcal{T}| \leq 3|\mathcal{S}|$.

Důkaz: Uvažme minimální řešení r soustavy $\mathcal{S}$. Soustavu $\mathcal{T}$ sestrojíme tak, že základě r vepíšeme všem neprázdným proměnným konstantu na začátek, následně všem neprázdným proměnným vepíšeme konstantu na konec a pak smažeme prázdné proměnné. Nakonec na základě toho řešení rozrůzníme konstanty a výsledek označíme za soustavu $\mathcal{T}$.. První dvě podmínky jsou zřejmě splněny, zbývá ukázat, že exponent periodicity $\mathcal{T}$ je roven jedné.

Uvažme tedy nějaké nejkratší řešení s této soustavy. ■

Důsledek 3.32. Pokud bychom měli polynomiální resp. jednoduše exponenciální mez pro ty kvadratické soustavy, jejichž exponent periodicity je roven jedné, měli bychom polynomiální resp. jednoduše exponenciální mez pro všechny kvadratické soustavy.

[makro]

Kapitola 4

Makro-operace s kvadratickou rovnicí aneb lámání a dosazování

4.1 Operace

4.1.1 Smazání prázdných proměnných

Definice 4.1. Uvažujme soustavu $\mathcal{S}$, která obsahuje alespoň jednu konstantu, a její řešení r. Definujeme soustavu $\mathcal{T}$ s řešením s, která vznikne *smazáním prázdných proměnných řešení* r jednoduše tak, že ze soustavy odstraníme všechny výskyty všech proměnných, které mají v řešení r prázdnou délku. Pokud má v řešení r nulovou délku celá některá rovnice, vyhodíme ji celou ze soustavy. Že nakonec něco zbude je zaručeno požadavkem na existenci konstanty v rovnici. Řešení s vznikne z r jednoduše tak, že zapomeneme hodnoty na vyhozených proměnných.

Pozorování 4.2. Z libovolného řešení s soustavy $\mathcal{T}$ je možné opět odvodit řešením r' soustavy $\mathcal{S}$. Splňující pro všechny proměnné x soustavy $\mathcal{T}$ rovnost $r'(x) = s(x)$ a pro ostatní proměnné soustavy $\mathcal{S}$ je $|x|_{r'} = 0$ Obráceně to můžeme provést právě pro ta řešení r', která splňují pro všechny proměnné soustavy x splňují implikaci $|x|_r = 0 \Rightarrow |x|_{r'} = 0$. Z toho plyne $\min T \geq \min S$. Pokud je navíc r minimální řešení, nastává rovnost.

Definice 4.3. Pojem *odvozené řešení* budeme v souvislosti se smazáním prázdných proměnných proměnné používat ve smyslu předchozího pozorování.

Pozorování 4.4. Smazáním prázdných proměnných se zachová množina konstant a nezvýší se počet zlomů.

4.1.2 Lámání rovnic

Definice 4.5. Uvažujme soustavu $\mathcal{S}$, její řešení r a její dva zlomy $(\boldsymbol{v}-1, \boldsymbol{v})$, $(\boldsymbol{w}-1, \boldsymbol{w})$ takové, že jsou to zlomy stejné rovnice na jejích různých stranách a navíc $\mathrm{Ind}_r(\boldsymbol{v}) = \mathrm{Ind}_r(\boldsymbol{w})$. Pak definujeme soustavu $\mathcal{T}$ vzniklou *rozdělením rovnice* $\mathcal{S}$ v těchto zlomech: Soustava $\mathcal{T}$ obsahuje stejné rovnice jako soustava $\mathcal{S}$ až na to, že neobsahuje rovnici výskytů $\boldsymbol{v}, \boldsymbol{w}$. Místo toho obsahuje dvě nové rovnice:

$$\ldots \overline{\boldsymbol{v}-3}\,\overline{\boldsymbol{v}-2}\,\overline{\boldsymbol{v}-1} = \ldots \overline{\boldsymbol{w}-3}\,\overline{\boldsymbol{w}-2}\,\overline{\boldsymbol{w}-1}, \quad \overline{\boldsymbol{v}}\,\overline{\boldsymbol{v}+1}\,\overline{\boldsymbol{v}+2}\ldots = \overline{\boldsymbol{w}}\,\overline{\boldsymbol{w}+1}\,\overline{\boldsymbol{w}+2}\ldots$$

Pozorování 4.6. Při použití značení z předchozí definice je r současně řešením soustavy $\mathcal{T}$. Dále libovolné řešení soustavy $\mathcal{T}$ je opět řešením soustavy $\mathcal{S}$. Obráceně to platí jen pro ta řešení, která splňují podmínku pro zlomy $\mathrm{Ind}_r(\boldsymbol{v}) = \mathrm{Ind}_r(\boldsymbol{w})$. Z toho plyne $\min T \geq \min S$. Pokud je navíc r minimální řešení, nastává rovnost.

Pozorování 4.7. Pokud v soustavě rozlomíme rovnici, snížíme počet celkový zlomů o 2 a zachováme množinu konstant.

Dokud lámání rovnic je možné, můžeme zvesela vytvářet složitější a složitější rovnice. Avšak hodilo by se moci rovnici „násilím“ zlomit tam, kde nám k tomu sama nedává příležitost. Musíme si tedy zlomy naproti sobě vyrobit.

4.1.3 Lámání proměnných

Definice 4.8. Uvažujme soustavu $\mathcal{S}$, její řešení r, výskyt $\boldsymbol{v}$ v soustavě její proměnné x a dvě nezáporné celá čísla z splňující

$$\mathrm{Ind}_r(\boldsymbol{v}) \le z \le \mathrm{Ind}_r(\boldsymbol{v}) + |x|_r.$$

$|x|_r$. Pak definujeme soustavu $\mathcal{T}$ s řešením s vzniklou *rozlomením výskytu* $\boldsymbol{v}$ *v bodě* z na základě řešení r: Soustava $\mathcal{T}$ vznikne nahrazením všech výskytů proměnné x za dvojici nových proměnných x_1x_2. V řešení s pak

$$s(x_1) = r(x)\,|_{\{0,\dots,z-\mathrm{Ind}_r(\boldsymbol{v})-1\}}, \qquad s(x_2) = r(x)\,|_{\{z-\mathrm{Ind}_r(\boldsymbol{v}),\dots,|x|_r-1\}}\,.$$

Dvojici výskytů v $\mathcal{T}$ vzniklou nahrazením výskytu $\boldsymbol{v}$ říkáme *zlom vzešlý z rozdělení proměnné*.

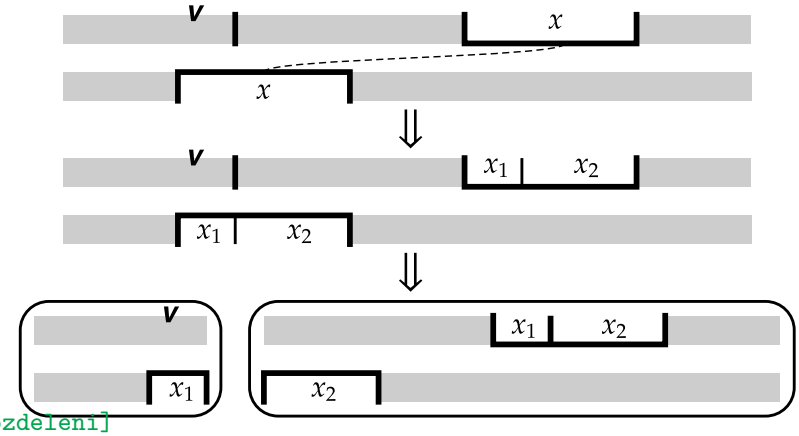
Pozorování 4.9. Při použití značení z předchozí definice lze z každého řešení s soustavy $\mathcal{T}$ odvodit řešení r' soustavy $\mathcal{S}$ splňující $s(y) = r'(y)$ pro každou proměnnou různou od x a dále $s(x_1x_2) = r(x)$. Totéž lze provést i obráceně. Tedy z libovolného r' odvodit s. Tedy například $\min T \ge \min S$.

Definice 4.10. Pojem *odvozené řešení* budeme v souvislosti v rozlomením proměnné používat ve smyslu předchozího pozorování.

Pozorování 4.11. Rozlomením proměnné v kvadratické soustavě zvýšíme celkový počet zlomů nejvýše o 2 a zachováme množinu konstant.

Definice 4.12. Uvažujme soustavu $\mathcal{S}$, její řešení r, výskyt $\boldsymbol{v} = (R, S, i)$ v soustavě její konstanty nebo neprázdné proměnné x. Předpokládáme, že je-li $\boldsymbol{v}+1$ definované, tak $|\boldsymbol{v}+1|_r > 0$. Definujeme soustavu $\mathcal{T}$ s řešením s vzniklou *rozlomením soustavy* $\mathcal{S}$ *za výskytem* $\boldsymbol{v}$ *na základě řešení* r. Je-li $\boldsymbol{v}$ poslední (tedy $\boldsymbol{v}+1$ není definováno), ponecháváme $\mathcal{T} = \mathcal{S}$ a $s = r$. V opačném případě se podíváme pozice $\boldsymbol{V}, \boldsymbol{W}$ definované jako $\boldsymbol{V} = \mathrm{Img}(\boldsymbol{v}) + (|\boldsymbol{v}|_r - 1)$ a $\boldsymbol{W} = \boldsymbol{V}^{\updownarrow}$ a konečně označíme výskyt $\boldsymbol{w} = \mathrm{Src}(\boldsymbol{W})$. Platí-li $\boldsymbol{w} \ne \mathrm{Src}(\boldsymbol{W}+1)$, rozlomíme rovnici ve dvojici zlomů $(\boldsymbol{v}, \boldsymbol{v}+1)$ a $(\boldsymbol{w}, \boldsymbol{w}+1)$. V opačném případě nejprve rozlomíme výskyt $\overline{w}$ v bodě $\mathrm{Ind}(\boldsymbol{W}+1)$ a pak teprve rozlomíme rovnici ve zlomu $(\boldsymbol{v}, \boldsymbol{v}+1)$ a zlomu vzešlém z rozlomení proměnné. Výsledek prohlásíme za soustavu $\mathcal{T}$ s řešením s.

Analogicky definujeme i *rozlomení soustavy* $\mathcal{S}$ *před výskytem* $\boldsymbol{v}$. Nakonec definujeme vylomení výskytu jako postupné rozlomení soustavy před a za ním.



[img-rozdeleni]

Obrázek 4.1. Rozlomení soustavy za výskytem $\boldsymbol{v}$.

Pozorování 4.13. Při použití značení z předchozí definice lze z každého řešení s soustavy $\mathcal{T}$ odvodit řešení r' soustavy $\mathcal{S}$ (ať už lámeme před výskytem, za výskytem, či dokonce vylamujeme výskyt). Opět tedy budeme hovořit o odvozených řešeních.

Pozorování 4.14. Rozlamování před/za výskyty v kvadratických soustavách (a tedy i vylamování výskytů) nezvýší počet zlomů a zachovává množinu konstant.

Ačkoli počet zlomů většinou zůstává stejný a každopádně neroste, počet rovnic může růst do aleluja. Je však nabíledni, že rovnice beze zlomů bude možné redukovat.

4.1.4 Dosazování

Definice 4.15. O rovnici R v soustavě $\mathcal{S}$ řekneme, že je *dosaditelná*, pokud alespoň jedna ze stran má délku 1 a navíc je tato strana tvořena proměnnou.

Definice 4.16. O dosaditelné rovnici R v soustavě $\mathcal{S}$ řekneme, že je *triviální*, pokud mají obě strany délku 1.

Definice 4.17. Uvažujme v soustavě $\mathcal{S}$ s řešením r dosaditelnou rovnici R a její stranu S délky 1 s proměnnou x. Definujeme *soustavu $\mathcal{T}$ vzniklou dosazením této strany resp. výskytu* $(R, S, 0)$. Označme S' druhou stranu rovnice R. Soustava $\mathcal{T}$ vznikne z $\mathcal{S}$ tak, že smažeme rovnici R a všechny výskyty proměnné x v rovnici nahradíme slovem S'.

Poznámka 4.18. Tedy, je-li rovnice tvaru $x = x$, je její dosazení jen jejím odstraněním.

Pozorování 4.19. Při použití značení z předchozí definice lze z každého řešení s soustavy $\mathcal{T}$ jednoznačně odvodit řešení r soustavy $\mathcal{S}$, aby pro všechny proměnné x soustavy $\mathcal{T}$ platilo $r(x) = s(x)$. Opět tedy budeme hovořit o odvozených řešeních. Délka odvozeného řešení v tomto případě splňuje nerovnost $|s|_{\mathcal{T}} \leq |r|_{\mathcal{R}} \leq 2|s|_{\mathcal{T}}$.

Pozorování 4.20. Dosazení v kvadratických soustavách nezmění počet zlomů, zato sníží počet rovnic i délku soustavy o 1.

Pozorování 4.21. Pokud provádíme dosazování, dokud to někde jde, tak po konečném čase skončíme a výsledek (až na isomorfismus soustav) nezávisí na pořadí, ve kterém jsme dosazovali.

Přirozený postup nyní je vylomit výskyt proměnné a následně jej dosadit. Zádrhel by mohl nastat, pokud bychom vylomením výskytu „zasáhli“ druhý výskyt této proměnné – tím bychom vylomením nezískali dosaditelnou rovnici. Ukážeme, že v minimálním řešení se to stát nemůže a dáme postup, jak se s takovou situací obecně vypořádat.

4.1.5 Krácení překryvů

Definice 4.22. Uvažujme kvadratickou soustavu $\mathcal{S}$ a její řešení r a proměnnou x. *Překryvem* proměnné x v řešení r rozumíme dvojici výskytů této proměnné $(\boldsymbol{v}, \boldsymbol{w})$ v soustavě, takovou, že tyto výskyty leží na různých stranách jedné rovnice a navíc

$$0 < |\operatorname{Ind}_r(\boldsymbol{v}) - \operatorname{Ind}_r(\boldsymbol{w})| \leq |x|_r.$$

Pokud nastává rovnost, nazýváme tento překryv triviálním, v případě ostré nerovnosti naopak mluvíme o netriviálním překryvu.

Pozorování 4.23. Pokud proměnná nemá překryv a kolem svých výskytů nemá prázdné proměnné, můžeme jeden její výskyt z rovnice vylomit, tím získáme dosaditelnou rovnici, tedy následně můžeme tento výskyt dosadit.

Mějme kvadratickou soustavu $\mathcal{S}$, její řešení r a překryv $(\boldsymbol{v} = (R, S, i), \boldsymbol{w} = (R, S', j)))$, BÚNO předpokládejme $\operatorname{Ind}_r(\boldsymbol{v}) \geq \operatorname{Ind}_r(\boldsymbol{w})$. Označme $k = \operatorname{Ind}_r(\boldsymbol{v}) - \operatorname{Ind}_r(\boldsymbol{w})$, dále pozici $\boldsymbol{V} = \operatorname{Img}(\boldsymbol{v})$ a definujeme homomorfismus s odstraněním pozic množiny

$$\boldsymbol{V}, \boldsymbol{V} + 1, \ldots, \boldsymbol{V} + (k - 1),$$

$$\boldsymbol{V}\!\!\downarrow, (\boldsymbol{V}+1)\!\!\downarrow, \ldots, (\boldsymbol{V}+(k-1))\!\!\downarrow.$$

Tedy proměnná x je v tomto řešení o k kratší.

Tvrzení 4.24. Homomorfismus s popsaný výše je skutečně řešením rovnice $\mathcal{S}$.

Důkaz: Pro libovolnou pozici $\boldsymbol{W}$ z množiny

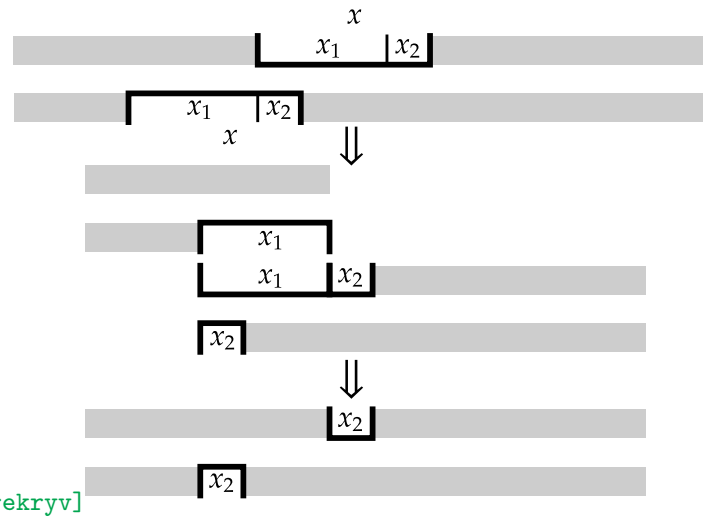
$$\boldsymbol{V}, \boldsymbol{V}+1, \ldots, \boldsymbol{V}+(k-1)$$

platí $\boldsymbol{W}\!\!\downarrow\!\updownarrow = \boldsymbol{W} - k$. Proto, označíme-li stranu rovnice pozice $\boldsymbol{V}$ jako S_1, máme

$$r(S_1)\,|_{\{\mathrm{Ind}(\boldsymbol{W})\mid \boldsymbol{W}\in r[S_1]\setminus\{\boldsymbol{V}+0,\ldots,V+(k-1)\}\}} =$$

$$= r(S_1)\,|_{\{\mathrm{Ind}(\boldsymbol{W})\mid \boldsymbol{W}\in r[S_1]\setminus\{\boldsymbol{V}-k,\ldots,V-1\}\}}$$

Podle pozorování 3.8 je tedy s řešením. ∎



[img-prekryv]

Obrázek 4.2. Pokrácení překryvu proměnné x

[neprekryv]

Důsledek 4.25. Pro každé řešení r kvadratické soustavy existuje řešení s bez překryvů splňující $s \leq_l r$, $s \leq_c r$. Žádné l-minimální ani c-minimální řešení tedy nemá překryv.

Definice 4.26. Uvažme řešení r kvadratické soustavy $\mathcal{S}$ a netriviální překryv $(\boldsymbol{v}, \boldsymbol{w})$ proměnné x v tomto řešení. Opakovaně odvozujeme menší řešení postupem popsaným výše, dokud má proměnná x netriviální překryv. Takto vzniklému řešení říkáme *řešení vzniklé pokrácením překryvu* $(\boldsymbol{v}, \boldsymbol{w})$.

Definice 4.27. Operace „smazání prázdných proměnných“, „lámání rovnic“, „lámání výskytu v bodě“, „lámání před / za výskytem“, „vylamování výskytů“, „dosazování“ a „pokrácení překryvu“ budeme souhrně označovat jako makro-operace.

4.2 Na které soustavy se stačí zaměřit – pokračování

Pozorování 4.28. Soustava $\mathcal{S}$ má nejvýše tolik netriviálních rovnic, kolik má celkem zlomů a konstant. Pokud tedy $\mathcal{S}$ nemá žádné triviální rovnice, můžeme odhadnout její délku na základě počtu zlomů z a počtu konstant c.

$$z = 2(|\mathcal{S}| - |\langle\!\langle \mathcal{S} \rangle\!\rangle|) \leq 2|\mathcal{S}| - 2(z+c) \Rightarrow |\mathcal{S}| \leq \frac{3}{2}z + c.$$

Pozorování 4.29. Máme-li řešitelnou kvadratickou soustavu $\mathcal{S}$, podosazováním všech možných dosaditelných rovnic získáme kvadratickou soustavu $\mathcal{S}_2$ bez dosaditelných rovnic, která splňuje

- $\mathscr{C}_{\mathcal{S}_2} = \mathscr{C}_{\mathcal{S}}$,
- $\langle\!\langle \mathcal{S}_2 \rangle\!\rangle \leq \langle\!\langle \mathcal{S} \rangle\!\rangle$,
- $\min \mathcal{S}_2 \leq \min \mathcal{S} \cdot |\mathcal{S}|$.

Poslední nerovnost plyne z toho, že pro libovolné řešení soustavy $\mathcal{S}_2$ se při odvození řešení soustavy $\mathcal{S}$ každá proměnná zkopíruje nejvýše $|\mathcal{S}|$-krát.

Důsledek 4.30. Pokud dokážeme polynomiální resp. jednoduše exponenciální mez pro 2-soustavy bez dosaditelných rovnic, budeme mít i polynomiální resp. jednoduše exponenciální mez pro všechny kvadratické soustavy.

[exp-exp-mez]

4.3 Dvojitě exponencionální mez

Pozorování 4.31. Existuje polynom p s následující vlastností. Kdykoli uvážíme pevné množiny $\mathscr{C}$ a $\mathscr{V}$, tak počet všech možných kvadratických soustav, jejíž množina konstant je podmnožina $\mathscr{C}$ a množina proměnných je podmnožina $\mathscr{V}$ nepřevyšuje

$$2^{p|\mathscr{C}\cup\mathscr{V}|}.$$

Pozorování 4.32. Uvažujme soustavu $\mathcal{S}$ s řešením r a z té pomocí makro-operací odvozenou soustavu $\mathcal{T}$ s řešením s. Platí $s \leq_c r$. Dále uvažme dvě řešení s_1, s_2 soustavy $\mathcal{T}$ a zpětně odvozená řešení r_1, r_2 soustavy $\mathcal{S}$. Pokud platí $s_1 \leq_l s_2$ tak platí i $r_1 \leq r_2$. Pokud tedy řešení s bylo l-minimální, bude takové i řešení s.

Poznámka 4.33. Rozšířit předchozí pozorování na další typy uspořádání nemůžeme. Dožadovat se $s \leq_l r$ by z principu nedávalo smysl, protože soustavy $\mathcal{S}$ a $\mathcal{T}$ mají při lámání proměnných různé množiny proměnných. Dále uveďme kvadratickou soustavu s jednou konstantou A.

$$A = xy, \quad x = z, \quad y = u, \quad u = v.$$

Její jediné c-minimální (současně nejmenší) řešení je $x = z = A$, zbylé proměnné jsou prázdné. Dosazením rovnic $u = v$ a $y = u$ odvodíme soustavu

$$A = xv, \quad x = z,$$

s odvozeným řešením $x = z = A$, v je prázdné. To ale není c-minimální řešení, c-menší řešení je takové, kde $v = A$ a x, z jsou prázdné.

Je tu tedy opět drobná nepříjemnost, že při postupném lámáním a dosazováním víme pouze, že řešení klesají v c-uspořádání, avšak minimalitu máme zaručenou pouze co se týče l-uspořádání. S tím se vypořádáme tak, že budeme rovnice lámat pouze specifickým způsobem, při kterém budeme mít kontrolu nad množinou $\mathscr{V}$ a budeme tak odvozovat dokonce řešení, která budou l-menší.

Tvrzení 4.34. Existuje polynom p takový, že pro libovolnou kvadratickou soustavu $\mathcal{S}$ má každé l-minimální řešení délku menší než $2^{2^{p|\mathcal{S}|}}$.

Důkaz: Dokážeme tvrzení pro 2-soustavu, pro zobecnění na všechny kvadratické rovnice stačí použít tvrzení ??.

Začneme s obecnou řešitelnou 2-soustavou $\mathcal{S}_1$ s z zlomy a jejím l-minimálním řešením r_1. Předpokládáme, že v r_1 není žádná proměnná prázdna (když tak smažeme

prázdné proměnné). Postupně odvozujeme 2-soustavy $\mathcal{S}_i$ s l-minimálními řešeními r_i bez prázdných proměnných tak, že vždy bude platit

- $\mathscr{C}_{\mathcal{S}_{i+1}} \subseteq \mathscr{C}_{\mathcal{S}_i}$,
- $\mathscr{V}_{\mathcal{S}_{i+1}} \subseteq \mathscr{C}_{\mathcal{S}_i}$,
- $r_{i+1} <_l r_i$,
- $r_i \leq 2r_{i+1}$.

V některém kroku k skončíme na soustavě $\mathcal{S}_k$ a řešení r_k délky 1. Vzhledem k tomu, že se bude jednat o ostře l-klesající posloupnost l-minimálních řešení, nesmí se v ní žádná 2-soustava opakovat. Různých soustav může být pouze omezeně mnoho, tedy $k \leq 2^{p|\mathcal{S}|}$ pro nějaký polynom nezávislý na $\mathcal{S}$. Poslední vlastnost posloupnosti nám zaručí dvojitě exponenciální mez

$$|r_1| \leq 2^{2^p|\mathcal{S}|}.$$

Zbývá popsat, jak odvozujeme jednotlivé soustavy. Pokud zbyla pouze jedna triviální rovnice, ukončíme činnost. Jinak v každém kroku vybereme jednu rovnici R soustavy $\mathcal{S}_i$. Je-li tato rovnice dosaditelná, dosadíme ji a jsme hotovi. V opačném případě uvážíme výskyty $\boldsymbol{v} = (R, S, 0)$, $\boldsymbol{w} = (R, S', 0)$, kde S, S' jsou různé strany rovnice R. Pokud $|\boldsymbol{v}|_{r_i} = |\boldsymbol{w}|_{r_i}$, rozlomíme rovnici R ve dvojici zlomů $(\boldsymbol{v}, \boldsymbol{v}+1)$ a $(\boldsymbol{w}, \boldsymbol{w}+1)$. Následně dosadíme vzniknuvší triviální rovnici a jsme opět hotovi.

Zbývá možnost, kdy mají $\boldsymbol{v}$ a $\boldsymbol{w}$ v r_i různé délky. Pak BÚNO $|\boldsymbol{v}|_r < |\boldsymbol{w}|_r$. V takovém případě rozlomíme rovnici za výskytem $\boldsymbol{v}$. To rozdělí proměnnou na dvě – xy. Pro proměnnou y použijeme jakožto prvek množiny $\mathscr{V}_{\mathcal{S}_i}$ odstraněnou proměnnou $\overline{\boldsymbol{w}}$ a provedeme dosazení výskytu $\boldsymbol{v}\updownarrow$ proměnné x, čímž tato nová proměnná zmizí a my dosáhneme požadované vlastnosti. ■

4.4 Co by stačilo pro jednoduše exponenciální mez – pokračování

[omez-vysk] **Tvrzení 4.35.** Jednoduše exponenciální mez pro kvadratické soustavy bychom měli již za následujícího předpokladu: „Existuje rostoucí polynom p s následující vlastností. Pro každou 2-soustavu $\mathcal{S}$ existuje její konstanta A a řešení r, že $\mathrm{Freq}_{r,\mathcal{S}}(A) \leq 2^{p|\mathcal{S}|}$." Jinými slovy stačí dokázat jednoduše exponenciální mez pro frekvence nejméně frekventované proměnné.

Důkaz: Stačí dokázat jednoduše exponenciální mez pro 2-soustavy bez dosaditelných rovnic. Vezměme si tedy takovou 2-soustavu $\mathcal{S}$, počet zlomů v ní označme n. Dále označme $\mathcal{S}_{|\mathscr{C}_\mathcal{S}|} = \mathcal{S}$ Popíšeme postup, jak z 2-soustavy $\mathcal{S}_i$ vytvořit soustavu $\mathcal{S}_{i-1}$ opět bez dosaditelných rovnic a s i konstantami. Tento postup budeme provádět, až do $\mathcal{S}_0$.

Z předpokladu máme řešení r soustavy $\mathcal{S}_i$ a konstantu A, že $\mathrm{Freq}_{r,\mathcal{S}_i}(A) \leq 2^{p|\mathcal{S}_i|}$. Označme $k_i = \mathrm{Freq}_{r,\mathcal{S}_i}(A)$. Na všechny pozice konstanty A v řešení tuto konstantu vepíšeme do soustavy – tím vznikne k nových proměnných a konstanta A se již v rovnici nebude vyskytovat dvakrát, ale $2k$-krát. Nová rovnice tak bude mít o $4k$ zlomů více. Dále podle řešení r kolem všech těchto konstant A rovnici rozdělíme a rovnice $A = A$ smažeme. Tím dostaneme soustavu, kterou označíme $\mathcal{T}_i$. Počet zlomů v soustavách $\mathcal{S}_i$ a $\mathcal{T}_i$ je stejný, jakkoli T_i může mít oproti S_i exponencionální množství rovnic. Nakonec definujeme S_{i-1} jako rovnici, která vznikne z T_i podosazováním dosaditelných rovnic.

Absence triviálních rovnic v S_i dává

$$\langle\!\langle \mathcal{S}_i \rangle\!\rangle \leq n, \quad |\mathcal{S}_i| = n/2 + \langle\!\langle \mathcal{S}_i \rangle\!\rangle \leq 2n.$$

Tedy $\langle\!\langle \mathcal{T}_i \rangle\!\rangle \leq n + 2^{p(2n)} \leq 2^{p_2(n)}$ pro nějaký rostoucí polynom p_2 nezávislý na n. Kdykoli vezmeme řešení r soustavy S_{i-1} a odvodíme z něj řešení r' soustavy T_i, bude tak $|r'|_{\mathcal{T}_i} \leq$

$2^{p_2(n)} \cdot |r|_{\mathcal{S}_{i-1}}$). Každá proměnná z S_{i-1} se totiž zkopíroje nejvýše $\langle\!\langle \mathcal{T}_i \rangle\!\rangle$-krát. Z toho plyne

$$\min(\mathcal{S}_i) \leq k_i + \min(\mathcal{T}_i) \leq 2^{p_2(n)} \cdot (\min \mathcal{S}_{i-1} + 1).$$

Rovnice S_0 má nulové řešení, takže indukcí snadno dostáváme

$$\min(\mathcal{S}_i) \leq 2^{p_2(n)\cdot 1} + \cdots + 2^{p_2(n)\cdot i}.$$

Celkem

$$\min(\mathcal{S}) \leq |\mathcal{S}| \cdot 2^{p_2(n)\cdot|\mathcal{S}|} \leq |\mathcal{S}| \cdot 2^{p_2(2|\mathcal{S}|)\cdot|\mathcal{S}|} \leq 2^{p_3|\mathcal{S}|}$$

pro nějaký polynom p_3 nezávislý na $\mathcal{S}$. ∎

[len-to-eq]

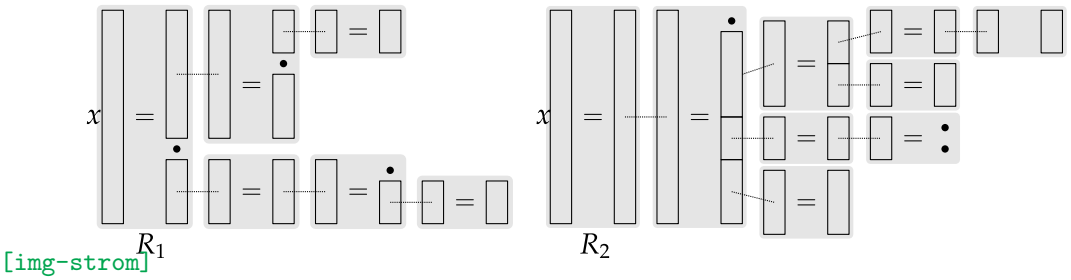
Tvrzení 4.36. Nechť je dána řešitelná 2-soustava $\mathcal{S}$ a její proměnná x. Pak existuje řešitelná 2-rovnice $\mathcal{R}$ (tj. soustava o jedné rovnici) a řešení r soustavy $\mathcal{S}$ splňující

- $\mathscr{C}_\mathcal{R} \subseteq \mathscr{C}_\mathcal{S}$,
- $|\mathcal{R}| \leq |\mathcal{S}|$,
- $|x|_r = \min \mathcal{R}$.

Důkaz: Uvažme libovolné řešení r_1 soustavy $\mathcal{S}$. Z něj následujícím postupem nezvyšujícím počet zlomů ani celkovou délku řešení odvodíme řešení r_2 soustavy $\mathcal{S}_2$. Nejprve odstraníme z rovnice všechny proměnné, které mají v r_1 nulovou délku. Pak v tomto řešení vylomíme oba výskyty proměnné x (pokud se tyto výskyty před tím překrývaly, tak je pokrátíme). Vzniklé dvě rovnice tvaru $x = \ldots$ označíme R_1, R_2 a orientujeme je tak, aby x bylo nalevo. Ještě (pro jistotu) smažeme prázdné proměnné, takže v dalším postupu již budou všechny proměnné neprázdné. Pokud by byla prázdná proměnná x, snadno vytvoříme kýženou rovnici $\mathcal{R}$, tedy předpokládáme, že x jsme nesmazali a je neprázdné. Dále budeme budovat dva zakořeněné stromy rovnic, jejichž kořeny budou R_1 a R_2. Každý prvek stromu bude tvaru

$$y = \ldots$$

pro nějakou proměnnou y a druhý výskyt proměnné y bude na pravé straně rodiče tohoto prvku. Tyto stromy budujeme jednoduše tak, že v každém kroku vybereme proměnnou y, jejíž jeden výskyt je v některém prvku stromu P a druhý výskyt $\boldsymbol{v}$ je mimo oba stromy. Vylomíme tedy výskyt $\boldsymbol{v}$ a rovnici „$y = \ldots$“ z toho vzešlou zařadíme do stromu za prvek P. Takto postupujeme, dokud můžeme. Tedy na konci máme soustavu $\mathcal{S}_2$, řešení r_2 a dva výše popsané zakořeněné stromy s vrcholy na rovnicích $\mathcal{S}_2$, že každá proměnná má buď oba své výskyty v těchto stromech nebo oba mimo ně.



[img-strom]

Obrázek 4.3. Ukázka sestavených stromů na základě postupu v důkazu

Uvažme soustavu $\mathcal{S}_3$ sestávající se pouze z rovnic z tohoto stromu – to je 2-soustava. Po podosazování všech dosaditelných rovnic v $\mathcal{S}_3$ dostáváme 2-rovnici $\mathcal{R}$ jejíž obě strany se mají rovnat již redukované proměnné x. První požadovaná vlastnost na rovnici $\mathcal{R}$ z tvrzení je splněna zřejmě, druhá plyne z toho, že $\mathcal{R}$ má nejvýše tolik zlomů i rovnic, co

soustava $\mathcal{S}$ Zbývá ukázat třetí vlastnost – tedy najít řešení r. Uvažme nejmenší řešení r_4 rovnice $\mathcal{R}$ a řešení r_3 jako řešení soustavy S_3 odvozené z r_4. To znamená

$$\min(\mathcal{R}) = |r_4|_{\mathcal{R}} = |x|_r.$$

Nyní v řešení r_2 použijeme u všech proměnných soustavy $\mathcal{S}_3$ hodnoty z řešení r_3 namísto původních hodnot z r_2, tím dostaneme řešení r_2' soustavy $\mathcal{S}_2$, ve kterém $|x|_{r_2} = \min(\mathcal{R})$. Odtud již jen zpětně odvodíme řešení soustavy $\mathcal{S}$ s touto vlastností. ■

Důsledek 4.37. Pokud bychom měli jednoduše exponenciální mez na nejmenší délku řešení pro řešitelné 2-rovnice, měli bychom z věty 3.22 i jednoduše exponenciální mez pro všechny řešitelné 2-soustavy, tedy i pro všechny kvadratické soustavy.

Důsledek 4.38. Pokud bychom měli jednoduše exponenciální mez na nejmenší možnou velikost nejmenší proměnné v řešení pro 2-rovnice, měli bychom na základě opětovného využití tvrzení 3.22 i jednoduše exponenciální mez pro všechny kvadratické soustavy.

[dualita]

Kapitola 5

Dualita mikro a makro přístupu

5.1 Motivace a intuice

Jak je patrno z předchozích dvou kapitol, postupy v mikro-přístupu jsou v jistém smyslu analogické k postupům v makro-přístupu, jakkoli oba přístupy vzešly ze značně odlišných úvah. Tabulka 5.1 zobrazuje pojmy a tvrzení z obou kapitol, které si odpovídají.

mikro-přístup	makro-přístup
délka proměnné	frekvence konstanty
l-uspořádání	*c*-uspořádání
c-uspořádání	*l*-uspořádání
prázdná proměnná	konstanta s jednou pozicí
počet proměnných	počet zlomů
konstanty	rovnice
slepené konstanty	triviální/dosaditelná rovnice
slití konstant	dosazení rovnice
vepsání a rozrůznění konstanty	rozpůlení proměnné
opakující se konstanta	překryv
tvrzení 3.22	tvrzení 4.35
tvrzení 3.23	tvrzení 4.36

[tab-dualita]

Tabulka 5.1. Intuitivně duální pojmy na základě mikro/makro analogie

Tato kapitola vysvětluje, kde se zde tato dualita bere a pro jistou třídu kvadratických rovnic zavádí pojem duální kvadratické rovnice – pro takzvané orientované 2-soustavy.

Definice 5.1. O 2-soustavě říkáme, že je orientovaná, pokud se každá proměnná vyskytuje právě jednou na levé straně některé rovnice a právě jednou na pravé. Na základě toho dělíme výskyty proměnných na *pravé* a *levé*.

Pozorování 5.2. Je-li výskyt $\boldsymbol{v}$ na levé straně rovnice, pak $\boldsymbol{v}^{\curvearrowleft\wr}$ je na pravé straně a obráceně.

Důsledek 5.3. V orientované 2-rovnici je každá konstanta právě jednou na levé straně některé rovnice a právě jednou na pravé.

Řešení r orientované 2-soustavy $\mathcal{S}$ je monoidový homomorfismus $(\mathscr{V}_\mathcal{S} \cup \mathscr{C}_\mathcal{S})^* \to \mathscr{C}_\mathcal{S}^*$ zachovávající konstanty. Na základě něj můžeme zavést „duální řešení“ r^T jakožto homomorfismus $(\mathscr{V}_\mathcal{S} \cup \mathscr{C}_\mathcal{S})^* \to \mathscr{V}_\mathcal{S}^*$ zachovávající proměnné. Přitom pro libovolnou konstantu A definujeme

$$r^T(A) = \overline{\mathrm{Src}(\boldsymbol{v}\updownarrow)}\,\overline{\mathrm{Src}(\boldsymbol{v}\updownarrow\!\upharpoonright\!\downarrow\updownarrow)}\,\ldots\,\overline{\mathrm{Src}(\boldsymbol{v}^{\curvearrowleft\wr}\updownarrow\!\upharpoonright\!\downarrow)}$$

Duální řešení bychom tedy měli, ale pro jakou rovnici je řešením?

5.2 2D rovnice

Definice 5.4. Nechť jsou dány dvě stejně mohutné disjunktní konečné množiny konstant $\mathscr{C}$ a proměnných $\mathscr{V}$. Pak 2D rovnicí na slovech rozumíme uspořádanou šestici

$$(\mathscr{C}^{\downarrow}, \mathscr{V}^{\rightarrow}, \mathscr{C}^{\uparrow}, \mathscr{V}^{\leftarrow}, \updownarrow, \approx),$$

kde první čtyři prvky jsou konečné neprázdné disjunktní množiny, $\updownarrow$ je postfixově značená bijekce množiny $(\mathscr{C}^{\downarrow} \cup \mathscr{V}^{\rightarrow} \cup \mathscr{C}^{\uparrow} \cup \mathscr{V}^{\leftarrow})$ do sebe, $\approx$ je symetrická reflexivní relace na té samé množině, a navíc platí následující požadavky.

- Bijekce $\updownarrow$ „posílá na opačný konec“, tedy
  - Pro $\alpha \in \mathscr{C}^{\downarrow}$ je $\alpha\updownarrow \in \mathscr{C}^{\uparrow}$,
  - Pro $\alpha \in \mathscr{V}^{\rightarrow}$ je $\alpha\updownarrow \in \mathscr{V}^{\leftarrow}$,
  - Pro $\alpha \in \mathscr{C}^{\uparrow}$ je $\alpha\updownarrow \in \mathscr{C}^{\downarrow}$,
  - Pro $\alpha \in \mathscr{V}^{\leftarrow}$ je $\alpha\updownarrow \in \mathscr{V}^{\rightarrow}$,
- Pro $\alpha \in \mathscr{C}^{\downarrow}$
  - existuje právě jedno $\beta \in \mathscr{C}^{\downarrow}$ splňující $\beta \approx \alpha$, konkrétně je $\beta = \alpha$,
  - existuje právě jedno $\beta \in \mathscr{V}^{\leftarrow}$ splňující $\beta \approx \alpha$,
  - existuje právě jedno $\beta \in \mathscr{V}^{\rightarrow}$ splňující $\beta \approx \alpha$,
  - neexistuje žádné $\beta \in \mathscr{C}^{\uparrow}$, které by splňovalo $\beta \approx \alpha$.
- Předchozí bod platí i pro zbylé 3 orientace šipek (tj. když se čtveřicí množin $(\mathscr{C}^{\uparrow}, \mathscr{V}^{\rightarrow}, \mathscr{C}^{\downarrow}, \mathscr{V}^{\leftarrow})$ provádíme cyklické záměny).

Z historických důvodů budeme prvkům množiny $\mathscr{C}^{\downarrow}$ říkat konstanty a prvkům množiny $\mathscr{V}^{\rightarrow}$ budeme říkat proměnné.

Definice 5.5. Velikostí 2D rovnice rozumíme celkový počet proměnných a konstant, tedy $|\mathscr{C}^{\downarrow} + \mathscr{V}^{\rightarrow}| = 2|\mathscr{V}^{\rightarrow}|$.

Definice 5.6. Řešením 2D rovnice rozumíme sedmici

$$(X, \pi_{\mathscr{C}}, \pi_{\mathscr{V}}, \downarrow, \rightarrow, \uparrow, \leftarrow),$$

kde X je konečná množina, $\pi_{\mathscr{C}}$, $\pi_{\mathscr{V}}$ jsou zobrazení a $\downarrow, \rightarrow, \uparrow, \leftarrow$ jsou postfixově značená zobrazení. Rozšiřme relaci $\approx$ jakožto minimální reflexivní ještě na množinu X. Řešení pak musí splňovat ještě následující podmínky.

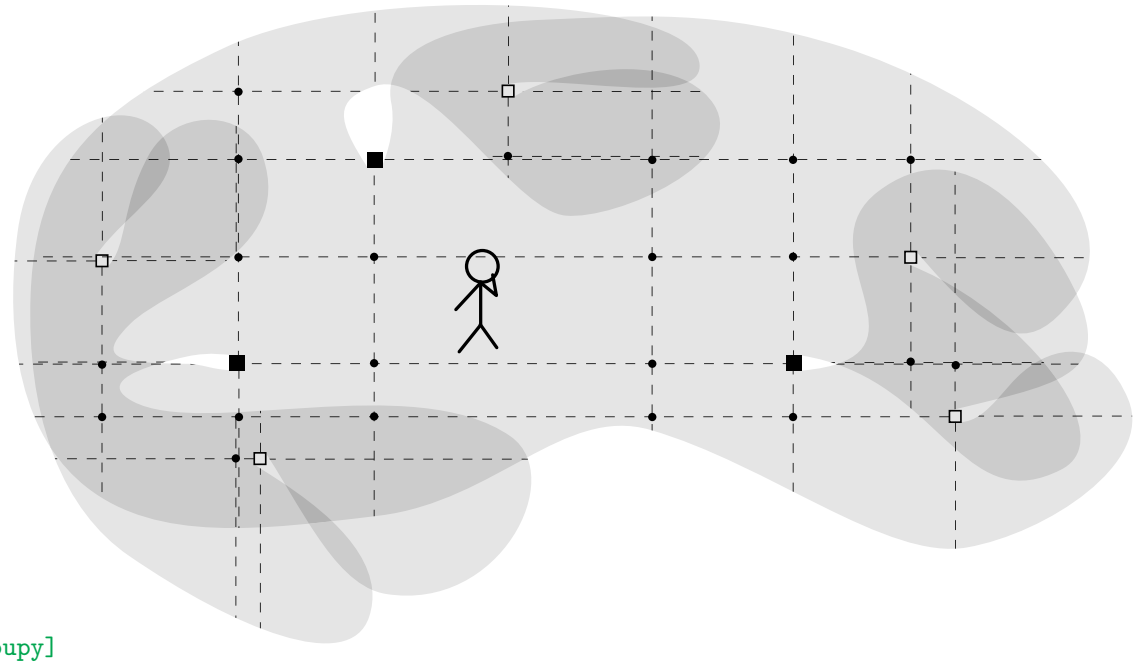
- Výše uvedená zobrazení jsou definována s těmito doménami a kodoménami:
  - $\pi_{\mathscr{V}} : (X \cup \mathscr{V}^{\rightarrow} \cup \mathscr{V}^{\leftarrow}) \to \mathscr{V}^{\rightarrow}$,
  - $\pi_{\mathscr{C}} : (X \cup \mathscr{C}^{\downarrow} \cup \mathscr{C}^{\uparrow}) \to \mathscr{C}^{\downarrow}$,
  - $\downarrow : (X \cup \mathscr{C}^{\downarrow} \cup \mathscr{V}^{\leftarrow} \cup \mathscr{V}^{\rightarrow}) \to (X \cup \mathscr{C}^{\uparrow})$,
  - $\rightarrow : (X \cup \mathscr{V}^{\rightarrow} \cup \mathscr{C}^{\downarrow} \cup \mathscr{C}^{\uparrow}) \to (X \cup \mathscr{V}^{\leftarrow})$,
  - $\uparrow : (X \cup \mathscr{C}^{\uparrow} \cup \mathscr{V}^{\rightarrow} \cup \mathscr{V}^{\leftarrow}) \to (X \cup \mathscr{C}^{\downarrow})$,
  - $\leftarrow : (X \cup \mathscr{V}^{\leftarrow} \cup \mathscr{C}^{\uparrow} \cup \mathscr{C}^{\downarrow}) \to (X \cup \mathscr{V}^{\rightarrow})$.
- Kdykoli je definováno
  - $x\downarrow$, platí $x\downarrow\uparrow \approx x$,
  - $x\rightarrow$, platí $x\rightarrow\leftarrow \approx x$,
  - $x\uparrow$, platí $x\uparrow\downarrow \approx x$,
  - $x\leftarrow$, platí $x\leftarrow\rightarrow \approx x$.
- Je-li definováno $x\downarrow$, platí $x\downarrow\rightarrow\uparrow\leftarrow \approx x$.

- Zobrazení $\pi_{\mathscr{C}}$ splňuje podmínky:
  - pro všechna $\alpha \in \mathscr{C}^{\downarrow}$ je $\pi_{\mathscr{C}}(\alpha) = \alpha$,
  - pro všechna $\alpha \in \mathscr{C}^{\uparrow}$ je $\pi_{\mathscr{C}}(\alpha) = \alpha\updownarrow$,
  - pro všechna $x \in X \cup \mathscr{C}^{\uparrow}$ je $\pi_{\mathscr{C}}(x) = \pi_{\mathscr{C}}(x\uparrow)$.
- Zobrazení $\pi_{\mathscr{V}}$ splňuje podmínky:
  - pro všechna $\alpha \in \mathscr{V}^{\rightarrow}$ je $\pi_{\mathscr{V}}(\alpha) = \alpha$,
  - pro všechna $\alpha \in \mathscr{V}^{\leftarrow}$ je $\pi_{\mathscr{V}}(\alpha) = \alpha\updownarrow$,
  - pro všechna $x \in X \cup \mathscr{V}^{\leftarrow}$ je $\pi_{\mathscr{V}}(x) = \pi_{\mathscr{V}}(x\leftarrow)$.

Velikost řešení chápeme jako mohutnost množiny X.

Poznámka 5.7. Situaci je možné si představovat takto: 2D rovnice nám definuje sadu „kouzelných" sloupů s čtvercovou podstavou. Kouzelnost sloupů spočívá ve vlastnosti, že obejitím sloupu se octneme na jiném místě. Sloupy je možné chápat jako třídy ekvivalence, která vznikne rozšířením relace $\approx$.

Řešením 2-rovnice se pak rozumí rozmístění těchto sloupů do konečného (přesněji řečeno periodického) světa tak, aby když rovně vyrazíme od stěny s nápisem α, dojdeme k jiné stěně s nápisem $\alpha\updownarrow$. Prvky množiny X jsou pak průsečíky takových cest.



[img-sloupy]

Obrázek 5.1. Umělecké ztvárnění světa řešení 2D rovnice, čtverečky značí sloupy, puntíky průsečíky cest od sloupů

Definice 5.8. Je-li dána 2D rovnice $\mathcal{R}$ a její řešení r, tak duální rovnici s duálním řešením značíme $\mathcal{R}^T$, r^T a vyrobíme je

- výměnou množin $\mathscr{C}^{\downarrow}$ a $\mathscr{V}^{\rightarrow}$,
- výměnou množin $\mathscr{V}^{\leftarrow}$ a $\mathscr{C}^{\uparrow}$,
- výměnou zobrazení $\pi_{\mathscr{C}}$ a $\pi_{\mathscr{V}}$,
- výměnou zobrazení $\leftarrow$ a $\uparrow$,
- výměnou zobrazení $\downarrow$ a $\rightarrow$.

Definice 5.9. délka proměnné resp. konstanty v 2D rovnici

Definice 5.10. 2D rovnice s prázdnými cestami a její řešení

Definice 5.11. *Složitostí* 2D rovnice s prázdnými cestami myslíme počet všech dvojic (x, A), kde $x \in \mathscr{V}^{\rightarrow} \cup \mathscr{V}^{\leftarrow}$, $A \in \mathscr{C}^{\downarrow} \cup \mathscr{C}^{\uparrow}$, platí $x \approx A$ a přitom ani x ani A není prázdné.

5.3 Souvislost s orientovanými 2-rovnicemi

Tvrzení 5.12. Z řešení orientované 2-soustavy umíme sestrojit 2D rovnici s prázdnými cestami a její řešení.

Tvrzení 5.13. Z řešení 2D rovnice s prázdnými cestami umíme sestrojit orientovanou 2-soustavu a její řešení.

Definice 5.14. Duální kvadratická rovnice vytvořená na základě řešení.

Důsledek 5.15. V orientovaných 2-soustavách existuje dvojitě exponenciální mez na c-minimální řešení.

Poznámka 5.16. Požadavek na orientovanost rovnice je ve 2D rovnicích pro to, abychom mohli orientovat cesty konstant shora dolů. Od tohoto požadavku by bylo možné upustit, pokud bychom nelpěli na orientované mřížce a připustili neorientovanou. Takovou „neorientovanou 2D rovnici“ bychom nemohli po dualizování převést na obyčejnou kvadratickou rovnici na slovech, ale mohli bychom ji převést na takovou, kde se místo některých proměnných vyskytuje jejich „zrcadlový obraz“. Tedy důsledek by bylo možné s jistým množstvím péče dokázat i pro obecné rovnice.

[slozitost]

Kapitola 6

Složitost řešení kvadratických soustav

6.1 Kvadratické soustavy jsou NP-těžké

Abychom ukázali, že řešitelnost kvadratických soustav je NP těžký problém, převedeme ji na NP-úplný problém – existenci nezávislé množiny v grafu[1]. Konkrétně ukážeme, že existuje algoritmus s polynomiální složitostí, který dostane na vstupu graf a číslo k a na základě toho sestaví kvadratickou soustavu, která je řešitelná právě když v grafu existuje nezávislá množina velikosti k.

Algoritmus funguje takto:

Označme V množinu vrcholů grafu, její velikost n a vrcholy budeme značit postupně $v_1, \ldots, v_n$. Sestrojíme kvadratickou soustavu s dvoubodovou množinou konstant $\mathscr{C} = \{A, B\}$. Dále pro každý vrchol v založíme proměnnou a_v, pro každou uspořádanou dvojici (v, w) různých vrcholů založíme proměnnou $b_{v,w}$. Nakonec ještě použijeme několik proměnných, které se budou v soustavě vyskytovat pouze jednou – každý výskyt takové proměnné budeme značit symbolem $\star$.

Pro každý vrchol v sestavíme rovnici

[vrchol-rovnice]

$$B^n A = b_{v,v_1} \ldots b_{v,v_n} a_v \star, \tag{1}$$

dále pro každou neuspořádanou dvojici vrcholů $\{v, w\}$ (je jedno, jak ji uspořádáme) za předpokladu, že mezi v a w vede hrana, sestavíme rovnici

[hrana-rovnice]

$$B = b_{v,w} b_{w,v} \star, \tag{2}$$

Pokud mezi těmito vrcholy naopak hrana nevede, sestavíme dvojici rovnic

[nehrana-rovnice]

$$B = b_{v,w} \star, \tag{3}$$
$$B = b_{w,v} \star.$$

A nakonec sestavíme rovnici

[nez-mnoz-rovnice]

$$A^k = a_{v_1} a_{v_2} \ldots a_{v_n}. \tag{4}$$

Pozorování 6.1. Sestavená rovnice je kvadratická. Každé a_v se totiž vyskytuje právě jednou v rovnici (4) a právě v jedné rovnici (1). Podobně každé $b_{v,w}$ se vyskytuje v právě jedné rovnici (1) a v jedné z rovnic (2) nebo (3).

Tvrzení 6.2. Existovala-li v grafu nezávislá množina N, pak je sestavená rovnice řešitelná.

Důkaz: Stačí volit $a_v = A$ pro všechna $v \in N$, $b_{v,w}$ pro všechna $v \in N$ a $w \in V$. Ostatní a_v a $b_{v,w}$ volíme prázdné. Proměnné $\star$ vhodně doplníme buď prázdné nebo totožné s levou stranou rovnice. ∎

Tvrzení 6.3. Je-li sestavená rovnice řešitelná, pak v grafu existuje nezávislá množina velikosti k.

Důkaz: Označme r řešení rovnice. Na základě rovnic (1) může každé $r(a_v)$ obsahovat nejvýše jedno A. Dále podle rovnice (4) se $r(a_v)$ mohou skládat pouze z konstant A, proto

[1] Ve zdroji [1] je ukázáno převedení na problém 3-SAT v podobném duchu.

$r(a_v) = A$ pro jistou množinu $N \subseteq V$, zbylá $r(a_v)$ jsou prázdná. Navíc podle téže rovnice $|N| = k$.

Ukážeme, že N je nezávislá množina, uvažujme dva vrcholy $v, w \in N$. Pak na základě rovnice (1) musí mít slovo $r(b_{v,v_1} b_{v,v_2} \dots b_{v,v_n})$ délku alespoň n. Navíc podle rovnic (2) resp. (3) je každá proměnná $b_{x,y}$ nejvýše jednoprvková, proto pro všechna $x \in V$ je $r(b_{v,x}) = B$ a podobně i $r(b_{w,x}) = B$. Tedy $r(b_{v,w} b_{w,v}) = BB$ a proto díky rovnici (2) mezi těmito vrcholy nevede hrana. ∎

Poznámka 6.4. Bylo by možné sestavením ještě další rovnice soustavu doplnit, aby se v ní každá proměnná vyskytovala právě jednou a přesto aby splňovala požadované vlastnosti. Na druhou stranu předvedený postup značně využívá skutečnosti, že v rovnici nejsou rozrůzněné konstanty, tedy není jasné, zda i řešitelnost 2-rovnic je NP-těžký problém.

6.2 Algoritmus pro ověření řešitelnosti rovnice s předepsanými délkami

Pokud bychom dostali kvadratickou soustavu a předepsané délky proměnných, mohli bychom v lineárním čase vzhledem k velikosti řešení ověřit, zda existuje řešení, ve kterém mají proměnné onu předepsanou délku – jednoduše projdeme jejich cesty, najdeme protějšky konstant a ověříme, zda si stejné konstanty odpovídají. Takový postup by byl lineární vzhledem k velikosti tohoto řešení. Ukážeme ale, že je možné postupovat ještě efektivněji – totiž polynomiálně vzhledem k velikosti soustavy a logaritmu velikosti řešení.

Definice 6.5. *Kvadratickou soustavou s předepsanými délkami proměnných* (zkráceně *KSPDP*) rozumíme dvojici $(\mathcal{S}, \varphi)$, kde $\mathcal{S}$ je kvadratická soustava a φ je monoidový homomorfismus z $(\mathcal{V}_\mathcal{S} \cup \mathcal{C}_\mathcal{S})^*$ do $\mathbb{N}_0$ se sčítáním, přičemž obrazem každé konstanty je 1. *Řešením KSPDP* rozumíme takové řešení soustavy $\mathcal{S}$, které pro každou proměnnou x splňuje $|r(x)| = \varphi(x)$. Říkáme, že soustava s předepsanými délkami je řešitelná, pokud má řešení. Délkou proměnné v KSPDP rozumíme právě její obraz v zobrazení φ.

Definice 6.6. O KSPDP $(\mathcal{S}, \varphi)$ říkáme, že je *smysluplná*, pokud pro každou její rovnici $S_1 = S_2$ platí $\varphi(S_1) = \varphi(S_2)$ Pro smysluplnou KSPDP můžeme mluvit o pozicích v řešení a operacích na nich, stejně tak o makro-operacích s rovnicí (i bez znalosti řešení, dokonce i pro neřešitelné). Definice pozic a makro-operací se totiž opírá pouze o délky proměnných v řešení a nevyužívá jeho dalších vlastností. Po provedení makro-operace dostaneme opět smysluplnou KSPDP.

Pozorování 6.7. Pokud není KSPDP smysluplná, není ani řešitelná.

Pozorování 6.8. Pokud provedeme makro-operaci na smysluplnou KSPDP $\mathcal{P}$, čímž dostaneme smysluplnou KSPDP $\mathcal{P}'$, bude $\mathcal{P}'$ řešitelná právě když byla $\mathcal{P}$ řešitelná.

Definice 6.9. Délkou KSPDP $(\mathcal{S}, \varphi)$ rozumíme výraz

$$|\mathcal{S}| + \sum_{\substack{x \in \mathcal{V}_\mathcal{S} \\ \varphi(x) \neq 0}} \log_2 \phi(x).$$

Tvrzení 6.10. Existuje algoritmus, který v polynomiálním čase (v závislosti na délce) ověří řešitelnost dané KSPDP. Tento algoritmus na vstupu dostane soustavu a pro každou proměnnou její délku zapsanou ve dvojkové soustavě. Na výstupu odpoví zda příslušná KSPDP je nebo není řešitelná.

Důkaz: Nejprve ověříme, zda je KSPDP smysluplná. Pokud ne, rovnou známe negativní odpověď. Dále tedy budeme předpokládat tuto vlastnost.

Délku KSPDP je možné chápat jako délku vstupu algoritmu. S délkami proměnných je možné v polynomiálním čase provádět aritmetické operace, tedy (konkrétně sčítání, odčítání, zbytek po dělení). Tedy i makro-operace je možné provádět v polynomiálním čase.

Na začátku ze soustavy smažeme prázdné proměnné. Dále budeme soustavu měnit pomocí makro operací, tedy prázdné proměnné se v ní již neobjeví a její řešitelnost se nezmění.

Algoritmus postupuje podle následujících bodů

1) Dokud existují dosaditelné rovnice, dosazuj je.
2) Najdi v soustavě rovnici R se stranami S_1, S_2, pro níž je $\varphi(S_1)$ nejvyšší možné.
3) Je-li $\varphi(S_1) = 1$, znamená to pouze, že v soustavě zbyly pouze rovnice, které mají na obou stranách jednu konstantu, tj. rovnice tvaru $A = B$. V tuto chvíli tedy stačí projít tyto rovnice a ověřit jejich platnost. Na základě toho odpověz na otázku řešitelnosti a ukonči činnost.
4) Jinak uvaž výskyt $\boldsymbol{v}$ jakožto zdroj pozice $(R, S_1, \lfloor \varphi(S_1)/2 \rfloor)$.
5) Má-li proměnná $\overline{\boldsymbol{v}}$ netriviální překryv, pokraťjej.
6) Vylom výskyt $\boldsymbol{v}$.
7) Vraťse na bod (1).

Zbývá si rozmyslet, že tento algoritmus skutečně běží v polynomiálním čase. Snadno nahlédneme, že každý krok skutečně probíhá v polynomiálním čase vzhledem k aktuální velikosti KSPDP. Tato velikost v průběhu může růst, nicméně je možné ji polynomiálně omezit vzhledem k tomu, že nezvyšujeme počet zlomů ani součet délek všech proměnných a po každém kroku (1) nemáme v soustavě žádné triviální rovnice.

Cyklem algoritmu označme postupné provádění kroků od bodu (1) po bod (6). Zbývá ukázat, že počet cyklů algoritmu je možné polynomiálně omezit.

Pro KSPDP $\mathcal{P} = (\mathcal{S}, \varphi)$ a přirozené číslo n označme $\langle\!\langle \mathcal{P} \rangle\!\rangle_n$ jako počet rovnic $S_1 = S_1$, pro které platí $\varphi(S_1) \leq n$.

Nakonec složitostí KSPDP $\mathcal{P}$ označme nejvyšší k, pro které $\langle\!\langle \mathcal{P} \rangle\!\rangle_{2^k} > 0$. Složitost KSPDP změříme při každém provádění kroku (2).

Pro dokončení důkazu stačí již jen několik pozorování:

- Složitost KSPDP je polynomiálně omezená délkou vstupní KSPDP.
- V každém kroku (2) je $\langle\!\langle \mathcal{P} \rangle\!\rangle_{2^k}$ menší nebo rovno počtu zlomů vstupní soustavy.
- Rovnice vznikající v kroku (6), které nejsou dosaditelné, budou mít ve zobrazení φ nejvýše poloviční délku oproti původní S_1. Díky tomu složitost aktuální KSPDP neroste a za předpokladu, že zůstane stejná, se sníží $\langle\!\langle \mathcal{P} \rangle\!\rangle_{2^k}$ o jedna.

Máme tak polynomiální omezení na počet cyklů mezi jednotlivými sníženími složitosti a současně polynomiální omezení na počet těchto snížení tedy i polynomiální omezení celkového algoritmu. ∎

Důsledek 6.11. Pokud platí hypotéza 2.19 o jednoduše exponenciální mezi, je řešitelnost kvadratických soustav NP-úplný problém.

Poznámka 6.12. Uvedený algoritmus má při důkladnějším zkoumání kvadratickou složitost, zdroji [1] uvádí dokonce algoritmus běžící v lineárním čase na základě optimalizované volby rovnic, které lámeme. Ve snaze o vyšší přehlednost zde předvádíme algoritmus pomalejší, avšak co se týče důsledků zcela postačující.

[palindrom]

Kapitola 7

O jisté konkrétní palindromické rovnici

Účelem této kapitoly je poskytnout sadu (proti)příkladů a na konkrétní rovnici demonstrovat, jak se kvadratické rovnice mohou chovat.

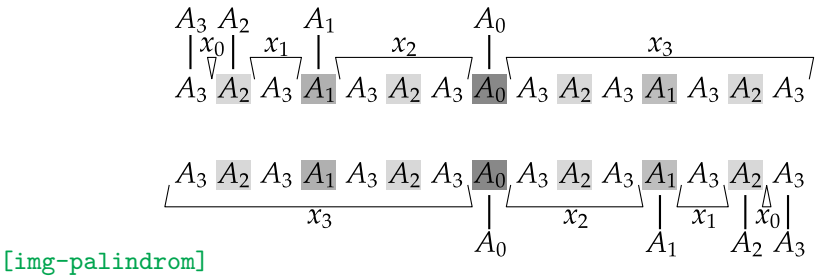
Definice 7.1. Pro dané $n \in \mathbb{N}$ definujeme 2-rovnici Pal(n) jako

$$A_{n-1}x_0A_{n-2}x_1 \dots A_0x_{n-1} = x_{n-1}A_0x_{n-2}A_1 \dots x_0A_{n-1},$$

kde $A_0, \dots, A_{n-1}$ jsou konstanty a $x_0, \dots, x_{n-1}$ jsou proměnné. Dále definujeme PalSol(n) jako řešení, ve kterém je proměnná x_0 prázdná a pro ostatní $i \in \{1, \dots, n-1\}$ je

$$\mathrm{PalSol}(()x_i) = x_{i-1}A_{n-i}x_{i-1}.$$

Pozorování 7.2. Zobrazení PalSol(n) je opravdu řešením rovnice Pal(n) a má délku $2^n - 1$. Dále $|\mathrm{PalSol}(n)(x_i)| = 2^i - 1$



[img-palindrom]

Obrázek 7.1. Rovnice Pal(4) s řešením PalSol(4)

Pozorování 7.3. Pro $n \geq 3$ není PalSol(n) nejmenším řešením, můžeme totiž definovat řešení r délky $2n$ jako

$$r(x_i) = A_i.$$

[palindrom-mikro]

7.1 Mikro-chování Pal(n)

Tvrzení 7.4. Pro libovolné n je řešení PalSol(n) l-minimální. Dokonce, uspořádáme-li řešení r abecedním uspořádáním podle n-tic

$$(|x_0|_r, |x_1|_r, \dots, |x_{n-1}|_r),$$

bude PalSol(n) nejmenší v tomto uspořádání.

Důkaz: Stačí ukázat druhou část tvrzení, první z ní okamžitě plyne. Dokážeme to indukcí podle n, pro $n = 1$ tvrzení zřejmě platí. Uspořádání na řešeních z tvrzení nazvěme l-abecední a pro $n \geq 2$ uvažme řešení $r_{\min}$, které je nejmenší l-abecedním uspořádání. Vzhledem k tomu, že PalSol(n)(x_0) je prázdné musí i $r_{\min}(x_0)$ být prázdné. Uvažujme tedy rovnici $\mathcal{R}$, která vznikne z Pal(n) smazáním proměnné x_0. Řešení PalSol(n)(x_0) i $r_{\min}$ je tak možné vnímat jako řešení $\mathcal{R}$.

Dále budeme zkoumat pozice v obecném řešení r. Na levé straně rovnice vezměme zlom mezi pozicemi konstant A_{n-1}, A_{n-2}. Protějšek tohoto zlomu nemůže obsahovat konstantu A_{n-1}, protože její druhý výskyt je na konci rovnice, tedy jeho protějškem je zlom mezi výskyty proměnné x_1 a konstanty A_{n-2}. Víme tedy, že x_1 končí na A_{n-1}. Uvážíme zlom mezi výskyty proměnné x_2 a konstanty A_{n-3} v levé straně rovnice. Ten musí mít za protějšek zlom mezi výskyty proměnné x_2 a konstanty A_{n-3} vzhledem k tomu, že druhá konstanta tohoto zlomu je A_{n-1}. Opakováním tohoto postupu dospějeme k tomu, že každá proměnná je neprázdná a končí na A_{n-1}. Obdobným symetrickým postupem dále zjistíme, že každá proměnná současně začíná na A_{n-1}.

Nyní, kdykoli vezmeme dvojici po sobě jdoucích pozic $(\boldsymbol{V}, \boldsymbol{W})$ v řešení r, musí být jejich protějškem některá dvojice pozic $(\boldsymbol{V}', \boldsymbol{W}')$, o které již víme, že právě jedna z konstant $\overline{\boldsymbol{V}'}, \overline{\boldsymbol{W}'}$ je A_{n-1}. Tedy i právě jedna z konstant $\overline{\boldsymbol{V}}, \overline{\boldsymbol{W}}$ je A_n. To nám dává, že konstanta A_{n-1} je v řešení „na střídačku“. Přesněji pro každé $i \in \{1, \ldots, n-1\}$ je $|x_i|_r$ liché a navíc pro každé $k \in \{0, \ldots |x_i|_r - 1\}$ je $r(x_i)[k] = A_n$ právě když k je sudé.

Popíšeme bijekci mezi řešeními rovnice $\mathcal{R}$ a řešeními rovnice Pal$(n-1)$, která zachovává l-abecední uspořádání. Tím z indukčního předpokladu přejde řešení $r_{\min}$ na PalSol$(n-1)$ (či přinejmenším na něco se stejnými délkami konstant jako PalSol$(n-1)$, avšak víme, že l-minimální řešení jsou délkami konstant jednoznačně určeny). K vítězství si pak stačí, že i PalSol(n) tato bijekce zobrazí na PalSol$(n-1)$.

Postup je následující. Z řešení r odstraníme všechny pozice konstanty A_{n-1}. Tím dostaneme řešení r' rovnice $\mathcal{R}'$, která vznikne z rovnice $\mathcal{R}$ smazáním obou výskytů konstanty A_{n-1}. Snížením indexu u každé proměnné x_i o jedna dostáváme rovnici Pal$(n-1)$ spolu s nějakým jejím řešením. Toto zobrazení je bijekce, protože má jednoznačně určený zpětný krok – konstantu A_{n-1} do řešení opět „na střídačku“ vrátíme. Dále skutečně zachovává l-abecední uspořádáaní, protože délku každé proměnné upraví rostoucí funkcí $f(n) = (n-1)/2$ a ani následným posunutím indexů nezmění vzájemné pořadí proměnných. Že i PalSol(n) se zobrazí na PalSol(n) je možné snadno ověřit porovnáním délek proměnných v řešení. ■

Tvrzení 7.5. Existuje systém řešitelných 2-rovnic, jejichž minimální řešení je kvadratické v závislosti na délce.

Důkaz: Volíme rovnice Pal(n), s následujícími dvěmi úpravami:

- Smažeme proměnnou x_0.
- Konstantu A_{n-1} nahradíme posloupností konstant $B_1 B_2 \ldots B_n$.

Tyto rovnice jsou řešitelné – stači v řešeních PalSol(r) provést patřičné nahrazení konstant. Dále z tvrzení [nestepici] existuje nejmenší neštěpící řešení, označme jej r. Jak bylo ukázáno v důkazu předchozího tvrzení, musí každá proměnná $x_1, \ldots, x_n$ končit na $B_1 \ldots B_n$. Toto řešení má tedy velikost alespoň n^2, zatímco délka rovnice je $3n-2$. ■

Tvrzení 7.6. Pro každé řešení r' rovnice Pal(n) existuje řešení r, které je v l-uspořádání i c-uspořádání menší než r' a navíc:

- Existuje právě jedna proměnná x_i, která je v řešení r prázdná.
- Všechny ostatní proměnné mají v tomto řešení lichou délku.

Důkaz: Předpokládáme $n \geq 2$, pro $n = 1$ je tvrzení splněno triviálně.

Podle důsledku 3.26 je pod každým řešením r' (v obou uspořádáních) takové řešení r, že pro každou dvojici pozic $(\boldsymbol{V}, \boldsymbol{V}+1)$ v řešení r platí $\overline{\boldsymbol{V}} \neq \overline{\boldsymbol{V}+1}$. Vezmeme tedy takové řešení r a ukážeme o něm požadované vlastnosti.

Porovnáním začátku stran rovnice shledáváme, že x_{n-1} musí začínat na konstantu A_{n-1}. Porovnáním konců pak i že touto konstantou musí končit. Tedy máme tak v řešení dvě uspořádané dvojice pozic odpovídající zlomu mezi A_0 a x_{n-1} na levé straně rovnice $(\boldsymbol{V}, \boldsymbol{V}+1)$ a na pravé straně $(\boldsymbol{W}, \boldsymbol{W}+1)$ takové, že $\left(\overline{\boldsymbol{V}}, \overline{\boldsymbol{V}+1}\right) = (A_0, A_{n-1})$ a $\left(\overline{\boldsymbol{W}}, \overline{\boldsymbol{W}+1}\right) = (A_{n-1}, A_0)$.

Začneme na dvojici po sobě jdoucích pozic v levé straně rovnice $(\boldsymbol{V}, \boldsymbol{V}+\boldsymbol{1}) = (\boldsymbol{V}_0, \boldsymbol{V}_0+1)$ a budeme odvozovat další takové dvojice následujícím algoritmem:

- Najdeme protějšek dvojice $\left(\overline{\boldsymbol{V}_i}, \overline{\boldsymbol{V}_i+1}\right)$, ten bude na pravé straně rovnice. označíme jej $(\boldsymbol{W}_i, \boldsymbol{W}_{i+1})$.
- Je-li $\boldsymbol{W}_i = \boldsymbol{W}$, ukončíme činnost.
- Jinak nastane jedna z následujících tří možností:
  - Je-li pro nějaké j $\left(\overline{\mathrm{Src}(\boldsymbol{W}_i)}, \overline{\mathrm{Src}(\boldsymbol{W}_i+1)}\right) = (A_j, A_{j+1})$, je proměnná x_{n-2-j} v řešení r prázdná. Pak volíme $(\boldsymbol{V}_{i+1}, \boldsymbol{V}_{i+1}+1) = ((\boldsymbol{V}_i+1)\!\downarrow\!, \boldsymbol{V}_i\!\updownarrow)$.
  - Je-li pro nějaké j $\left(\overline{\mathrm{Src}(\boldsymbol{W}_i)}, \overline{\mathrm{Src}(\boldsymbol{W}_i+1)}\right) = (A_j, x_{n-2-j})$, volíme $\boldsymbol{V}_{i+1}+1 = (\boldsymbol{V}_i+1)\!\downarrow$. V takovém okamžiku je $\overline{\boldsymbol{V}_{i+1}} = A_{j+1}$ a proměnná x_{n-2-j} v řešení r není prázdná.
  - Je-li pro nějaké j $\left(\overline{\mathrm{Src}(\boldsymbol{W}_i)}, \overline{\mathrm{Src}(\boldsymbol{W}_i+1)}\right) = (x_{n-1-j}, A_j)$, nemůže být j rovno nule – to bychom podle pravidla výše ukončili algoritmus. Volíme $\boldsymbol{V}_{i+1} = \boldsymbol{V}_i\!\downarrow$. V takovém okamžiku je $\overline{\boldsymbol{V}_{i+1}+1} = A_{j-1}$ a proměnná x_{n-2-j} v řešení r není prázdná.
- Pokračujeme ze začátku s dvojicí $(\boldsymbol{V}_{i+1}, \boldsymbol{V}_{i+1}+1)$

Definujeme posloupnosti σ, ρ, aby pro každé i z průběhu algoritmu bylo

$$\left(\overline{\boldsymbol{V}_i}, \overline{\boldsymbol{V}_i+1}\right) = (A_{\sigma_i}, A_{\rho_i})$$

Všimneme si, že $\sigma_0 - \rho_0 = -(n-1)$, a pokud v kroku i nenastala podmínka (i), je

$$(\sigma_{i+1} - \rho_{i+1}) = (\sigma_i - \rho_i) + 1.$$

Navíc podmínka (i) může nastat pouze za situace i, kdy $\sigma_i - \rho_i = -1$ a v takovém okamžiku je pak $\sigma_{i+1} - \rho_{i+1} = 1$. Algoritmus skončí až když $\sigma_i - \rho_i = (n-1)$ a nikdy se nesmí rovnat $\sigma_i - \rho_i = 0$ – to by byl spor s vlastností r uvedenou v úvodu. Proto algoritmus skončí v kroku $2n-3$, přičemž podmínka (i) nastane právě jednou – v kroku $n-2$.

Označíme j stejné jako v podmínce (i) v kroku $n-1$. Odstraníme z rovnice Pal(n) prázdnou proměnnou $x_{n-2-\sigma_{n-2}}$ se zachováním řešení. Všechny ostatní proměnné jsou díky průběhu algoritmu neprázdné. Nová rovnice má celkem $4(n-1)$ zlomů. Algoritmus prošel $2(n-1)$ kroky, přitom jednotlivé dvojice $(\boldsymbol{V}_i, \boldsymbol{V}_{i+1})$ a $(\boldsymbol{W}_i, \boldsymbol{W}_{i+1})$ přísluší různým zlomům nové rovnice. Algoritmus nimi tak postupně pokryje všechny zlomy v rovnici.

Dále pro $i \le n-2$ je $\sigma_i \le \sigma_{n-2}$ a $\rho_i \ge \sigma_{n-2}+1$, zatímco pro $i \le n-1$ je $\sigma_i \ge \sigma_{n-2}+1$ a $\rho_i \le \sigma_{n-2}$. Z toho plyne, že na všech zlomech je právě jedna konstanta z množiny $M = \{A_0, \ldots, A_{\sigma_{n-2}}\}$. Tedy konstanty z M a mimo ni se musí střídat (podobný argument jako v důkazu tvrzení o l-minimalitě PalSol(n)). Pro $j < x_{n-2-\sigma_{n-2}}$ musí proměnná x_j končit i začínat konstantou z množiny M, naopak pro $j > x_{n-2-\sigma_{n-2}}$ musí proměnná x_j končit i začínat konstantou mimo množinu M – tedy v obou případech bude mít lichou délku. ■

[palindrom-makro]

7.2 Makro-chování Pal(n)

Tvrzení 7.7. Pro libovolné n je řešení PalSol(n) c-minimální. Dokonce, uspořádáme-li řešení r abecedním uspořádáním podle n-tic

$$(\mathrm{Freq}_{r,\mathrm{Pal}(n)}(A_1), \mathrm{Freq}_{r,\mathrm{Pal}(n)}(A_2), \ldots, \mathrm{Freq}_{r,\mathrm{Pal}(n)}(A_n)),$$

bude PalSol(n) nejmenší v tomto uspořádání.

Důkaz: Stačí ukázat druhou část tvrzení, první z ní okamžitě plyne. Dokážeme to indukcí podle n, pro $n = 1$ tvrzení zřejmě platí. Uspořádání na řešeních z tvrzení nazvěme c-abecední a pro $n \ge 2$ uvažme řešení $r_{\min}$, které je nejmenší c-abecedním uspořádání.

Vzhledem k tomu, že $\mathrm{Freq}_{\mathrm{PalSol}(n),\mathrm{Pal}(n)}(A_0) = 1$ musí i $\mathrm{Freq}_{r_{\min},\mathrm{Pal}(n)}(A_0) = 1$. Obě pozice konstanty A_0 tak mají v řešení r_{min} stejný index. Uvažujme tedy soustavu tří rovnic $\mathcal{S}$, která vznikne vyříznutím konstanty A_0 (je jedno, kterého výskytu). Rovnici $A_0 = A_0$ můžeme ze soustavy zahodit bez vlivu na řešení a jejich c-uspořádání.

Zbývá soustava

$$A_{n-1}x_0A_{n-2}x_1\ldots A_1x_{n-2} = x_{n-1}, \quad x_{n-1} = x_{n-2}A_1x_{n-3}A_2\ldots x_0A_{n-1}.$$

Dosazením proměnné x_{n-1} zbude rovnice $\mathcal{R}$, která je až na posunutí indexů u konstant shodná s rovnicí $\mathrm{Pal}(n-1)$. Navíc, máme-li obecné řešení r soustavy S a označíme-li r' odvozené řešení rovnice $\mathcal{R}$, bude pro libovolné $i \in \{1,\ldots,n-1\}$

$$\mathrm{Freq}_{r,\mathrm{Pal}(n)}(A_i) = 2\mathrm{Freq}_{r',\mathrm{Pal}(n)}(A_i).$$

Proto $r_{\min}$ po dosazení odpovídá nejmenšímu řešení $\mathrm{Pal}(n-1)$ v c-abecedním uspořádání, tedy $\mathrm{PalSol}(n-1)$. Zpětným krokem dostaneme $\mathrm{PalSol}(n-1)$, což jsme chtěli dokázat. ∎

Tvrzení 7.8. Pro každé řešení r' rovnice $\mathrm{Pal}(n)$ existuje řešení r, které je v l-uspořádání i c-uspořádání menší než r' a navíc:

- Existuje právě jedna konstanta A_i, pro kterou $\mathrm{Freq}_{r,\mathrm{Pal}(n)}(A_i) = 1$ vyskytuje právě jednou.
- Pro všechny ostatní konstanty A_i je $\mathrm{Freq}_{r,\mathrm{Pal}(n)}(A_i)$ sudé.

Důkaz: Předpokládáme $n \geq 2$, pro $n = 1$ je tvrzení splněno triviálně.

Podle důsledku 4.25 je pod každým řešením r' (v obou uspořádáních) takové řešení r, že pro každou proměnnou x dvojice jejích výskytů $(\boldsymbol{v}, \boldsymbol{w})$ v rovnici splňuje

$$|\mathrm{Ind}_r(\boldsymbol{v}) - \mathrm{Ind}_r(\boldsymbol{w})| > |x|_r$$

nebo se indexy v řešení r obou překryvů rovnají. Druhá varianta ale z podstaty rovnice $\mathrm{Pal}(n)$ nemůže nastat – byly by proti sobě různé konstanty. Můžeme tedy vzorec výše předpokládat pro všechny proměnné.

Označme $\boldsymbol{v}$ největší výskyt proměnné v levé straně rovnice splňující $\mathrm{Ind}_r(\boldsymbol{v}) < \mathrm{Ind}_r(\boldsymbol{v}^{\uparrow\downarrow})$. Poznamenejme, že výskyt proměnné x_0 tuto podmínku splňuje, zatímco výskyt x_{n-1} nikoli. Výskyty $\boldsymbol{v}+2$ a $w = \boldsymbol{v}^{\updownarrow} - 2$ jsou definované a jedná se opět o výskyty jedné proměnné, a opět $\mathrm{Ind}_r(\boldsymbol{w}) < \mathrm{Ind}_r(\boldsymbol{w}^{\uparrow\downarrow})$. Výskyty $\boldsymbol{v}+1$ a $\boldsymbol{w}+1$ jsou navíc výskyty téže konstanty, označme ji C. Využitím nerovnosti výše

$$\mathrm{Ind}_r(\boldsymbol{v}^{\uparrow\downarrow}) > |\boldsymbol{v}|_r + \mathrm{Ind}_r(\boldsymbol{v}) = \mathrm{Ind}_r(\boldsymbol{v}+1),$$

tedy $\mathrm{Ind}_r(\boldsymbol{v}^{\uparrow\downarrow}) - 1 = \mathrm{Ind}_r(w+1) \geq \mathrm{Ind}_r(\boldsymbol{v}+1)$. Obdobně dostaneme i $\mathrm{Ind}_r(v+1) \geq \mathrm{Ind}_r(\boldsymbol{w}+1)$, čili celkem $\mathrm{Ind}_r(\boldsymbol{v}+1) = \mathrm{Ind}_r(\boldsymbol{w}+1)$. Konstanta C se v řešení vyskytuje přesně jednou, okolo C tak můžeme rovnici rozdělit na tři. Rovnici, která říká pouze rovnost konstanty C samy se sebou rovnou smažeme – zbudou dvě rovnice $\mathcal{R}, \mathcal{S}$.

Zbývá si všimnout, že množina všech konstant a proměnných levé strany rovnice $\mathcal{R}$ je totožná s množinou všech konstant a proměnných pravé strany rovnice $\mathcal{S}$. Z toho pro libovolnou konstantu $A \neq C$ je $\mathrm{Freq}_{r,\mathcal{R}}(A) = \mathrm{Freq}_{r_S,\mathcal{S}}(A)$, tedy celkem je $\mathrm{Freq}_{r,\mathrm{Pal}(n)}(A)$ sudé. ∎

7.3 Možné vylepšení – přidání podmínek

Samotná rovnice $\mathrm{Pal}(n)$ jako kandidát na „protipříklad na polynomiální mez pro řešení kvadratických rovnic“ neobstojí. Jakkoli je $\mathrm{PalSol}(n)$ v jistých pohledech optimální, nebrání se rovnice řešení s lineární délkou. Sice v minimálních řešeních rovnice musí být

konstanta, která ji dělí na dvě a proměnná, která má nulovou délku, ale nic rovnici nenutí, aby dělící konstanta byla právě A_0 a prázdná proměnná právě x_0. S touto motivací rovnici Pal(n) drobně vylepšíme.

Definice 7.9. *Rovnicí s podmínkami* rozumíme rovnici na slovech a dodatečný seznam podmínek tvaru „proměnná x musí obsahovat konstantu A“. Přitom proměnné se v seznamu nesmí opakovat a nemusí se použít všechny. Řešením rovnice s podmínkami pak rozumíme takové řešení, ve kterém navíc proměnné splňují zadané podmínky.

Tvrzení 7.10. Pro danou kvadratickou rovnici s p podmínkami $\mathcal{R}$ existuje obyčejná kvadratická rovnice $\mathcal{R}'$, která splňuje

- $\min(\mathcal{R}') = \min(\mathcal{R})$,
- $|\mathcal{R}'| \leq |\mathcal{R}'| + 2p$.

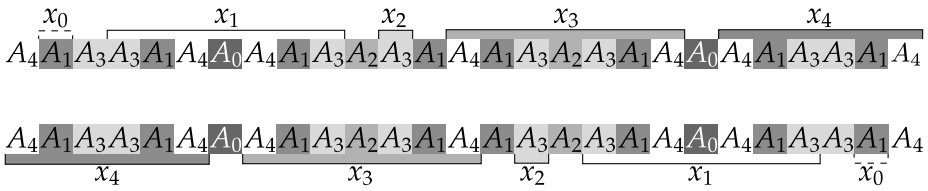
Důkaz: Na základě podmínky „proměnná x musí obsahovat konstantu A“ nahradíme všechny výskyty proměnné x trojicí $x_1 A x_2$, kde x_1, x_2 jsou nové proměnné. ∎

Pozorování 7.11. Řešení PalSol(n) je řešením rovnice Pal(n) s podmínkami „proměnná x_i musí obsahovat konstantu A_{n-i}“ pro $i \in \{1, \ldots, n-1\}$.

Hypotéza 7.12. Velikost nejmenšího řešení rovnice s podmínkami z předchozího pozorování není možné polynomiálně odhadnout.

Poznámka 7.13. Ani s takovými podmínkami není nutně řešení PalSol(n) nejmenší. Například velikost řešení PalSol(5) je 31, zatímco nejmenší velikost řešení rovnice Pal(5) s takovými podmínkami je 27 (strojově ověřeno). Toto řešení je vyobrazeno na obrázku 7.2.

[img-pal5sol]



Obrázek 7.2. Obrázek demonstrující, že ani po přidání podmínek k rovnici Pal(5) nemusí být PalSol(5) nejmenší řešení.

Poznamenejme, že ve zobrazeném řešení není žádná proměnná prázdná, ani žádná konstanta pouze jednou – to proto, že tu máme vedle sebe dvě konstanty A_3 a proměnné x_3 se překrývají. Avšak pokud bychom zrušili dvojitou cestu konstanty A_3, proměnná x_2 by se stala nulovou a nesplňovala by podmínku na obsahování konstanty A_3. Podobně, pokud bychom pokrátili proměnnou x_3, přestala by obsahovat konstantu A_2.

Literatura

[diekert] [1] John Michael Robson, Volker Diekert, *On Quadratic Word Equations*, STACS 1999: 217-226

[kufleitner] [2] Volker Diekert, Manfred Kufleitner, *A Remark about Quadratic Trace Equations*, Developments in Language Theory 2002: 59-66

[grz] [3] Grzegorz Rozenberg, Arto Salomaa (eds.), *Volume 1 Handbook of Formal Languages*, Springer 1997