Bakalářská práce

Kvadratické rovnice na slovech

Miroslav Olšák

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

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

MATHEMATICA PHYSICAQUE
FACULTAS

Poděkování

Díky všem, kteří mi pomohli s touto prací.

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval 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 1. 8. 2013

.....

Přehled

Název práce: Kvadratické rovnice na slovech
Autor: Miroslav Olšák
Katedra algebry
Vedoucí práce: doc. Mgr. Štěpán Holub, Ph.D.
Abstrakt: Práce se zabývá řešitelnosti kvadratických rovnic na slovech. V upravené podobě opakuje výsledky Robsona a Diekerta a navazuje na otázku jednoduše exponenciální meze na velikost nejkratšího řešení kvadratických rovnic. Kladná odpověď na tuto otázku by znamenala, že je řešitelnost kvadratických rovnic na slovech NP úplný problém. Hypotézu o jednoduše exponenciální mezi se dokázat nepodařilo, ale podařilo se například zúžit třídu rovnic, kterým je třeba se věnovat, a dále ukázat, že se stačí zabývat mezí pro nejkratší proměnnou. V závěru práce je ukázáno chování jisté konkrétní rovnice a dále vysvětlena dualita dvou přístupů ke kvadratickým soustavám.
Klíčová slova: kombinatorika na slovech, kvadratické rovnice, jednoduše exponenciální mez

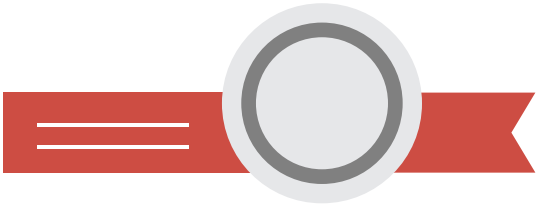
Summary

Title: Quadratic word equations
Author: Miroslav Olšák
Department of Algebra
Supervisor: doc. Mgr. Štěpán Holub, Ph.D.
Abstract: The article discuses satisfiability of quadratic word equations. It reproduces results of Robson and Diekert and explores the question about simple exponential bound of the shortest solution of quadratic word equations. The positive answer to this question would mean NP completeness of satisfiability of quadratic word equations. The simple exponential bound hypothesis was not solved but some results were given: for example a smaller class of equations which needs to be investigated or a proposition saying that it is sufficient to prove a bound of the smallest variable. At the end of this work the behavior of a particular equations is shown and afterwards the duality of two concepts of quadratic word equations handling is explained.
Keywords: combinatorics on words, quadratic equations, simple exponential bound

Obsah

Tabulky

Obrázky



Úvod

Rovnice na slovech je například $AxCABC = ABCCAx$, kde A, B, C jsou konstanty a x je proměnná. Řešení této rovnice je například $x = BCCABC$. Kvadratická soustava rovnic na slovech je soustava rovnic na slovech, ve které se každá proměnná vyskytuje nejvýše dvakrát. Je známa *hypotéza o jednoduše exponenciální mezi na délku řešení*, ve které se tvrdí, že existuje polynom p takový, že nejkratší řešení každé řešitelné soustavy rovnic na slovech má délku omezenou výrazem

$$2^{p(|\mathcal{S}|)},$$

kde $|\mathcal{S}|$ je délka soustavy, tedy počet všech proměnných a konstant v soustavě. V [1] je dokonce pro kvadratické soustavy vyslovena silnější hypotéza o omezení délky řešení výrazem $p(|\mathcal{S}|)$. Ve zdroji [1] je uvedeno, že řešitelnost kvadratických soustav na slovech je NP-těžká a za platnosti hypotézy o jednoduše exponenciální mezi by byla NP-úplná.

Původní záměr této práce byl prověřit platnost hypotézy o jednoduše exponenciální mezi aspoň pro kvadratické soustavy na slovech. To se sice nepodařilo, ale práce přináší několik výsledků, které by mohly inspirovat další výzkum v této oblasti. Například je zde ukázáno, že pro pokoření hypotézy se stačí zabývat menší třídou kvadratických soustav, tzv. 2-soustav.

V kapitole 1 jsou zavedeny základní pojmy. Kapitoly 2 a 3 přinášejí originální terminologii pro metody úprav rovnic a soustav, kterým souhrnně říkáme *mikro-operace* (viz definici 2.53) a *makro-operace* (viz definici 3.28). Dalším výsledkem v těchto kapitolách je například tvrzení 2.66, které říká, že pro pokoření hypotézy o jednoduše exponenciální mezi pro kvadratické rovnice stačí omezit délku nejkratší proměnné.

Kapitola 4 sice opakuje známé výsledky o výpočetní složitosti kvadratických soustav, ale přistupuje k nim jinak. Tvrzení o tom, že kvadratické soustavy jsou NP-těžké, je zde dokázáno pomocí převedení soustavy na graf, zatímco v [1] je toto tvrzení obhájeno pomocí 3-SAT. Dále je v této kapitole prezentován algoritmus, který na rozdíl od algoritmu v [1] není sice lineární, ale je daleko názornější a opírá se o pojmy zavedené v předchozích kapitolách. Pro klíčové důsledky o jednoduše exponenciální mezi je tento názornější algoritmus postačující.

Kapitolu 5 je možno chápat jako odbočku, zabývající se jedním konkrétním problémem palindromické rovnice. Výsledky v této kapitole jsou nové. Stejně tak je v poslední kapitole 6 zcela nově zavedena algebraická struktura 2D rovnic a prokázána její souvislost s balancovanými 2-soustavami. Pomocí toho je pak snadno objasněna dualita mezi některými makro-operacemi a mikro-operacemi pro balancované soustavy.

Tato práce byla formátována za použití TEXového makra CUstyle[1], ovšem pro MFF UK byla nakonec odevzdána podle požadavků fakulty v obvyklém vzhledu. Shodný text této práce včetně nové typografické úpravy podle CUstyle je k dispozici na WWW stránce[2].

[1] `http://petr.olsak.net/custyle.html`
[2] `http://www.olsak.net/mirek/bakalarka/`

Kapitola 1

Základní pojmy a značení

Definice 1.1. *Slovem* s nad danou abecedou $\mathscr{A}$ (konečnou množinou „písmen“) rozumíme konečnou posloupnost písmen z této abecedy. Symbolem $|s|$ značíme délku slova s a symbolem $s[i]$ značíme $(i+1)$-ní písmeno pro $0 \leq i \leq |s|-1$, tedy indexujeme od nuly jako v programovacím jazyce C. Je-li $|s| = 0$, říkáme, že je toto slovo *prázdné*. Množinu všech slov nad danou abecedou $\mathscr{A}$ značíme $\mathscr{A}^*$. Množinu $\mathscr{A}^*$ slov navíc považujeme za monoid s operací skládání slov za sebe (konkatenace). Všechna zobrazení z nějaké abecedy $\mathscr{A}$ do nějaké množiny slov $\mathscr{B}^*$ (nebo jen jiné abecedy $\mathscr{B}$) jednoznačným způsobem rozšiřujeme na homomorfismus $\mathscr{A}^* \to \mathscr{B}^*$.

Definice 1.2. 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 speciální prázdnou množinu, je-li $i_1 > i_2$.

Definice 1.3. Dále *faktorem* slova a rozumíme jakékoli slovo b, pro které existují slova x, y splňující $xby = a$. Dále kdykoli rozdělíme slovo $a = xy$, nazýváme slovo x prefixem slova a a slovo y nazýváme jeho suffixem. Prefix slova a délky k značíme $\mathrm{Pref}_k(a)$ a suffix délky k značíme $\mathrm{Suff}_k(a)$.

Definice 1.4. Mějme danou neprázdnou množinu proměnných $\mathscr{V}$ a s ní disjunktní neprázdnou množinu konstant $\mathscr{C}$. *Rovnice na slovech* je uspořádaná dvojice (S_0, S_1), kde $S_0, S_1 \in (\mathscr{V} \cup \mathscr{C})^*$. Rovnici zapisujeme $S_0 = S_1$. *Soustavou rovnic na slovech* $\mathcal{S}$ (krátce jenom soustavou) rozumíme uspořádanou n-tici takovýchto rovnic $\mathcal{S} = (R_0, R_1, \ldots, R_{n-1})$. Standardně budeme značit jednotlivé rovnice soustavy jako $R_r = (S_{r,0}, S_{r,1})$.

Dále vyžadujeme, aby pro každou rovnici R_r bylo alespoň jedno ze slov $S_{r,0}, S_{r,1}$ neprázdné. Pokud bychom nějakou operací se soustavou vytvořili prázdnou rovnici, okamžitě jej z indexové množiny odstraníme. Podobně vyžadujeme, aby každá proměnná i konstanta byla někde v soustavě použita, pokud bychom ji nějakou operací ze soustavy zcela odstranili, okamžitě ji odstraníme i z příslušné množiny $\mathscr{V}$ nebo $\mathscr{C}$.

Definice 1.5. Pro danou soustavu $\mathcal{S}$ budeme značit množinu proměnných resp. konstant jako $\mathscr{V}_\mathcal{S}$ resp. $\mathscr{C}_\mathcal{S}$. Nakonec počet rovnic soustavy $\mathcal{S}$ značíme $\langle\langle \mathcal{S} \rangle\rangle$.

Definice 1.6. *Délku rovnice* R definujeme jako součet délek obou stran a *délku soustavy* $\mathcal{S}$ jako součet délek všech jejích rovnic. Tyto délky značíme $|R|, |\mathcal{S}|$.

Definice 1.7. *Řešením soustavy* $\mathcal{S}$ nazýváme homomorfismus σ z monoidu $(\mathscr{V}_\mathcal{S} \cup \mathscr{C}_\mathcal{S})^*$ do nějakého nadmonoidu monoidu $\mathscr{C}_\mathcal{S}^*$ splňující:

- Zobrazení σ zachovává konstanty, tedy na množině $\mathscr{C}_\mathcal{S}^*$ je identitou.
- Pro každou rovnici $R_r = (S_{r,0}, S_{r,1})$ soustavy $\mathcal{S}$ platí $\sigma(S_{r,0}) = \sigma(S_{r,1})$.

Pokud homomorfismus splňuje pouze první uvedenou podmínku (a druhou nemáme ověřenou), nazýváme jej *kandidátem na řešení*. Soustavu $\mathcal{S}$ nazýváme *řešitelná*, pokud má nějaké řešení. Dále *délkou proměnné* x *v řešení* σ rozumíme $|\sigma(x)|$.

Poznámka 1.8. Typicky bude obor hodnot řešení přímo $\mathscr{C}_\mathcal{S}^*$. Nadmonoid je v definici proto, abychom nebyli zbytečně omezováni, pokud by například byla množina konstant prázdná.

Definice 1.9. Pro rovnici $R = (S_0, S_1)$ a homomorfismus σ definujeme dosazení $\sigma(R) = (\sigma(S_0), \sigma(S_1))$. Podobně tak pro soustavu $\mathcal{S} = (R_0, \ldots, R_{n-1})$ máme $\sigma(\mathcal{S}) = (\sigma(R_0), \ldots, \sigma(R_{n-1}))$. *Délkou řešení* σ soustavy $\mathcal{S}$ pak rozumíme výraz $|\sigma(\mathcal{S})|$ chápaný jako délka soustavy $\sigma(\mathcal{S})$.

Definice 1.10. *Nejkratší* řešení dané soustavy je takové, že žádné jiné řešení nemá ostře menší délku. Pro řešitelné rovnice v takovém případě definujeme $\min(\mathcal{S})$ jako délku nejkratšího řešení.

Příklad 1.11. Rovnice

$$xABCy = yCBAx$$

délky 10, kde x, y jsou neznámé, A, B, C jsou konstanty, má dvě různá nejkratší řešení délky 14:

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

Navíc se v této rovnici každá proměnná vyskytuje nejvýše dvakrát, což je přesně druh rovnic, kterými se tato práce zabývá.

1.1 Porovnávání řešení

1.1.1 Porovnávání délkami – *l*-uspořádání

Definice 1.12. Na řešeních zavedeme *l-uspořádání* jako následující kvaziuspořádání: Uvažujme řešení σ_1 soustavy $\mathcal{S}_1$ a řešení σ_2 soustavy $\mathcal{S}_2$. Pokud $\mathcal{V}_{\mathcal{S}_1} \neq \mathcal{V}_{\mathcal{S}_2}$ rozšíříme zobrazení σ_1, σ_2 na množinu $\mathcal{V}_{\mathcal{S}_1} \cup \mathcal{V}_{\mathcal{S}_2}$, přitom pro proměnné x, na kterých σ_1 resp. σ_2 nebylo definované, dodefinujeme $\sigma_1(x)$ resp. $\sigma_2(x)$ jako prázdné slovo. Pak píšeme $\sigma_1 \leq_l \sigma_2$, pokud pro každou proměnnou x soustavy platí $|\sigma_1(x)| \leq |\sigma_2(x)|$.

Definujeme *l-minimální řešení* soustavy $\mathcal{S}$ jako minimální prvek v tomto kvaziuspořádání. Tedy σ je l-minimální právě když pro každé řešení σ' soustavy $\mathcal{S}$ splňující $\sigma' \leq_l \sigma$ platí $\sigma \leq_l \sigma'$.

Pozorování 1.13. Pod každým řešením je nějaké l-minimální a každé nejkratší řešení je l-minimální.

1.1.2 Uspořádání podle počtu konstant – *c*-uspořádání

Definice 1.14. Pro slovo s nad abecedou $\mathscr{A}$ a symbol a této abecedy definujeme $|s|_a$ jako počet symbolů a ve slově s. Podobně pro soustavu $\mathcal{S}$ a $a \in \mathcal{V}_{\mathcal{S}} \cup \mathcal{C}_{\mathcal{S}}$ definujeme $|\mathcal{S}|_a$ jako počet výskytů symbolu a v soustavě $\mathcal{S}$ – tedy součet přes všechny strany $S_{r,b}$ všech rovnic.

Pozorování 1.15. Pro slovo s nad abecedou $\mathscr{A}$ a dále pro soustavu $\mathcal{S}$ s řešením σ platí

- $|s| = \sum_{a \in \mathscr{A}} |s|_a$,
- $|\mathcal{S}| = \sum_{x \in \mathcal{V}_{\mathcal{S}}} |\mathcal{S}|_x + \sum_{A \in \mathcal{C}_{\mathcal{S}}} |\mathcal{S}|_A$,
- $|\sigma(\mathcal{S})| = \sum_{A \in \mathcal{C}_{\mathcal{S}}} |\sigma(\mathcal{S})|_A$.

Definice 1.16. Na řešeních zavedeme *c-uspořádání* jako následující kvaziuspořádání. Uvažme řešení σ_1 soustavy $\mathcal{S}_1$ a řešení σ_2 soustavy $\mathcal{S}_2$ Píšeme $\sigma_1 \leq_c \sigma_2$, pokud pro každou konstantu

$$A \in (\mathcal{C}_{\mathcal{S}_1} \cup \mathcal{C}_{\mathcal{S}_2})$$

platí nerovnost

$$|\sigma_1(\mathcal{S}_1)|_A \leq |\sigma_2(\mathcal{S}_2)|_A$$

Pak *c-minimální řešení* soustavy $\mathcal{S}$ je minimální prvek v tomto kvaziuspořádání.

Pozorování 1.17. Pod každým řešením je nějaké c-minimální a každé nejkratší řešení je c-minimální.

Definice 1.18. Prvek $a \in (\mathscr{V}_\mathcal{S} \cup \mathscr{C}_\mathcal{S})$ v soustavě $\mathcal{S}$ nazýváme *unikátní*, pokud se v soustavě vyskytuje právě jednou, tedy $|\mathcal{S}|_a = 1$.

Soustavu $\mathcal{S}$ rovnic nazýváme *kvadratická*, pokud se v ní každá proměnná $x \in \mathscr{V}_\mathcal{S}$ vyskytuje nejvýše dvakrát, tedy $|\mathcal{S}|_x \leq 2$.

Konečně uvedeme hlavní problematiku, kterou se tato práce zabývá.

1.2 Meze pro nejkratší řešení

Tvrzení 1.19. 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 nejkratšího řešení nazýváme *dvojitě exponenciální*.

Důkaz tohoto tvrzení je známý a není složitý. Je uveden v kapitole 3.2, kde je tvrzení zopakováno pod číslem 3.33.

Tvrzení 1.20. Existuje reálná konstanta $c > 0$ taková, že existuje nekonečně mnoho neisomorfních řešitelných kvadratických soustav $\mathcal{S}$, pro které platí $\min(\mathcal{S}) \geq c \cdot |\mathcal{S}|^2$.

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

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

Jako konstantu c v takovém případě můžeme volit $c = 2/9$, totiž $\min(\mathcal{S}) = 2n^2$ a $|\mathcal{S}| = 3n - 1 \leq 3n$, takže

$$\min(\mathcal{S}) \geq \frac{2}{9}|\mathcal{S}|^2.$$

■

Poznámka 1.21. Předchozí tvrzení platí, i pokud se omezíme na soustavy o jedné rovnici. Příklad je dán v tvrzení 5.5 na konci podkapitoly 5.1, případně jiný je snadno odvoditelný z poznámky 2.73 na konci podkapitoly 2.5.

Lepší odhady známy nejsou, tedy můžeme spekulovat.

Hypotéza 1.22. (slabší verze) Existuje polynom p takový, že pro libovolnou řešitelnou kvadratickou soustavu $\mathcal{S}$ má nejkratší řešení délku menší než $2^{p(|\mathcal{S}|)}$. Toto omezení na délku nejkratšího řešení nazýváme *jednoduše exponenciální*.

Tato skutečnost by stačila k tomu, aby bylo řešení kvadratických rovnic NP-úplné (viz kapitola 4). Nicméně absence protipříkladů umožňuje být ještě odvážnější.

Hypotéza 1.23. (silnější verze) Existuje polynom p takový, že pro libovolnou řešitelnou kvadratickou soustavu $\mathcal{S}$ má nejkratší řešení délku menší než $p(|\mathcal{S}|)$. Toto omezení na délku nejkratšího řešení nazýváme *polynomiální*.

1.3 Typ řešení

Definice 1.24. Nechť je dána soustava $\mathcal{S}$ s abecedou proměnných $\mathscr{V}_\mathcal{S} = \{x_0, \ldots, x_{n-1}\}$. Typem řešení (nebo jen kandidáta na řešení) σ pak rozumíme n-tici délek

$$(|\sigma(x_0)|, \ldots, |\sigma(x_{n-1})|).$$

Obecně délkovým typem v soustavě rozumíme takovou n-tici nezáporných celých čísel „předepsaných délek“ bez ohledu na to, zda příslušné řešení existuje.

Dále zavedeme typový homomorfismus t. Pro každou konstantu i proměnnou a soustavy zavedeme novou abecedu $\{a_0, \ldots, a_{k-1}\}$, kde k je předepsaná délka proměnné a nebo $k = 1$, jedná-li se o konstantu. V typovém homomorfismu t je pak slovo $t(a) = a_0 \ldots a_{k-1}$. Konstanty mají přitom délku 1, tedy pro konstantu A je $t(A) = A_0$. Pro řešení σ pak definujeme t_σ jako typový homomorfismus příslušný typu řešení σ, tedy $|t_\sigma(s)| = |\sigma(s)|$ pro každé slovo s. Nakonec na abecedě obrazu zobrazení t_σ definujeme zobrazení $\tilde{\sigma}$ předpisem $\tilde{\sigma}(a_i) = \sigma(a)[i]$.

Definice 1.25. *Délkou typu* či příslušného typového homomorfismu t v soustavě $\mathcal{S}$ rozumíme výraz $|t(\mathcal{S})|$.

Pozorování 1.26. Pro řešení σ platí $\tilde{\sigma} t_\sigma = \sigma$.

Definice 1.27. Říkáme, že typ v soustavě $\mathcal{S} = (R_0, \ldots R_{n-1})$ (případně jeho typový homomorfismus) je *smysluplný*, pokud pro jeho typový homomorfismus t platí

$$|t(S_{r,0})| = |t(S_{r,1})|$$

pro každou rovnici $R_r = (S_{r,0}, S_{r,1})$.

Pozorování 1.28. Typ každého řešení je smysluplný.

Kapitola 2

Mikro-operace s kvadratickou soustavou aneb skákání po pozicích

V této kapitole pracujeme s řešením kvadratických rovnic na bázi malých (i o jedné konstantě) úseků v řešení. Hlavním přínosem je tvrzení 2.58, které omezuje třídu soustav, na které se stačí zaměřit při zkoumání polynomiální resp. jednoduše exponenciální meze, a dále tvrzení 2.66, které ukazuje, že při dokazování jednoduše exponenciální meze se stačí věnovat mezi pro nejkratší proměnnou.

Závěrečná podkapitola 2.5 této kapitoly věnovaná exponentu periodicity ukazuje analogický výsledek, jako je uveden v článku [2], za použití jiného nástroje.

2.1 Pojmy a skoky

2.1.1 Pojmy ohledně výskytů a pozic

Definice 2.1. Uvažujme soustavu $\mathcal{S} = (R_0, \ldots, R_{n-1})$. Pak zavedeme novou abecedu $\mathscr{Q}^{\mathcal{S}}$, která „indexuje výskyty v soustavě". Tedy pro každou rovnici $R_r = (S_{r,0}, S_{r,1})$ soustavy a její stranu $S_{r,b}$ zavedeme písmena

$$\mathscr{Q}^{\mathcal{S}}_{r,b,0}, \ldots, \mathscr{Q}^{\mathcal{S}}_{r,b,|S_{r,b}|-1}.$$

Výskytem v soustavě $\mathcal{S}$ pak myslíme prvek této abecedy. Na množině $\mathscr{Q}^{\mathcal{S}}$ definujeme zobrazení Φ, které výskytům přiřazuje zpátky proměnné a konstanty

$$\Phi\left(\mathscr{Q}^{\mathcal{S}}_{r,b,i}\right) = S_{r,b}[i].$$

Také říkáme, že výskyt $\mathscr{Q}^{\mathcal{S}}_{r,b,i}$ je výskytem proměnné nebo konstanty $S_{r,b}[i]$.

Definice 2.2. Uvažujme soustavu $\mathcal{S} = (R_0, \ldots, R_{n-1})$ a nějaký typový homomorfismus t. Pak zavedeme novou abecedu $\mathscr{P}^{\mathcal{S},t}$, která „indexuje pozice v řešení". Tedy pro každou rovnici $R_r = (S_{r,0}, S_{r,1})$ soustavy a její stranu $S_{r,b}$ zavedeme písmena

$$\mathscr{P}^{\mathcal{S},t}_{r,b,0}, \ldots, \mathscr{P}^{\mathcal{S},t}_{r,b,|t(S_{r,b})|-1}.$$

Pozicí v typu s typovým homomorfismem t (resp. v kandidátu na řešení σ) rozumíme prvek abecedy $\mathscr{P}^{\mathcal{S},t}$ (resp. $\mathscr{P}^{\mathcal{S},t_\sigma}$). Na množině $\mathscr{Q}^{\mathcal{S}}$ definujeme zobrazení Ψ, které pozicím přiřazuje příslušné obrazy typového zobrazení, tedy

$$\Psi_t\left(\mathscr{P}^{\mathcal{S},t}_{r,b,0}\right) = t(S_{r,b})[i].$$

Nakonec pro kandidáta na řešení σ definujeme zkratky $\mathscr{P}^{\mathcal{S},\sigma} = \mathscr{P}^{\mathcal{S},t_\sigma}$ a $\Psi_\sigma = \Psi_{t_\sigma}$.

Definice 2.3. Uvažujme soustavu $\mathcal{S}$ a typový homomorfismus t. Pak definujeme homomorfismus $Z_t : (\mathscr{Q}^{\mathcal{S}})^* \to (\mathscr{P}^{\mathcal{S},t})^*$ tak, aby platilo

- Pro každý výskyt $\boldsymbol{v} \in \mathscr{Q}^{\mathcal{S}}$ jsou slova $t\Phi(\boldsymbol{v})$ a $Z_t(\boldsymbol{v})$ stejně dlouhá.

- Pro každou rovnici $R_r = (S_{r,0}, S_{r,1})$ soustavy $\mathcal{S}$ a její stranu $S_{r,b}$ platí

$$Z_t\left(\mathcal{Q}^{\mathcal{S}}_{r,b,0}\dots\mathcal{Q}^{\mathcal{S}}_{r,b,|S_{r,b}|-1}\right) = \mathcal{P}^{\mathcal{S},t}_{r,b,0}\dots\mathcal{P}^{\mathcal{S},t}_{r,b,|t(S_{r,b})|-1}.$$

Definice 2.4. Pro pozici $\boldsymbol{p} \in \mathcal{P}^{\mathcal{S},t}$ definujeme $Z_t^{-1}(\boldsymbol{p})$ jako takový výskyt $\boldsymbol{v} \in \mathcal{Q}^{\mathcal{S}}$, pro který je znak $\boldsymbol{p}$ obsažen ve slově $Z_t(\boldsymbol{v})$.

Poznámka 2.5. Předchozí definice je jisté násilí na značení – ve skutečnosti se nejedná o vzor v homomorfismu Z_t, ač tomu jak význam, tak značení napovídá. Spíše by se naopak dalo říci, že zobrazení Z_t je vlastně vzorem zobrazení Z_t^{-1} (pokud vnímáme $Z_t(\boldsymbol{v})$ jako množinu a zapomeneme na písmenech strukturu slova). Ve snaze o nematení čtenáře ale necháváme směr zobrazení Z_t stejný jako v případě kandidátů na řešení a typových homomorfismů.

Definice 2.6. Na výskytech $\mathcal{Q}^{\mathcal{S}}_{r,b,i}$ a pozicích $\mathcal{P}^{\mathcal{S},t}_{r,b,i}$ zavedeme přičítání celočíselné konstanty jako její přičítání k poslednímu indexu i. Takový součet definujeme pouze, je-li posunutý výskyt nebo pozice stále v příslušné abecedě.

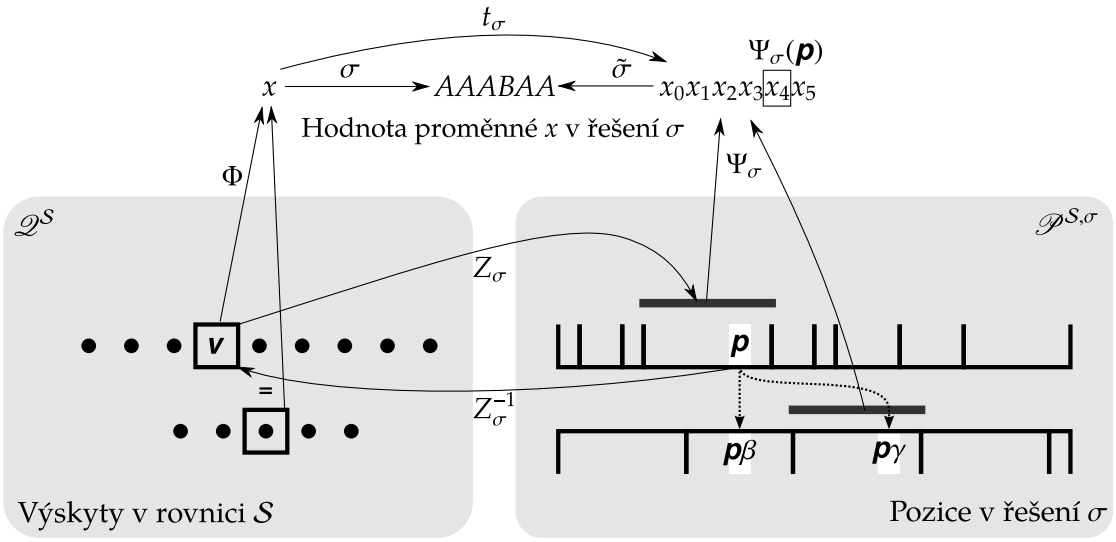
Podobně definujeme porovnávání výskytů a pozic. Píšeme $\boldsymbol{v} < \boldsymbol{w}$, pokud existuje kladné celé k, pro které $\boldsymbol{w} = \boldsymbol{v} + k$.

2.1.2 Jednoduché skoky

V celé sekci o skocích budeme předpokládat pevnou soustavu $\mathcal{S}$, a její smysluplný typový homomorfismus t.

Definice 2.7. Na množině $\mathcal{P}^{\mathcal{S},t}$ definujeme dvě postfixově značená zobrazení β a γ. Pro pozici $\boldsymbol{p} = \mathcal{P}^{\mathcal{S},t}_{r,b,i}$ definujeme $\boldsymbol{p}\beta = \mathcal{P}^{\mathcal{S},t}_{r,1-b,i}$.

Druhé zobrazení definujeme jen pro ty pozice $\boldsymbol{p}$, pro něž je $\Phi Z_t^{-1}(\boldsymbol{p})$ proměnná, která má v soustavě právě dva výskyty. V takovém případě definujeme $\boldsymbol{p}\gamma$ jako pozici různou od $\boldsymbol{p}$ splňující $\Psi_t(\boldsymbol{p}) = \Psi_t(\boldsymbol{p}\gamma)$. Taková pozice je díky předpokladu jednoznačně určená.



Obrázek 2.1. Značení okolo výskytů a pozic

Pozorování 2.8. Zobrazení β je samo k sobě inverzní.

Pozorování 2.9. Zobrazení γ je samo k sobě inverzní.

Pozorování 2.10. Je-li σ řešení a $\boldsymbol{p}$ pozice, platí

$$\tilde{\sigma}\Psi_\sigma(\boldsymbol{p}) = \tilde{\sigma}\Psi_\sigma(\boldsymbol{p}\beta) = \tilde{\sigma}\Psi_\sigma(\boldsymbol{p}\gamma),$$

přičemž druhá rovnost platí jen za předpokladu, že je pravá strana definovaná.

Definice 2.11. Mějme danou soustavu $\mathcal{S}$ a v ní typový homomorfismus t. *Faktorem pozic* (v t) rozumíme faktor nějakého slova

$$\mathscr{P}^{\mathcal{S},t}_{r,b,0} \cdots \mathscr{P}^{\mathcal{S},t}_{r,b,|t(S_{r,b})|-1}$$

pro nějakou stranu $S_{r,b}$ nějaké rovnice R_r.

Definice 2.12. Skoky jakožto homomorfismy **nerozšiřujeme** na všechna slova nad abecedou $\mathscr{P}^{\mathcal{S},t}$, ale jen na **faktory pozic**. To znamená, že pro slovo $\boldsymbol{F} \in (\mathscr{P}^{\mathcal{S},t})^*$ definujeme $\boldsymbol{F}\beta$ jen za předpokladu, že $\boldsymbol{F}$ je faktor pozic. Navíc γ definujeme jen za předpokladu, že $\boldsymbol{F}$ je faktor nějakého slova $Z_t(\boldsymbol{v})$ pro nějaký výskyt $\boldsymbol{v} \in \mathscr{Q}^{\mathcal{S}}$. Z toho pak plyne, že i $\boldsymbol{F}\gamma$ bude faktor pozic.

2.1.3 Pokročilejší skoky – protějšek

Definice 2.13. Buď $\boldsymbol{F}$ neprázdný faktor pozic. Pak uvažujme posloupnost faktorů pozic

$$\boldsymbol{F}, \boldsymbol{F}\gamma\beta, \boldsymbol{F}(\gamma\beta)^2, \ldots,$$

a předpokládejme, že tuto posloupnost není možné definovat nekonečně dlouhou. Pak označme symbolem $\boldsymbol{F}(\gamma\beta)^{\mathrm{MAX}}$ poslední prvek této posloupnosti, který je definovaný. Přitom mějme na paměti, že $\boldsymbol{F}\gamma$ není definovaný pokud nastane libovolná z podmínek

- $\Phi Z_t^{-1}(\boldsymbol{F}[0])$ je proměnná x, pro kterou $|\mathcal{S}|_x = 1$,
- $\Phi Z_t^{-1}(\boldsymbol{F}[0])$ je konstanta,
- Slovo $Z_t^{-1}(\boldsymbol{F})$ obsahuje více různých výskytů.

Protějšek faktoru pozic $\boldsymbol{F}$ definujeme jako $\boldsymbol{F}\beta(\gamma\beta)^{\mathrm{MAX}}$.

2.2 Operace se soustavou a jejím řešením

2.2.1 Mazání prázdných proměnných

Definice 2.14. Uvažujme soustavu $\mathcal{S}$, typový homomorfismus t a proměnnou x, pro kterou je slovo $t(x)$ prázdné. Takovou proměnnou nazýváme *prázdnou v* t. Definujeme soustavu $\mathcal{T}$ s typem t_2, která vznikne *smazáním prázdné proměnné* x jednoduše tak, že všechny její výskyty ze soustavy odstraníme. Typ t_2 pak vznikne z t jednoduše tak, že zapomeneme hodnoty na vyhozených proměnných. Častěji ale budeme používat smazání všech prázdných proměnných v typu t.

Pozorování 2.15. Po smazání všech prázdných proměnných má každá rovnice soustavy neprázdné obě strany.

Pozorování 2.16. Při použití značení z předchozí definice je možné z libovolného řešení τ soustavy $\mathcal{T}$ odvodit řešení σ soustavy $\mathcal{S}$ splňující pro všechny proměnné x soustavy $\mathcal{T}$ rovnost $\sigma(x) = \tau(x)$ a pro ostatní proměnné x soustavy $\mathcal{S}$ je $|\sigma(x)| = 0$. Obráceně to můžeme provést právě pro ta řešení σ, která pro všechny proměnné soustavy x splňují implikaci $|t(x)| = 0 \Rightarrow |\sigma(x)| = 0$.

Speciálně můžeme odvodit řešení τ z řešení σ, pokud jsme odstranili prázdné proměnné typu řešení σ. V takovém případě je $\mathcal{T}$ řešitelná, platí $\min T \geq \min S$ a pokud je navíc σ nejkratší řešení, nastává rovnost.

Definice 2.17. Pojem *odvozené řešení* budeme v souvislosti se smazáním prázdných proměnných používat ve smyslu předchozího pozorování.

2.2.2 Odstranění množiny pozic

Definice 2.18. Uvažujme soustavu $\mathcal{S}$, její řešení σ a $M \subseteq \mathscr{P}^{\mathcal{S},\sigma}$ množinu pozic v tomto řešení, která splňuje

- Pro každou pozici $\boldsymbol{p} \in M$ je $\Phi Z_\sigma^{-1}(\boldsymbol{p})$ proměnná (a nikoli konstanta).
- Existuje množina $N \subseteq t_\sigma(\mathcal{S})$, pro kterou $M = \Psi_\sigma^{-1}(N)$.

Pak definujeme kandidáta na řešení τ vzniklého *odstraněním pozic množiny* M tak, že z řešení σ odstraníme symboly příslušející množině N v typovém homomorfismu t_σ.

Pozorování 2.19. Zobrazení definované v předchozí definici je skutečně kandidát na řešení.

Pozorování 2.20. Druhá podmínka předchozí definice je v kvadratických rovnicích ekvivalentní podmínce: pro každou pozici $\boldsymbol{p} \in M$ je i $\boldsymbol{p}\gamma \in M$.

Definice 2.21. Říkáme, že z řešení σ *lze odstranit množinu pozic* M, pokud tato množina jednak splňuje požadavky na odstranění množiny pozic z předchozí definice a navíc je vzniklý kandidát na řešení opět řešením.

Pozorování 2.22. Splňuje-li množina M pozic podmínky na odstranění z řešení a navíc pro každou pozici $\boldsymbol{p} \in M$ je i $\boldsymbol{p}\beta \in M$, pak kandidát na řešení vzniklý odstraněním této množiny je řešením.

Pozorování 2.23. Řešení vzniklé odstraněním neprázdné množiny výskytů je ostře l-menší i c-menší než původní řešení.

Tvrzení 2.24. Předpokládejme, že v řešení σ kvadratické soustavy $\mathcal{S}$ některá pozice $\boldsymbol{p}$ nemá definované $\boldsymbol{p}(\gamma\beta)^{\mathrm{MAX}}$. Pak je možné z řešení σ odstranit neprázdnou množinu pozic.

Důkaz: Pozic je jen konečně mnoho, proto se v posloupnosti

$$\boldsymbol{p}, \boldsymbol{p}\gamma\beta, \boldsymbol{p}(\gamma\beta)^2, \ldots$$

některá pozice zopakuje. Navíc krok $(\gamma\beta)$ je možné vrátit – krokem $(\beta\gamma)$, proto existuje i, pro které $\boldsymbol{p} = \boldsymbol{p}(\gamma\beta)^i$. Jako množinu pozic, které odstraníme, pak stačí volit

$$\left\{\boldsymbol{p}, \boldsymbol{p}\gamma, \boldsymbol{p}(\gamma\beta)^1, \boldsymbol{p}(\gamma\beta)^1\gamma, \boldsymbol{p}(\gamma\beta)^2, \ldots, \boldsymbol{p}(\gamma\beta)^{i-1}, \boldsymbol{p}(\gamma\beta)^{i-1}\gamma\right\}.$$

Tuto množinu lze odstranit, jelikož splňuje podmínky pozorování 2.22. ∎

Důsledek 2.25. V každém l-minimálním i c-minimálním řešení má každá pozice (a tedy i každý neprázdný faktor pozic) definovaný protějšek.

Tvrzení 2.26. Uvažujme pozici $\boldsymbol{p} \in \mathscr{P}^{\mathcal{S},\sigma}$ v řešení σ. Předpokládejme, že $\Phi Z_\sigma^{-1}(\boldsymbol{p})$ je unikátní proměnná a stejně tak pro protějšek je $\Phi Z_\sigma^{-1}\left(\boldsymbol{p}\beta(\gamma\beta)^{\mathrm{MAX}}\right)$ unikátní proměnná. Pak je možné z řešení σ odstranit neprázdnou množinu pozic.

Důkaz: Odstraníme množinu

$$\left\{\boldsymbol{p}, \boldsymbol{p}\beta, \boldsymbol{p}\beta\gamma, \boldsymbol{p}\beta(\gamma\beta)^1, \boldsymbol{p}\beta(\gamma\beta)^1\gamma, \ldots, \boldsymbol{p}\beta(\gamma\beta)^{\mathrm{MAX}}\right\}.$$

Tato množina splňuje podmínky pozorování 2.22. ∎

Důsledek 2.27. Předpokládejme, že z řešení σ nelze odstranit neprázdnou množinu pozic. Pak pro každou pozici $\boldsymbol{p}$ v řešení σ je alespoň jedna z hodnot $\Phi Z_\sigma^{-1}\left(\boldsymbol{p}(\gamma\beta)^{\mathrm{MAX}}\right)$ nebo $\Phi Z_\sigma^{-1}\left(\boldsymbol{p}\beta(\gamma\beta)^{\mathrm{MAX}}\right)$ konstantou.

Důsledek 2.28. Každé l-minimální i c-minimální řešení σ soustavy $\mathcal{S}$ je jednoznačně určené svým typem. Kdykoli totiž máme pozici $\boldsymbol{p} \in \mathscr{P}^{\mathcal{S},\sigma}$ a $\Phi Z_\sigma^{-1}\left(\boldsymbol{p}(\gamma\beta)^{\mathrm{MAX}}\right) = A \in \mathscr{C}_\mathcal{S}$, je jednoznačně určeno $\tilde{\sigma}\Psi_\sigma(\boldsymbol{v}) = A$.

2.2.3 Vepisování konstant

Definice 2.29. Mějme soustavu $\mathcal{S}$, její řešení σ a její proměnnou x, pro kterou je $\sigma(x)$ neprázdné. Pak definujeme *soustavu $\mathcal{T}$ s řešením τ vzniklou vepsáním konstanty na konec proměnné x na základě řešení σ* tak, že všechny výskyty proměnné x v soustavě $\mathcal{S}$ nahradíme dvojicí xA, kde A je poslední písmeno ve slově $\sigma(x)$. Řešení τ splňuje $\tau(y) = \sigma(y)$ pro všechny proměnné $y \neq x$ a dále $\tau(x)A = \sigma(x)$.

Pozorování 2.30. Vepsáním konstanty na konec proměnné vznikne ostře l-menší řešení stejné délky.

Pozorování 2.31. Kdykoli máme nějaké řešení τ' soustavy $\mathcal{T}$ z předchozí definice, můžeme z něj zpětně odvodit řešení σ' soustavy $\mathcal{S}$ stejné délky tak, že na konec proměnné x vrátíme konstantu A. Z toho plyne $\min(\mathcal{S}) \leq \min(\mathcal{T})$. Pokud navíc bylo σ nejkratší řešení, nastává rovnost.

Definice 2.32. Pojem *odvozené řešení* budeme v souvislosti s vepsáním konstanty na konec proměnné používat ve smyslu předchozího pozorování.

Pozorování 2.33. Mějme soustavu $\mathcal{S}$ a její řešení σ. Na základě řešení σ vepíšeme konstantu na konec nějaké proměnné, tím vznikne soustava $\mathcal{T}$ s řešením τ. Dále uvažujme dvě řešení $\tau_1 \leq_l \tau_2$ soustavy $\mathcal{T}$. Pak z nich odvozená řešení σ_1, σ_2 opět splňují $\sigma_1 \leq_l \sigma_2$. Z toho plyne, že bylo-li σ l-minimální řešení, bude i τ l-minimální řešení.

Definice 2.34. Buď $\mathcal{S}$ soustava, σ její řešení a x její proměnná. Soustava vzniklá *dosazením obrazu* $\sigma(x)$ je soustava, ve které nahradíme každý výskyt proměnné x slovem $\sigma(x)$.

Pozorování 2.35. Dosazení obrazu $\sigma(x)$ je možné považovat za opakované vepisování konstanty na konec proměnné x, dokud není proměnná x prázdná, a nakonec smazání prázdné proměnné x. Ve stejném smyslu budeme tedy hovořit o odvozených řešeních.

2.2.4 Rozrůznění konstant

Definice 2.36. Mějme soustavu $\mathcal{S}$ a typový homomorfismus t nějakého řešení v této soustavě. Pak definujeme *soustavu $\mathcal{T}$ vzniklou rozrůzněním konstant na základě typového homomorfismu t* jako soustavu, ve které

- stále existuje řešení typu t,
- existuje zobrazení $\varphi : \mathscr{C}_\mathcal{T} \to \mathscr{C}_\mathcal{S}$, které když rozšíříme identitou na množinu $\mathscr{V}_\mathcal{S}$, splňuje $\mathcal{S} = \varphi(\mathcal{T})$,
- velikost množiny $\mathscr{C}_\mathcal{T}$ je za daných podmínek je nejvyšší možná.

Pozorování 2.37. Soustava vzniklá rozrůzněním konstant má stejnou délku jako původní soustava.

Pozorování 2.38. Pokud provedeme rozrůznění konstant dvakrát po sobě (se stejným typem) dostaneme až na isomorfismus stejnou soustavu, jako kdybychom jej provedli pouze jednou.

Pozorování 2.39. Kdykoli máme řešení τ soustavy $\mathcal{T}$ z předchozí definice, je homomorfismus $\varphi\tau$ řešením soustavy $\mathcal{S}$ stejného typu jako τ. Z toho plyne $\min(\mathcal{S}) \leq \min(\mathcal{T})$.

Tvrzení 2.40. Po rozrůznění konstant v kvadratické soustavě se ve výsledné soustavě každá konstanta vyskytuje nejvýše dvakrát.

Důkaz: Předpokládejme, že se konstanta A vyskytuje v soustavě $\mathcal{T}$ alespoň třikrát. Ukážeme, že je možné soustavu ještě více rozrůznit – přidat alespoň jednu další konstantu se zachováním podmínky, že existuje řešení typu t. Označme τ řešení soustavy $\mathcal{T}$ typu t. Dále vezměme některý výskyt $\boldsymbol{v} \in \mathscr{D}^\mathcal{T}$ konstanty A a označme pozici $\boldsymbol{p} = Z_\tau(\boldsymbol{v})$. Ve výskytu $\boldsymbol{v}$ nahradíme proměnnou A nově založenou proměnnou B. Dále uvažme protějšek $\boldsymbol{q}$ pozice $\boldsymbol{p}$. Je-li $Z_\tau^{-1}(\boldsymbol{q})$ opět výskytem konstanty A, nahradíme i v něm konstantu

A konstantou B. Nově vzniklou soustavu označme $\mathcal{T}_2$. Jelikož měla konstanta A alespoň 3 výskyty, nějaký výskyt jí zbyl a v soustavě $\mathcal{T}_2$ se tak vyskytuje více konstant než v soustavě $\mathcal{T}$, zbývá ukázat, že soustava $\mathcal{T}$ je stále řešitelná.

To je dáno tím, že můžeme sestrojit její řešení τ_2 tím, že uvážíme množinu pozic

$$M = \{\boldsymbol{p}\beta, \boldsymbol{p}\beta\gamma, \boldsymbol{p}\beta\gamma\beta, \ldots\},$$

kde posledním prvkem je buď $\boldsymbol{q}$, pokud jsme v soustavě nahradili pouze jednu konstantu, nebo $\boldsymbol{q}\beta$, pokud jsme nahradili dvě. K sestrojení τ_2 zbývá v řešení τ nahradit na všech pozicích odpovídajících množině $\Psi_\tau(M)$ konstantu A za konstantu B. ■

2.2.5 Slepené konstanty

Definice 2.41. Buď $\mathcal{S}$ soustava a $s = A_1 \ldots A_n$ slovo složené z n různých konstant. Pak řekneme, že konstanty slova s jsou v soustavě $\mathcal{S}$ *slepené*, pokud každý výskyt každé konstanty ze slova s je součástí celého slova s. Tedy pro každý výskyt $\boldsymbol{v}$ a index i, pro který $\Phi(\boldsymbol{v}) = s[i]$, platí

$$\Phi\Big((\boldsymbol{v} - i)(\boldsymbol{v} - i + 1)\ldots(\boldsymbol{v} + (n - i - 1))\Big) = s.$$

Pozorování 2.42. Relace „konstanty A, B jsou spolu v nějakém slepeném slově“ je ekvivalence. Navíc třídy této ekvivalence je možné zapsat opět jako slepená slova.

Definice 2.43. Ekvivalenci z předchozího pozorování nazveme ekvivalencí *slepenosti*. Říkáme, že je soustava *bez slepených konstant*, je-li každá třída této ekvivalence jednoprvková.

Definice 2.44. Nechť $A_1 \ldots A_n$ jsou konstanty slepené v soustavě $\mathcal{S}$ a σ je její řešení. Řekneme, že konstanty slova $s = A_1 \ldots A_n$ jsou *slepené i v řešení* σ, pokud pro každou proměnnou x platí, že kdekoli se ve slově $\sigma(x)$ vyskytuje některá konstanta slova s, pak se zde vyskytuje jako součást celého slova s.

Definice 2.45. Řekneme, že řešení σ soustavy $\mathcal{S}$ je *neštěpící*, pokud splňuje: kdykoli je nějaké slovo z konstant slepené v soustavě $\mathcal{S}$, je toto slovo slepené i v řešení σ.

Tvrzení 2.46. Pro každé řešení σ soustavy $\mathcal{S}$ existuje neštěpící řešení σ_2, které splňuje $\sigma_2 \leq_c \sigma$.

Důkaz: Rozložíme konstanty na třídy ekvivalence slepenosti v soustavě (příslušné konstanty nemusí být slepené v řešení σ). Pak z každé třídy vybereme jednu takovou konstantu A, pro kterou je $|\sigma(\mathcal{S})|_A$ nejmenší. Následně sestrojíme soustavu $\mathcal{S}_1$ s řešením σ_1 tak, že ze soustavy $\mathcal{S}$ a z řešení σ odstraníme všechny nevybrané konstanty. Následně z těchto řešení vyrobíme řešení σ_2 soustavy $\mathcal{S}_2$ tak, že v soustavě i v řešení nahradíme každou konstantu celým příslušným slovem slepeným v $\mathcal{S}$. Tím ale $\mathcal{S}_2 = \mathcal{S}$ a řešení σ_2 je neštěpící splňuje požadovanou c-nerovnost. ■

2.2.6 Slití slepených konstant

Definice 2.47. Buď $\mathcal{S}$ soustava. Uvažujme v této soustavě třídy ekvivalence slepenosti. Pak definujeme soustavu $\mathcal{T}$ vzniklou *slitím slepených konstant soustavy* $\mathcal{S}$ tak, že pro každou třídu ekvivalence slepenosti, která odpovídá slepeným konstantám $A_1 \ldots A_n$ nahradíme v soustavě každé slovo $A_1 \ldots A_n$ jednou novou proměnnou. Celkem tedy bude mít soustava $\mathcal{T}$ tolik konstant, kolik měla $\mathcal{S}$ tříd ekvivalence slepenosti.

Pozorování 2.48. Slitím slepených konstant vznikne soustava bez slepených konstant.

Pozorování 2.49. Kdykoli máme řešení τ soustavy $\mathcal{T}$ v předchozí definici, můžeme v tomto řešení nahradit každou konstantu odpovídajícím slepeným slovem $A_1 \ldots A_n$ a odvodit tak řešení σ soustavy $\mathcal{S}$. Naopak to můžeme provést právě pro neštěpící řešení.

Definice 2.50. O odvozených řešeních budeme v souvislosti se slitím slepených konstant mluvit ve smyslu předchozího pozorování.

Pozorování 2.51. Neštěpící řešení σ v soustavě $\mathcal{S}$ je c-minimální právě když odvozené řešení τ v soustavě $\mathcal{T}$ vzniklé slitím slepených konstant je c-minimální.

Pozorování 2.52. Pro soustavu $\mathcal{S}$ a z ní vzniklou soustavu $\mathcal{T}$ vzniklou slitím slepených konstant platí

$$\min(\mathcal{T}) \leq \min(\mathcal{S}) \leq \min(\mathcal{T}) \cdot |\mathscr{C}_{\mathcal{S}}|.$$

Definice 2.53. Operace „smazání prázdných proměnných“, „odstranění množiny pozic“, „vepsání konstanty na konec proměnné“, „dosazení obrazu proměnné“, „rozrůznění konstant“, „slití slepených konstant“ budeme souhrnně označovat *mikro-operace*.

Pozorování 2.54. Uvažujme soustavu $\mathcal{S}$ s řešením σ a z té pomocí mikro-operací odvozenou soustavu $\mathcal{T}$ s řešením τ. Platí $\tau \leq_l \sigma$. Dále uvažme dvě řešení τ_1, τ_2 soustavy $\mathcal{T}$ a zpětně odvozená řešení σ_1, σ_2 soustavy $\mathcal{S}$. Pokud platí $\tau_1 \leq_c \tau_2$ tak platí i $\sigma_1 \leq_c \sigma_2$. Pokud tedy řešení σ bylo c-minimální, bude takové i řešení τ.

2.3 2-soustavy

Zkoumat soustavy, ve kterých se každá proměnná vyskytuje nejvýše dvakrát je zbytečně obecné. Ukážeme, že bohatě stačí se zaměřit na „ještě více kvadratické“ soustavy.

Definice 2.55. Definujeme *2-soustavu* jako takovou kvadratickou soustavu na slovech, ve které se každá proměnná i konstanta vyskytuje právě dvakrát.

Pozorování 2.56. Pro 2-soustavu $\mathcal{S}$ platí $|\mathcal{S}| = 2(|\mathscr{V}_{\mathcal{S}}| + |\mathscr{C}_{\mathcal{S}}|)$.

Tvrzení 2.57. Uvažujme kvadratickou soustavu $\mathcal{S}$ a její řešení σ, z kterého nelze odstranit neprázdnou množinu pozic. Pokud do soustavy dosadíme obraz σ všech unikátních proměnných, získáme soustavu, delší (je-li vůbec delší) nejvýše o počet výskytů všech konstant původní soustavy $\mathcal{S}$.

Důkaz: Existuje totiž prosté zobrazení z množiny

$$\left\{\boldsymbol{p} \in \mathscr{P}^{\mathcal{S},\sigma} \;\middle|\; \Phi Z_\sigma^{-1}(\boldsymbol{p}) \text{ je unikátní proměnná}\right\}$$

do množiny

$$\left\{\boldsymbol{p} \in \mathscr{P}^{\mathcal{S},\sigma} \;\middle|\; \Phi Z_\sigma^{-1}(\boldsymbol{p}) \in \mathscr{C}_{\mathcal{S}}\right\}.$$

Tímto zobrazením je samotný protějšek. K tomu, že opravdu každý takový protějšek $\boldsymbol{p}$ splňuje $\Phi Z_\sigma^{-1}(\boldsymbol{p}) \in \mathscr{C}_{\mathcal{S}}$, použijeme tvrzení 2.26. ∎

Tvrzení 2.58. Je-li dána kvadratická soustava $\mathcal{S}$ s řešením σ, ze kterého nelze odstranit neprázdnou množinu pozic, pak existuje 2-soustava $\mathcal{T}$ s řešením τ, pro kterou platí

- $|\mathcal{T}| \leq 2|\mathcal{S}|$,
- $|\sigma(\mathcal{S})| = |\tau(\mathcal{T})|$.
- Bylo-li řešení σ nejkratší resp. l-minimální resp. c-minimální, bude i řešení τ nejkratší resp. l-minimální resp. c-minimální.

Důkaz: Nejprve dosadíme obraz σ všech unikátních proměnných. Díky předchozímu tvrzení tak soustavu zvětšíme nejvýše dvakrát. Druhým (a posledním) krokem k soustavě $\mathcal{T}$ s řešením τ je rozrůznění konstant.

Díky tvrzení 2.40 bude v $\mathcal{T}$ každá konstanta nejvýše dvakrát. Žádná konstanta A se ale v soustavě nemůže vyskytovat právě jednou – pro výskyt $\boldsymbol{v}$ konstanty A musí být i $Z_\sigma^{-1}\left(Z_\tau(\boldsymbol{v})\beta(\gamma\beta)^{\mathrm{MAX}}\right)$ výskytem konstanty A různým od $\boldsymbol{v}$. ∎

Důsledek 2.59. Kdybychom měli polynomiální resp. jednoduše exponenciální mez na 2-soustavy, měli bychom polynomiální resp. jednoduše exponenciální mez na všechny kvadratické soustavy.

Definice 2.60. *Zlomem* v soustavě $\mathcal{S}$ rozumíme dvojici po sobě jdoucích výskytů $z = (\mathcal{Q}^{\mathcal{S}}_{r,b,i}, \mathcal{Q}^{\mathcal{S}}_{r,b,i+1})$. Představa zlomu je „ta mezera mezi písmeny", tedy nazýváme jej zlomem mezi výskyty $\mathcal{Q}^{\mathcal{S}}_{r,b,i}$ a $\mathcal{Q}^{\mathcal{S}}_{r,b,i+1}$.

Pozorování 2.61. Pokud je každá strana každé rovnice v soustavě $\mathcal{S}$ neprázdná, je počet zlomů roven $|\mathcal{S}| - 2\langle\langle\mathcal{S}\rangle\rangle$.

Definice 2.62. Mějme faktor pozic $\boldsymbol{F}$. O zlomu $(\boldsymbol{v}, \boldsymbol{v}+1)$ v soustavě $\mathcal{S}$ řekneme, že leží v tomto faktoru, pokud existují ve slově $\boldsymbol{F}$ jsou pozice $\boldsymbol{p}, \boldsymbol{q}$ splňující

$$Z_t^{-1}(\boldsymbol{p}) \leq \boldsymbol{v} < \boldsymbol{v}+1 \leq Z_t^{-1}(\boldsymbol{q}).$$

Pozorování 2.63. Uvažujme soustavu $\mathcal{S}$ bez unikátních proměnných, v ní smysluplný typový homomorfismus t a faktor pozic $\boldsymbol{F}$ délky alespoň 2. Pak ve $\boldsymbol{F}(\gamma\beta)^{\mathrm{MAX}}$ leží alespoň jeden zlom.

Pozorování 2.64. Máme-li dva faktory pozic $\boldsymbol{F}_1, \boldsymbol{F}_2$, ve kterých leží jeden společný zlom, pak se tyto faktory musí protínat v alespoň dvou pozicích.

Tvrzení 2.65. Uvažme řešitelnou 2-soustavu $\mathcal{S}$ bez slepených konstant. Taková soustava pak splňuje nerovnost

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

Důkaz: Uvažujme nějaké řešení σ soustavy $\mathcal{S}$ a ze soustavy $\mathcal{S}$ vyhodíme rovnice, které mají některou stranu prázdnou. Tím mohou vzniknout unikátní proměnné, ale všechny budou v řešení σ prázdné. Pokud v nové soustavě dokážeme nerovnost z tvrzení, budeme ji mít dokázanou i pro původní soustavu, protože jsme nezměnili levou stranu nerovnosti a pravou jsme snížili.

Množinu všech zlomů soustavy $\mathcal{S}$ označme Y a její podmnožinu složenou ze všech zlomů mezi dvěma výskyty konstant označme $Y_{\mathscr{C}}$.

Definujeme prosté zobrazení $f : Y_{\mathscr{C}} \to Y$. Pro každý zlom $z = (\boldsymbol{v}, \boldsymbol{v}+1) \in Y_{\mathscr{C}}$ uvažme faktor pozic délky 2: $\boldsymbol{F}_z = Z_\sigma(\boldsymbol{v}(\boldsymbol{v}+1))$. Jeho protějšek označme $\boldsymbol{G}_z$. Pak existuje alespoň jeden zlom ležící v $\boldsymbol{G}_z$, některý takový označme $f(z)$.

Zobrazení f je prosté, pokud totiž $f(z_1) = f(z_2)$, tak i $\boldsymbol{G}_{z_1} = \boldsymbol{G}_{z_2}$, tím se rovnají jejich protějšky $\boldsymbol{F}_{z_1} = \boldsymbol{F}_{z_2}$ a tak i samotné $z_1 = z_2$.

Navíc každé $f(z) \in Y \setminus Y_{\mathscr{C}}$. Kdyby totiž $f(z)$ byl zlom mezi dvěma výskyty konstant, byly by tyto konstanty v soustavě slepené.

Označme $c = |Y_{\mathscr{C}}|$ počet zlomů mezi dvěma konstantami a $d = |Y| - c$ počet ostatních zlomů. Pak máme nerovnost $c \leq d$, ekvivalentně $2c \leq |Y|$.

Každý výskyt každé konstanty může být buď poslední v nějaké straně nějaké rovnice, nebo za sebou mít výskyt proměnné, nebo je mezi ním a následujícím výskytem zlom mezi dvěma konstantami. To dává nerovnost

$$2|\mathscr{C}_{\mathcal{S}}| \leq 2\langle\langle\mathcal{S}\rangle\rangle + 2|\mathscr{V}_{\mathcal{S}}| + c.$$

Dosazením nerovnosti $2c \leq |Y| \leq 2(|\mathscr{C}_{\mathcal{S}}| + |\mathscr{V}_{\mathcal{S}}| - \langle\langle\mathcal{S}\rangle\rangle)$ dostáváme

$$2\mathscr{C}_{\mathcal{S}} \leq 2\langle\langle\mathcal{S}\rangle\rangle + 2\mathscr{V}_{\mathcal{S}} + c \leq \langle\langle\mathcal{S}\rangle\rangle + 3\mathscr{V}_{\mathcal{S}} + \mathscr{C}_{\mathcal{S}}$$
$$\mathscr{C}_{\mathcal{S}} \leq \langle\langle\mathcal{S}\rangle\rangle + 3\mathscr{V}_{\mathcal{S}}.$$

2.4 Co by stačilo pro jednoduše exponenciální mez

Tvrzení 2.66. Předpokládejme, že existuje rostoucí polynom p s vlastností: pro každou řešitelnou 2-soustavu $\mathcal{S}$ s alespoň jednou proměnnou existuje řešení σ a nějaká proměnná této soustavy x, že platí $|\sigma(x)| \leq 2^{p(|\mathcal{S}|)}$. Jinými slovy předpokládáme, že máme jednoduše exponenciální omezení na nejmenší možnou délku nejkratší proměnné. Za takového předpokladu platí hypotéza 1.22 o jednoduše exponenciální mez pro délku celého řešení pro kvadratické soustavy.

Navíc, pokud by předpoklad platil pouze pro všechny 2-soustavy o jedné rovnici, platila by i hypotéza 1.22 pro všechny kvadratické soustavy o jedné rovnici.

Důkaz: Díky důsledku 2.59 a pozorování 2.52 stačí ukázat jednoduše exponenciální mez pro 2-soustavy bez slepených konstant. Vezměme si tedy takovou 2-soustavu $\mathcal{S}$, označme $j = |\mathcal{V}_S|$ a $\mathcal{S}_j = \mathcal{S}$. Popíšeme postup, jak z 2-soustavy $\mathcal{S}_{i+1}$ vytvořit 2-soustavu $\mathcal{S}_i$ opět bez slepených konstant a s i proměnnými. Tento postup budeme provádět až do $\mathcal{S}_0$.

Z předpokladu máme řešení σ_{i+1} soustavy $\mathcal{S}_{i+1}$ a proměnnou x_{i+1} takovou, že platí $|\sigma_{i+1}(x_{i+1})| \leq 2^{p(|\mathcal{S}_{i+1}|)}$. Pak vytvoříme soustavu $\mathcal{T}_i$ tak, že do $\mathcal{S}_{i+1}$ dosadíme obraz $\sigma_{i+1}(x_{i+1})$ a následně na základě odvozeného σ_{i+1} rozrůzníme konstanty. Poté v soustavě $\mathcal{T}_i$ slijeme slepené konstanty, čímž dostaneme soustavu $\mathcal{S}_i$.

Při tomto odvozování platí

- $\langle\langle \mathcal{S}_i \rangle\rangle = \langle\langle \mathcal{T}_i \rangle\rangle \leq \langle\langle \mathcal{S}_{i+1} \rangle\rangle$, rovnice může při kroku zmizet, pokud dosadíme prázdný obraz proměnné, která je v nějaké rovnici samotná,
- $|\mathscr{C}_{\mathcal{S}_i}| \leq 3j + \langle\langle \mathcal{S}_i \rangle\rangle$ podle tvrzení 2.65 a díky skutečnosti, že $\mathcal{S}_i$ nemá slepené proměnné,
- $|\mathcal{S}_i| = 2(|\mathcal{V}_{\mathcal{S}_i}| + |\mathscr{C}_{\mathcal{S}_i}|) \leq 2j + 2(3j + \langle\langle \mathcal{S} \rangle\rangle) = 8j + 2\langle\langle \mathcal{S} \rangle\rangle \leq 10|\mathcal{S}|$, první rovnost je pozorování 2.56, následující nerovnost pak vyplývá z předchozího bodu, poslední nerovnost vyplývá z triviálních odhadů hodnot j a $\langle\langle \mathcal{S} \rangle\rangle$,
- $|\mathscr{C}_{\mathcal{T}_i}| = |\mathscr{C}_{\mathcal{S}_{i+1}}| + |\sigma_{i+1}(x_{i+1})| \leq 3j + \langle\langle \mathcal{S}_{i+1} \rangle\rangle + 2^{p(|\mathcal{S}_{i+1}|)} \leq 4|\mathcal{S}| + 2^{p(10|\mathcal{S}|)}$, počet konstant $\mathcal{T}_i$ vyplývá z definice této 2-soustavy, další odhady plynou z předpokladu o řešení σ_{i+1} a předchozích dvou bodů,
- $|\mathscr{C}_{\mathcal{T}_i}| \leq 4|\mathcal{S}| + 2^{p(10|\mathcal{S}|)} \leq 2^{p_2(|\mathcal{S}|)}$ pro nějaký polynom p_2 nezávislý na $\mathcal{S}$, tedy odvodíme z polynomu p polynom p_2 splňující pro každé $x \in \mathbb{R}$ nerovnost $4x + 2^{p(10x)} \leq 2^{p_2(x)}$,
- $\min(\mathcal{S}_{i+1}) \leq \min(\mathcal{T}_i) \leq \min(\mathcal{S}_i) \cdot |\mathscr{C}_{\mathcal{T}_i}| \leq \min(\mathcal{S}_i) \cdot 2^{p_2(|\mathcal{S}|)}$, první nerovnost plyne z pozorování 2.31 a 2.39, druhá z pozorování 2.52 a poslední z předchozího bodu,
- $\min(\mathcal{S}_0) = |\mathcal{S}_0| = 2\langle\langle \mathcal{S}_0 \rangle\rangle$, protože v této 2-soustavě nejsou žádné proměnné ani slepené konstanty.

Použijeme poslední dva body, tedy rovnost $\min(\mathcal{S}_0) = 2\langle\langle \mathcal{S}_0 \rangle\rangle$ a nerovnost $\min(\mathcal{S}_{i+1}) \leq \min(\mathcal{S}_i) \cdot 2^{p_2(|\mathcal{S}|)}$. Celkem tak pro všechna $i \in \{0, \ldots, j\}$ indukcí dostáváme

$$\min(\mathcal{S}_i) \leq 2\langle\langle \mathcal{S}_0 \rangle\rangle \cdot \left(2^{p_2(|\mathcal{S}|)}\right)^i = 2\langle\langle \mathcal{S}_0 \rangle\rangle \cdot 2^{i \cdot p_2(|\mathcal{S}|)},$$

a tak

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

což dává kýžený odhad. Dodatek k tvrzení plyne ze skutečnosti, že při manipulaci se soustavou $\mathcal{S}$ nezvyšujeme počet rovnic. ∎

2.5 Periodicita

Definice 2.67. O výskytu $\mathscr{Q}^{\mathcal{S}}_{r,b,i}$ říkáme, že je *levý*, pokud $b = 0$. Naopak, je-li $b = 1$, mluvíme o *pravém* výskytu. O soustavě říkáme, že je *balancovaná*, pokud má každá proměnná i konstanta stejný počet pravých výskytů jako levých.

Pozorování 2.68. Je-li v kvadratické rovnici $\mathcal{S}$ výskyt $\boldsymbol{v} \in \mathcal{Q}^{\mathcal{S}}$ levý, pak jeho protějšek je pravý výskyt a obráceně.

2.5.1 Opakující se konstanta

Tvrzení 2.69. Uvažujme kvadratickou soustavu $\mathcal{S}$ bez unikátních proměnných, její řešení σ a její konstantu A. Číslem n označme počet výskytů konstanty A a předpokládejme, že existuje faktor $\boldsymbol{F}$ pozic délky $n+1$, pro který $\tilde{\sigma}\Psi_\sigma(\boldsymbol{F}) = A^{n+1}$. Pak z tohoto řešení lze odstranit neprázdnou množinu pozic. Je-li navíc $\mathcal{S}$ balancovaná soustava, stačí n volit jako počet výskytů konstanty A na jen levých nebo jen pravých stranách.

Důkaz: BÚNO předpokládejme, že faktor $\boldsymbol{F}$ je na levé straně rovnice. Pro každou pozici $\boldsymbol{p} \in \boldsymbol{F}$ najdeme výskyt $Z_\sigma^{-1}\left(\boldsymbol{p}(\gamma\beta)^{\mathrm{MAX}}\right)$. Každý takový výskyt $\boldsymbol{v}$ bude výskytem konstanty A. Těch je pouze n, najdeme tedy dvě různé pozice $\boldsymbol{p}_1, \boldsymbol{p}_2 \in \boldsymbol{F}$, pro které

$$Z_\sigma^{-1}\left(\boldsymbol{p}_1(\gamma\beta)^{\mathrm{MAX}}\right) = Z_\sigma^{-1}\left(\boldsymbol{p}_2(\gamma\beta)^{\mathrm{MAX}}\right).$$

Pokud jsme navíc měli soustavu balancovanou, je každá pozice $\boldsymbol{p}(\gamma\beta)^{\mathrm{MAX}}$ na stejné straně jako $\boldsymbol{p}$. Proto k nalezení takové dvojice $\boldsymbol{p}_1, \boldsymbol{p}_1$ stačilo mít délku faktoru $\boldsymbol{F}$ poloviční.

Jelikož je $\left|Z_\sigma\left(Z_\sigma^{-1}\left(\boldsymbol{p}_1(\gamma\beta)^{\mathrm{MAX}}\right)\right)\right| = 1$, máme dokonce

$$\boldsymbol{p}_1(\gamma\beta)^{\mathrm{MAX}} = \boldsymbol{p}_2(\gamma\beta)^{\mathrm{MAX}}.$$

Označme i, j taková, že

$$\boldsymbol{p}_1(\gamma\beta)^{i} = \boldsymbol{p}_1(\gamma\beta)^{\mathrm{MAX}} = \boldsymbol{p}_2(\gamma\beta)^{\mathrm{MAX}} = \boldsymbol{p}_2(\gamma\beta)^{j}.$$

BÚNO $i > j$, čili

$$\boldsymbol{p}_1(\gamma\beta)^{i-j} = \boldsymbol{p}_2$$

Pak sestrojíme kandidáta na řešení σ' odstraněním množiny M definované jako

$$M = \{\boldsymbol{p}_1, \boldsymbol{p}_1\gamma, \boldsymbol{p}_1(\gamma\beta)^1, \boldsymbol{p}_1(\gamma\beta)^1\gamma, \ldots \boldsymbol{p}_1(\gamma\beta)^{i-j}\beta = \boldsymbol{p}_2\beta\}.$$

Zbývá ověřit, že σ' je řešením.

Pro každou $\boldsymbol{p} \in M \setminus \{\boldsymbol{p}_1, \boldsymbol{p}_2\}$ platí $\boldsymbol{p}\beta \in M$. Proto stačí ověřit rovnost levé a pravé strany jen pro rovnici faktoru $\boldsymbol{F}$. A ta také platí, protože ve slově A^{n+1} nezáleží, na kterých místech konstantu A odstraníme. ∎

2.5.2 Lineární mez na exponent periodicity

V článku [2] je dokázaná lineární mez na exponent periodicity. Ukážeme zde tentýž výsledek pomocí protějšku.

Definice 2.70. Buď $\mathcal{S}$ soustava na slovech a σ její řešení. Pak *exponentem periodicity řešení* σ rozumíme největší přirozené číslo n takové, že existuje rovnice $R_r = (S_{r,0}, S_{r,1})$ a slova $a, b, p \in \mathcal{C}_{\mathcal{S}}^*$ splňující

$$\sigma(S_{r,0}) = ap^n b,$$

přičemž p je neprázdné.

Exponent periodicity řešitelné soustavy $\mathcal{S}$ pak je největší možný exponent periodicity přes všechna možná nejkratší řešení.

Tvrzení 2.71. Uvažujme kvadratickou soustavu $\mathcal{S}$ bez unikátních proměnných a její řešení σ. Dále označme z jako počet zlomů soustavy. Předpokládejme, že existuje faktor pozic $\boldsymbol{F}$ splňující $\tilde{\sigma}\Psi_\sigma(\boldsymbol{F}) = a^{z+3}$ pro nějaké slovo a délky alespoň 2, a navíc, že pro slovo a délky 1 již takový faktor neexistuje. Pak z řešení σ lze odstranit neprázdnou množinu pozic.

Důkaz: Volme nejkratší možné slovo a a označme $k = |a|$, $\boldsymbol{G} = \mathrm{Pref}_k(\boldsymbol{F})$. Pak budou slova

$$\tilde{\sigma}\Psi_\sigma(\boldsymbol{G}), \tilde{\sigma}\Psi_\sigma(\boldsymbol{G}+1), \ldots, \tilde{\sigma}\Psi_\sigma(\boldsymbol{G}+(k-1)),$$

navzájem různá.

Uvažme nyní $z+1$ faktorů pozic:

$$(\boldsymbol{G}+k)(\gamma\beta)^{\mathrm{MAX}}, (\boldsymbol{G}+2k)(\gamma\beta)^{\mathrm{MAX}}, \ldots, (\boldsymbol{G}+(z+1)k)(\gamma\beta)^{\mathrm{MAX}}$$

Každý z těchto faktorů obsahuje nějaký zlom, existují tedy čísla $m, n \in \{1, \ldots, z+1\}$, pro která se protínají faktory

$$(\boldsymbol{G}+mk)(\gamma\beta)^{\mathrm{MAX}}, \quad (\boldsymbol{G}+nk)(\gamma\beta)^{\mathrm{MAX}}.$$

Dále najdeme exponenty i, j splňující

$$(\boldsymbol{G}+mk)(\gamma\beta)^{i} = (\boldsymbol{G}+mk)(\gamma\beta)^{\mathrm{MAX}}, \quad (\boldsymbol{G}+nk)(\gamma\beta)^{j} = (\boldsymbol{G}+nk)(\gamma\beta)^{\mathrm{MAX}}.$$

BÚNO $i > j$. Pak se protínají i faktory pozic $(\boldsymbol{G}+mk)(\gamma\beta)^{i-j}$ a $(\boldsymbol{G}+nk)$, což jsou dokonce faktory slova $\boldsymbol{F}$. Vzhledem k vlastnostem slova $\tilde{\sigma}\Psi_\sigma(\boldsymbol{F})$ a tomu, že a je nejkratší možné, tak máme $(\boldsymbol{G}+mk)(\gamma\beta)^{i-j} = (\boldsymbol{G}+nk)$.

Z řešení σ tak lze (obdobně jako v případě důkazu předchozího tvrzení) odstranit množinu

$$(\boldsymbol{G}+mk) \cup (\boldsymbol{G}+mk)\gamma \cup (\boldsymbol{G}+mk)(\gamma\beta)^1 \cup (\boldsymbol{G}+mk)(\gamma\beta)^1\gamma \cup \ldots \cup (\boldsymbol{G}+mk)(\gamma\beta)^{i-j}\beta = (\boldsymbol{G}+nk)\beta.$$

■

Důsledek 2.72. Kombinace předchozího tvrzení a tvrzení 2.69 dává lineární (vzhledem k délce) mez exponentu periodicity pro kvadratické soustavy bez unikátních proměnných. Pomocí tvrzení 2.57 můžeme tuto lineární mez na exponent periodicity rozšířit dokonce na všechny kvadratické soustavy,

Poznámka 2.73. Lepší než lineární mez pro exponent periodicity nemůže být, 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 rozpad rovnice na dvě neřešitelné 2-rovnice. Jediné nejkratší řešení σ tedy je, když pro všechna i je $\sigma(a_i) = A$ a $\sigma(b_i) = B$. Její exponent periodicity tak je $n+1$.

Kapitola 3

Makro-operace s kvadratickou soustavou aneb lámání a dosazování

V této kapitole především budujeme terminologii pro přehlednou formulaci algoritmu v podkapitole 4.2, který je ovšem značně inspirován algoritmem v [1]. Dále je zde uveden standardní důkaz dvojitě exponenciální meze.

Novým přínosem je pak hlavně tvrzení 3.35, které navazuje na předchozí tvrzení 2.66 a ještě více omezuje nutný okruh zkoumání pro dokazování jednoduše exponenciální meze.

3.1 Operace

3.1.1 Lámání rovnic

Definice 3.1. Uvažujme soustavu $\mathcal{S}$, smysluplný typový homomorfismus t v této soustavě a dva zlomy $z_1 = \left(\mathcal{Q}^{\mathcal{S},t}_{r,0,i}, \mathcal{Q}^{\mathcal{S},t}_{r,0,i+1}\right)$, $z_2 = \left(\mathcal{Q}^{\mathcal{S},t}_{r,1,j}, \mathcal{Q}^{\mathcal{S},t}_{r,1,j+1}\right)$ splňující

$$\left|t(\mathrm{Pref}_{i+1}(S_{r,0})\right| = \left|t(\mathrm{Pref}_{j+1}(S_{r,1})\right|.$$

Pak definujeme soustavu $\mathcal{T}$ vzniklou *rozlomením rovnice R_r v soustavě $\mathcal{S}$ ve zlomech z_1, z_2*. Soustava $\mathcal{T}$ obsahuje stejné rovnice jako soustava $\mathcal{S}$ až na to, že neobsahuje rovnici R_r. Místo ní obsahuje dvě nové rovnice

$$\mathrm{Pref}_{i+1}(S_{r,0}) = \mathrm{Pref}_{j+1}(S_{r,1}), \quad \mathrm{Suff}_{|S_{r,0}|-i-1}(S_{r,0}) = \mathrm{Suff}_{|S_{r,0}|-j-1}(S_{r,1}).$$

Pozorování 3.2. Při použití značení z předchozí definice je t typovým homomorfismem nějakého řešení soustavy $\mathcal{S}$ právě tehdy, když je typovým homomorfismem toho samého řešení v soustavě $\mathcal{T}$. Dále libovolné řešení soustavy $\mathcal{T}$ je opět řešením soustavy $\mathcal{S}$. Obráceně to platí jen pro ta řešení σ, která splňují podmínku z definice pro zlomy z_1, z_2.

Pozorování 3.3. Pokud v soustavě rozlomíme rovnici, snížíme počet celkových zlomů o 2 a zachováme množinu konstant.

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

3.1.2 Lámání proměnných

Definice 3.4. Uvažujme soustavu $\mathcal{S}$, smysluplný typový homomorfismus t v této soustavě, výskyt $\mathcal{Q}^{\mathcal{S}}_{r,b,i}$ proměnné x a nezáporné celé číslo j splňující

$$\left|t(\mathrm{Pref}_i(S_{r,b}))\right| \leq j \leq \left|t(\mathrm{Pref}_{i+1}(S_{r,b}))\right|.$$

Pak definujeme soustavu $\mathcal{T}$ s typovým homomorfismem t_2 vzniklou *rozlomením výskytu $\mathcal{Q}^{\mathcal{S}}_{r,b,i}$ v bodě j* na základě typu t. Soustava $\mathcal{T}$ vznikne nahrazením všech výskytů proměnné x za dvojici nových proměnných x_1x_2. V novém typovém homomorfismu t_2 pak

$$|t_2(x_1)| = j - \left|t(\mathrm{Pref}_i(S_{r,b}))\right|, \quad |t_2(x_2)| = \left|t(\mathrm{Pref}_{i+1}(S_{r,b}))\right| - j.$$

Na všech ostatních proměnných se typový homomorfismus t_2 shoduje s t. Dvojici výskytů v $\mathcal{T}$ vzniklou nahrazením výskytu $\boldsymbol{v}$ říkáme *zlom vzešlý z rozlomení proměnné*.

Pozorování 3.5. Při použití značení z předchozí definice lze z každého řešení τ soustavy $\mathcal{T}$ odvodit řešení σ soustavy $\mathcal{S}$ splňující $\tau(y) = \sigma(y)$ pro každou proměnnou y různou od x a dále $\tau(x_1x_2) = \sigma(x)$. Totéž lze provést i obráceně. Tedy z libovolného σ odvodit nějaké takové řešení τ (máme dokonce na výběr). Platí tak $\min S = \min T$.

Definice 3.6. Pojem *odvozené řešení* budeme v souvislosti s rozlomením proměnné používat ve smyslu předchozího pozorování.

Pozorování 3.7. Uvažujme situaci z definice 3.4. Kdykoli máme řešení τ typu t_2 soustavy $\mathcal{T}$, je odvozené řešení σ soustavy $\mathcal{S}$ typu t. Naopak pro každé řešení σ typu t soustavy $\mathcal{S}$ existuje právě jedno řešení τ typu t_2 soustavy $\mathcal{T}$, z nějž odvozené řešení je právě σ.

Pozorování 3.8. Rozlomením proměnné v kvadratické soustavě zvýšíme celkový počet zlomů nejvýše o 2 a zachováme množinu konstant.

Definice 3.9. Uvažujme soustavu $\mathcal{S}$, smysluplný typový homomorfismus t, výskyt $\mathcal{Q}^{\mathcal{S}}_{r,b,i}$ v soustavě $\mathcal{S}$ její konstanty nebo proměnné neprázdné v t. Předpokládáme, že existuje-li $\mathcal{Q}^{\mathcal{S}}_{r,b,i+1}$, tak $\left|Z_t\left(\mathcal{Q}^{\mathcal{S}}_{r,b,i+1}\right)\right| > 0$. Definujeme soustavu $\mathcal{T}$ s typovým homomorfismem t_2 vzniklou *rozlomením soustavy $\mathcal{S}$ za výskytem $\mathcal{Q}^{\mathcal{S}}_{r,b,i}$ na základě typu t*. Je-li $\mathcal{Q}^{\mathcal{S}}_{r,b,i}$ poslední (tedy $\mathcal{Q}^{\mathcal{S}}_{r,b,i+1}$ není definováno), ponecháváme $\mathcal{T} = \mathcal{S}$ a $t_2 = t$. V opačném případě se podíváme na pozici $\mathcal{P}^{\mathcal{S},t}_{r,b,k}$ definovanou jako

$$\mathcal{P}^{\mathcal{S},t}_{r,b,k} = \mathrm{Suff}_1\left(Z_t\left(\mathcal{Q}^{\mathcal{S}}_{r,b,i}\right)\right)$$

a označíme výskyt $\mathcal{Q}^{\mathcal{S}}_{r,1-b,j} = Z_t^{-1}\left(\mathcal{P}^{\mathcal{S},t}_{r,1-b,k}\right)$.

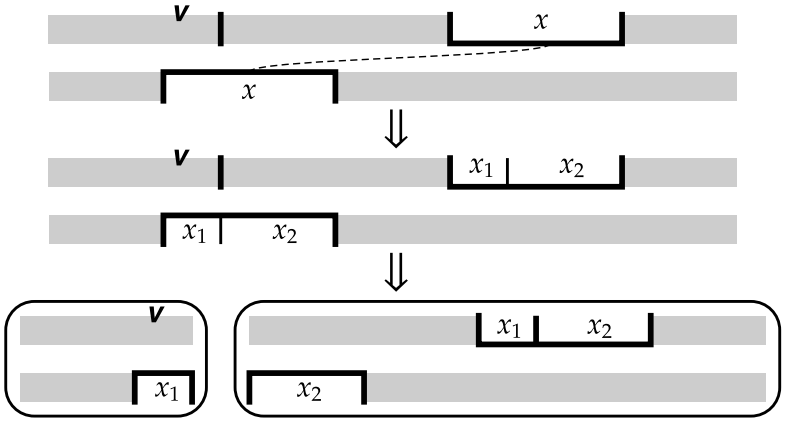
Pokud

$$Z_t^{-1}\left(\mathcal{P}^{\mathcal{S},t}_{r,1-b,k+1}\right) \neq \mathcal{Q}^{\mathcal{S}}_{r,1-b,j},$$

rozlomíme rovnici ve dvojici zlomů $(\mathcal{Q}^{\mathcal{S}}_{r,b,i}, \mathcal{Q}^{\mathcal{S}}_{r,b,i+1})$ a $(\mathcal{Q}^{\mathcal{S}}_{r,1-b,j}, \mathcal{Q}^{\mathcal{S}}_{r,1-b,j+1})$.

V opačném případě nejprve rozlomíme výskyt $\mathcal{Q}^{\mathcal{S}}_{r,1-b,j}$ v bodě $k+1$ a pak teprve rozlomíme rovnici ve zlomu $(\mathcal{Q}^{\mathcal{S}}_{r,b,i}, \mathcal{Q}^{\mathcal{S}}_{r,b,i+1})$ a zlomu vzešlém z rozlomení proměnné. Výsledek prohlásíme za soustavu $\mathcal{T}$ s typovým homomorfismem t_2.

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



Obrázek 3.1. Rozlomení soustavy za výskytem $\boldsymbol{v}$.

Pozorování 3.10. Při použití značení z předchozí definice lze z každého řešení τ soustavy $\mathcal{T}$ odvodit řešení σ soustavy $\mathcal{S}$ (ať už lámeme před výskytem, za výskytem, či dokonce vylamujeme výskyt). Opět tedy budeme hovořit o odvozených řešeních.

Pozorování 3.11. Je-li σ řešení typu t soustavy $\mathcal{S}$ ve smyslu předchozí definice, existuje právě jedno řešení τ typu t_2 takové, že řešení zpětně odvozené z τ je σ.

Pozorování 3.12. 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.

3.1.3 Dosazování

Definice 3.13. O rovnici $R_r = (S_{r,0}, S_{r,1})$ v soustavě $\mathcal{S}$ řekneme, že je *dosaditelná*, pokud alespoň jedna ze stran $S_{r,0}, S_{r,1}$ má délku 1 a navíc je tato strana tvořena proměnnou, která se na druhé straně této rovnice nevyskytuje.

Definice 3.14. Uvažujme v soustavě $\mathcal{S}$ dosaditelnou rovnici R_r a její stranu $S_{r,b} = x$, kde x je proměnná. Definujeme *soustavu $\mathcal{T}$ vzniklou dosazením strany $S_{r,b}$* (případně výskytu $\mathcal{Q}^{\mathcal{S}}_{r,b,0}$) jako soustavu vzniklou z $\mathcal{S}$ tak, že odstraníme rovnici R_r a všechny zbylé výskyty proměnné x v rovnici nahradíme slovem $S_{r,1-b}$.

Poznámka 3.15. Může se stát, že proměnná x nemá v soustavě jiný výskyt, tedy dosazení rovnice r je pouze jejím smazáním. Pokud i na druhé straně rovnice byla nějaká proměnná nebo konstanta, která se v soustavě vyskytovala pouze jednou, zmizí i ta ze soustavy (tedy nejen proměnná x).

Pozorování 3.16. Dosazení v kvadratických soustavách nezvýší počet zlomů, zato sníží počet rovnic o jedna.

Pozorování 3.17. 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.

Pozorování 3.18. Uvažujme soustavu $\mathcal{S}$ se smysluplným typovým homomorfismem t. Provedeme několik dosazení, čímž vznikne soustava $\mathcal{T}$ s odvozeným typem t_2. Přitom, dosazujeme pouze takové proměnné, které mají i další výskyt. Při jednom dosazení se vždy odstraní jedna rovnice, která navíc není v typovém homomorfismu t nejdelší. Pokud tedy počet dosazení označíme d, platí nerovnost

$$|t(\mathcal{S})| \leq |t_2(\mathcal{T})| \cdot (d+1),$$

speciálně

$$|t(\mathcal{S})| \leq |t_2(\mathcal{T})| \cdot \langle\langle \mathcal{S} \rangle\rangle .$$

Definice 3.19. O rovnici $R_r = (S_{r,0}, S_{r,1})$ v soustavě $\mathcal{S}$ řekneme, že je *triviální*, pokud obě strany $S_{r,0}, S_{r,1}$ mají délku 1.

Definice 3.20. Buď $\mathcal{S}$ řešitelná soustava a r její triviální rovnice. *Eliminací triviální rovnice* r myslíme buď její dosazení, je-li rovnice dosaditelná (jsou-li obě strany tvořeny proměnnou, nezáleží na výběru proměnné, kterou dosadíme, výsledek bude až na isomorfismus soustav stejný), případně odstranění této rovnice ze soustavy, je-li rovnice tvaru $a = a$ pro nějakou konstantu nebo proměnnou a.

Poznámka 3.21. Jediná možná triviální rovnice, kterou není možné eliminovat, je tvaru $A = B$, kde A, B jsou různé konstanty. Existence takové rovnice ovšem znamená, že soustava není řešitelná.

Pozorování 3.22. Uvažujme soustavu $\mathcal{S}$. Dále uvažujme soustavu $\mathcal{T}$, která vznikla dosazením dosaditelné rovnice nebo eliminací triviální rovnice. Pak pro libovolné řešení σ soustavy $\mathcal{S}$ snadno sestrojíme řešení τ soustavy $\mathcal{T}$ tak, že zapomeneme hodnoty na vyhozených proměnných. Naopak pro řešení τ soustavy $\mathcal{T}$ můžeme zpětně sestrojit příslušné řešení σ. Toto řešení je určeno jednoznačně právě tehdy, když jsme provedli dosazení a odstranili pouze jednu proměnnou. Opět tedy budeme hovořit o odvozených řešeních.

Dále uvažujme smysluplný typový homomorfismus t soustavy $\mathcal{S}$ a typový homomorfismus t_2 soustavy $\mathcal{T}$ vzniklý zapomenutím hodnot na odstraněných proměnných. Pak existuje řešení typu t soustavy $\mathcal{S}$, právě když existuje řešení typu t_2 soustavy $\mathcal{T}$.

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.

3.1.4 Krácení překryvů

Definice 3.23. Uvažujme kvadratickou soustavu $\mathcal{S}$, typový homomorfismus t v této soustavě a proměnnou x. *Překryvem* proměnné x v řešení σ rozumíme dvojici výskytů této proměnné $\left\{\mathcal{Q}^{\mathcal{S}}_{r,b,i}, \mathcal{Q}^{\mathcal{S}}_{r,1-b,j}\right\}$ takovou, že tyto výskyty leží na různých stranách jedné rovnice a navíc

$$0 < \left||t(\mathrm{Pref}_i(S_{r,b}))| - |t(\mathrm{Pref}_j(S_{r,1-b}))|\right| \leq |t(x)|.$$

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

Pozorování 3.24. Pokud proměnná nemá netriviální 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.

Poznámka 3.25. Mějme kvadratickou soustavu $\mathcal{S}$, její řešení σ a překryv $\left\{\mathcal{Q}^{\mathcal{S}}_{r,b,i}, \mathcal{Q}^{\mathcal{S}}_{r,1-b,j}\right\}$ proměnné x v řešení σ. BÚNO předpokládejme $k = |t(\mathrm{Pref}_i(S_{r,b}))| - |t(\mathrm{Pref}_j(S_{r,1-b}))|$ kladné (jinak prohodíme strany). Pak definujeme kandidáta na řešení τ odstraněním pozic

$$\mathrm{Pref}_k\left(Z_t\left(\mathcal{Q}^{\mathcal{S}}_{r,b,i}\right)\right) \cup \mathrm{Pref}_k\left(Z_t\left(\mathcal{Q}^{\mathcal{S}}_{r,1-b,j}\right)\right)$$

Tedy proměnná x je v tomto řešení o k kratší.

Tvrzení 3.26. Kandidát τ na řešení popsaný v předchozí poznámce je skutečně řešením rovnice $\mathcal{S}$.

Důkaz: Slovo

$$\left(\mathrm{Pref}_k\left(Z_t\left(\mathcal{Q}^{\mathcal{S}}_{r,1-b,j}\right)\right)\beta\right)\mathrm{Pref}_k\left(Z_t\left(\mathcal{Q}^{\mathcal{S}}_{r,b,i}\right)\right)$$

je faktorem pozic. Pak obraz tohoto slova v homomorfismu $\tilde{\sigma}\Psi_t$ je roven

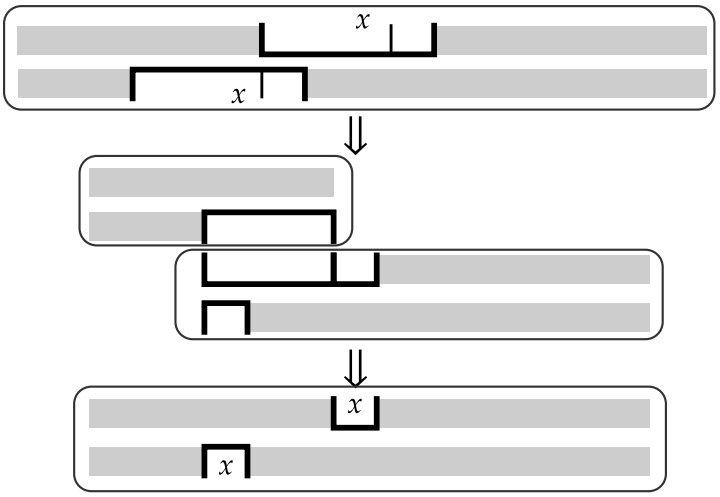
$$\left(\mathrm{Pref}_k(\sigma(x))\right)^2.$$

Nezáleží, kterou polovinu dvakrát zopakovaného slova odebereme, proto

$$\tau(S_{r,b}) = \tau(S_{r,1-b}).$$

∎

Definice 3.27. Uvažme řešení σ kvadratické soustavy $\mathcal{S}$ a v tomto řešení netriviální překryv $\left\{\mathcal{Q}^{\mathcal{S}}_{r,b,i}, \mathcal{Q}^{\mathcal{S}}_{r,1-b,j}\right\}$ proměnné x. Opakovaně odvozujeme menší řešení postupem popsaným v předchozí poznámce, dokud má proměnná x netriviální překryv. Takto vzniklému řešení říkáme *řešení vzniklé pokrácením překryvu* $\left\{\mathcal{Q}^{\mathcal{S}}_{r,b,i}, \mathcal{Q}^{\mathcal{S}}_{r,1-b,j}\right\}$.



Obrázek 3.2. Pokrácení překryvu proměnné x

Stejně tak, pokud máme pouze typový homomorfismus t v kvadratické soustavě $\mathcal{S}$ a netriviální překryv $\left\{\mathcal{Q}^{\mathcal{S}}_{r,b,i}, \mathcal{Q}^{\mathcal{S}}_{r,1-b,j}\right\}$ proměnné x, definujeme typový homomorfismus t_2 vzniklý pokrácením překryvů jako takový typový homomorfismus, který se na všech proměnných různých od x shoduje s t a dále $|t_2(x)|$ je nejmenší možná kladná délka splňující

$$|t_2(x)| \equiv |t(x)| \pmod{k},$$

kde $k = |t(\mathrm{Pref}_i(S_{r,b}))| - |t(\mathrm{Pref}_j(S_{r,1-b}))|$.

Definice 3.28. Operace „smazání prázdných proměnných“, „lámání rovnic“, „lámání výskytu v bodě“, „lámání před / za výskytem“, „vylamování výskytů“, „dosazování“, „eliminace triviální rovnice“ a „pokrácení překryvu“ budeme souhrnně označovat jako *makro-operace*.

3.2 Dvojitě exponenciální mez

Pozorování 3.29. Kvadratická soustava $\mathcal{S}$, ve které je každá strana každé rovnice neprázdná, má nejvýše tolik netriviálních rovnic, kolik má celkem zlomů. Pokud tedy $\mathcal{S}$ nemá žádné dosaditelné triviální rovnice, můžeme odhadnout její délku na základě počtu z zlomů a počtu c výskytů konstant.

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

Pozorování 3.30. 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 3.31. Uvažujme kvadratickou soustavu $\mathcal{S}$ s řešením σ a z té pomocí makro-operací odvozenou soustavu $\mathcal{T}$ s řešením τ. Platí $\tau \leq_c \sigma$. Dále uvažme dvě řešení τ_1, τ_2 soustavy $\mathcal{T}$ a zpětně odvozená řešení σ_1, σ_2 soustavy $\mathcal{S}$. Pokud platí $\tau_1 \leq_l \tau_2$, tak platí i $\sigma_1 \leq_l \sigma_2$. Pokud tedy řešení τ bylo l-minimální, bude takové i řešení τ.

Poznámka 3.32. Rozšířit předchozí pozorování na další druhy uspořádání nemůžeme. Dožadovat se $\tau \leq_l \sigma$ 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 nejkratší) 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 lámání a dosazování 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 3.33. 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 2.58.

Začneme s obecnou řešitelnou 2-soustavou $\mathcal{S}_1$ s z zlomy a jejím l-minimálním řešením σ_1. Předpokládáme, že v σ_1 není žádná proměnná prázdná (když tak smažeme prázdné proměnné). Postupně odvozujeme 2-soustavy $\mathcal{S}_i$ s l-minimálními řešeními σ_i bez prázdných proměnných tak, že vždy bude platit

- $\mathscr{C}_{\mathcal{S}_{i+1}} \subseteq \mathscr{C}_{\mathcal{S}_i}$,
- $\mathscr{V}_{\mathcal{S}_{i+1}} \subseteq \mathscr{V}_{\mathcal{S}_i}$,
- $\sigma_{i+1} \leq_l \sigma_i$,
- $|\sigma_{i+1}(\mathcal{S}_{i+1})| < |\sigma_i(\mathcal{S}_i)|$.
- $|\sigma_i(\mathcal{S}_i)| \leq 2|\sigma_{i+1}(\mathcal{S}_{i+1})|$.

V některém kroku k skončíme na soustavě $\mathcal{S}_k$ a jejím řešení σ_k délky 2. Vzhledem k tomu, že se bude jednat o l-nerostoucí posloupnost l-minimálních řešení, které se ostře zkracují, nesmí se v této posloupnosti žádná 2-soustava opakovat. Různých 2-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

$$|\sigma_1| \leq 2 \cdot 2^{2^p|\mathcal{S}|}.$$

Zbývá popsat, jak odvozujeme jednotlivé soustavy. Pokud zbyla pouze jedna rovnice tvaru $A = A$ pro nějakou konstantu A, ukončíme činnost. Jinak v každém kroku vybereme jednu rovnici $R_r = (S_{r,0}, S_{r,1})$ soustavy $\mathcal{S}_i$. Je-li tato rovnice dosaditelná resp. triviální, tak ji dosadíme resp. eliminujeme a jsme hotovi.

V opačném případě, pokud $|\sigma_i(S_{r,0}[0])| = |\sigma_i(S_{r,1}[0])|$, rozlomíme rovnici r ve dvojici zlomů

$$\left(\mathscr{Q}^{\mathcal{S}}_{r,0,0}, \mathscr{Q}^{\mathcal{S}}_{r,0,1}\right), \quad \left(\mathscr{Q}^{\mathcal{S}}_{r,1,0}, \mathscr{Q}^{\mathcal{S}}_{r,1,1}\right),$$

Následně eliminujeme vzniknuvší triviální rovnici a jsme opět hotovi.

Zbývá možnost, kdy mají $|\sigma_i(S_{r,0}[0])|$ a $|\sigma_i(S_{r,1}[0])|$ různé délky. Zvolme b tak, aby $|\sigma_i(S_{r,b}[0])| < |\sigma_i(S_{r,1-b}[0])|$. V takovém případě rozlomíme rovnici za výskytem $\mathscr{Q}^{\mathcal{S}}_{r,b,0}$. To rozdělí proměnnou $\Phi\left(\mathscr{Q}^{\mathcal{S}}_{r,1-b,0}\right)$ na dvě – nazveme je xy. Pro proměnnou y použijeme jakožto prvek množiny $\mathscr{V}_{\mathcal{S}_i}$ odstraněnou proměnnou $\Phi\left(\mathscr{Q}^{\mathcal{S}}_{r,1-b,0}\right)$ a provedeme dosazení výskytu proměnné x v nově vzniklé dosaditelné triviální rovnici. Tím i nová proměnná x zmizí a my dosáhneme požadované vlastnosti řešení σ_{i+1}. ∎

3.3 Co by stačilo pro jednoduše exponenciální mez – pokračování

Tvrzení 3.34. Platnost hypotézy 1.22 o jednoduše exponenciální mezi pro kvadratické soustavy bychom měli potvrzenu již za tohoto předpokladu: „Existuje rostoucí polynom p s následující vlastností. Pro každou 2-soustavu $\mathcal{S}$ s alespoň jednou konstantou existuje její konstanta A a řešení σ, že $|\sigma(\mathcal{S})|_A \leq 2^{p(|\mathcal{S}|)}$." Jinými slovy, pro důkaz jednoduše exponenciální meze na velikost nejkratšího řešení by stačilo dokázat jednoduše exponenciální mez pro počet pozic nejméně frekventované konstanty v řešení.

Důkaz: Díky pozorováním 3.18 a 2.52 stačí dokázat jednoduše exponenciální mez pro 2-soustavy bez dosaditelných rovnic. Vezměme si tedy takovou 2-soustavu $\mathcal{S}$, počet zlomů v ní označme z. Dále označme $j = |\mathscr{C}_S|$ a $\mathcal{S}_j = \mathcal{S}$. Předpokládáme $j \geq 1$, jinak by bylo $\min(\mathcal{S}) = 0$.

Popíšeme postup, jak z 2-soustavy $\mathcal{S}_{i+1}$ vytvořit soustavu $\mathcal{S}_i$ opět bez dosaditelných rovnic a s i konstantami. Tento postup budeme provádět, až do $\mathcal{S}_1$.

Z předpokladu máme řešení σ soustavy $\mathcal{S}_{i+1}$ a konstantu A takovou, že platí $|\sigma(\mathcal{S}_{i+1})|_A \leq 2^{p(|\mathcal{S}_{i+1}|)}$. Jsou-li v σ prázdné proměnné, smažeme je ze soustavy. Označme $k_{i+1} = |\sigma(\mathcal{S}_{i+1})|_A$. Opakujme krok „vylom jeden ze dvou výskytů konstanty A a následně vezmi vzniklou triviální rovnici, jejíž jedna strana je tvořena konstantou A, a eliminuj ji", dokud se konstanta A v soustavě nachází. Výsledek označíme $\mathcal{T}_i$ s řešením τ bez prázdných proměnných. Počet zlomů v soustavě $\mathcal{T}_i$ nepřevyšuje počet zlomů v soustavě $\mathcal{S}_{i+1}$, jakkoli $\mathcal{T}_i$ může mít oproti $\mathcal{S}_{i+1}$ exponenciální množství rovnic. Nakonec definujeme $\mathcal{S}_i$ s řešením σ' jako soustavu, která vznikne z $\mathcal{T}_i$ podosazováním všech dosaditelných rovnic. Nemohla zbýt žádná rovnice tvaru $x = x$ pro nějakou proměnnou x. Tato proměnná by totiž musela mít v σ' nenulovou délku, protože nemáme prázdné proměnné. To by ale znamenalo, že σ' není nejkratší řešení, tedy ani τ ani σ.

Krok „vylomení a eliminace triviální rovnice" provedeme celkem $(k_{i+1}/2)$-krát (v každém kroku snížíme $|\sigma(\mathcal{S}_{i+1})|_A$ o dva), tedy můžeme odhadnout počet rovnic v soustavě $\mathcal{T}_i$:

$$\langle\langle \mathcal{T}_i \rangle\rangle \leq \langle\langle \mathcal{S}_{i+1} \rangle\rangle + (k_{i+1}/2) \leq \langle\langle \mathcal{S}_{i+1} \rangle\rangle + 2^{p(|\mathcal{S}_{i+1}|)}.$$

Absence dosaditelných triviálních rovnic v $\mathcal{S}_i$ dává dle pozorování 3.29

$$\langle\langle \mathcal{S}_i \rangle\rangle \leq z + i, \quad |\mathcal{S}_i| \leq z + 2\langle\langle \mathcal{S}_i \rangle\rangle \leq 3z + 2j.$$

Odhadneme tak

$$\langle\langle \mathcal{T}_i \rangle\rangle \leq \langle\langle \mathcal{S}_{i+1} \rangle\rangle + 2^{p(|\mathcal{S}_{i+1}|)} \leq z + j + 2^{p(3z+2j)} \leq 2^{p_2(z+j)}$$

pro nějaký rostoucí polynom p_2 nezávislý na $\mathcal{S}$. Polynom p_2 lze na základě polynomu p zvolit například tak, aby splňoval $\forall x \in \mathbb{R} : x + 2^{p(3x)} \leq 2^{p_2(x)}$.

Dále díky pozorování 3.18 máme

$$2^{p_2(j+z)} \cdot \min(\mathcal{S}_i) \geq \min(\mathcal{T}_i) \geq \min(\mathcal{S}_{i+1}) - 2^{p(|\mathcal{S}_{i+1}|)} \geq \min(\mathcal{S}_{i+1}) - 2^{p(3z+2j)},$$

přitom druhá nerovnost plyne z toho, že do nejmenšího řešení soustavy $\mathcal{T}_i$ můžeme vrátit $k_{i+1}/2$ triviálních rovnic, které mají v řešení σ délku 2 a zpětně odvodit stejně dlouhé řešení soustavy $\mathcal{S}_{i+1}$. Poslední nerovnost plyne odhadů výše.

Úpravou dostáváme

$$\min(\mathcal{S}_{i+1}) \leq 2^{p_2(z+j)} \cdot \min(\mathcal{S}_i) + 2^{p(3z+2j)} \leq 2^{p_3(z+j)} \cdot \min(\mathcal{S}_i)$$

Pro nějaký rostoucí polynom p_3 nezávislý na $\mathcal{S}$.

Velikost nejmenšího řešení soustavy $\mathcal{S}_1$ shora opět odhadneme pomocí předpokladu, protože se v jeho řešení vyskytuje jen jedna konstanta.

$$\min(\mathcal{S}_1) \leq 2^{p(|\mathcal{S}_1|)} \leq 2^{p_3(z+j)}.$$

K odhadnutí $\min(\mathcal{S})$ již stačí použít nerovnost $\min(\mathcal{S}_{i+1}) \leq 2^{p_3(z+j)} \cdot \min(\mathcal{S}_i)$:

$$\min(\mathcal{S}) \leq (2^{p_3(z+j)})^j = 2^{j \cdot p_3(z+j)}.$$

Tvrzení 3.35. Nechť je dána řešitelná 2-soustava $\mathcal{S}$ a její proměnná x. Pak existuje řešitelná 2-soustava $\mathcal{R}$ o jedné rovnici a řešení σ soustavy $\mathcal{S}$ splňující

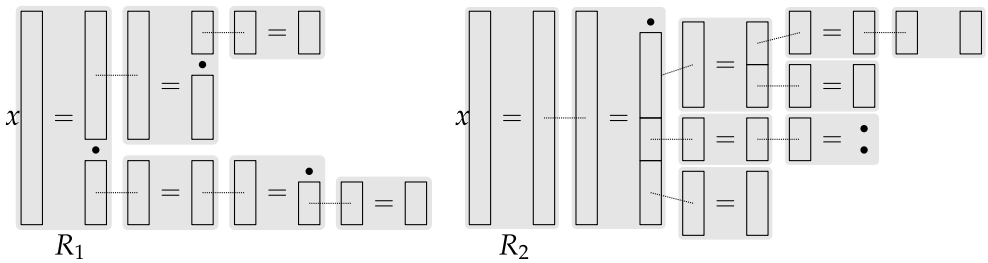
- $\mathscr{C}_\mathcal{R} \subseteq \mathscr{C}_\mathcal{S}$,
- $|\mathcal{R}| \leq |\mathcal{S}|$,
- $2|\sigma(x)| = \min \mathcal{R}$.

Důkaz: Předpokládáme, že neexistuje řešení soustavy $\mathcal{S}$, ve kterém by měla x nulovou délku. Kdyby takové řešení existovalo, stačilo by jej zvolit za σ a za soustavu $\mathcal{R}$ kupříkladu $x = x$.

Uvažme libovolné řešení σ_1 soustavy $\mathcal{S}$. Z něj následujícím postupem nezvyšujícím počet zlomů ani celkovou délku řešení odvodíme řešení σ_2 soustavy $\mathcal{S}_2$. Nejprve odstraníme ze soustavy $\mathcal{S}$ na základě σ_1 všechny prázdné proměnné. V dalším postupu již budou všechny proměnné neprázdné. 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. 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 nějakém 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 σ_2 a dva výše popsané zakořeněné stromy s vrcholy na rovnicích $\mathcal{S}_2$ takové, že každá proměnná má buď oba své výskyty v těchto stromech nebo oba mimo ně.



Obrázek 3.3. Ukázka sestavených stromů na základě postupu v důkazu

Uvažme soustavu $\mathcal{S}_3$ sestávající pouze z rovnic těchto stromů – to je 2-soustava. Po podosazování všech dosaditelných rovnic v $\mathcal{S}_3$ dostáváme 2-soustavu $\mathcal{R}$ o jedné rovnici jejíž obě strany se mají rovnat již redukované proměnné x. První požadovaná vlastnost na soustavu $\mathcal{R}$ 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 σ. Uvažme nejkratší řešení σ_4 soustavy $\mathcal{R}$ a řešení σ_3 jako řešení soustavy $\mathcal{S}_3$ odvozené z σ_4. To znamená

$$\min(\mathcal{R}) = |\sigma_4(\mathcal{R})| = 2|\sigma_3(x)|.$$

Nyní v řešení σ_2 použijeme u všech proměnných soustavy $\mathcal{S}_3$ hodnoty z řešení σ_3 namísto původních hodnot z σ_2. Tím dostaneme řešení σ_2' soustavy $\mathcal{S}_2$, ve kterém $2|\sigma_2(x)| = \min(\mathcal{R})$. Odtud již jen zpětně odvodíme řešení soustavy $\mathcal{S}$ s touto vlastností. ■

Důsledek 3.36. Pokud bychom měli jednoduše exponenciální mez na nejkratší délku řešení pro řešitelné 2-soustavy o jedné rovnici, měli bychom z věty 2.66 i jednoduše exponenciální mez pro všechny řešitelné 2-soustavy, tedy i pro všechny řešitelné kvadratické soustavy.

Důsledek 3.37. Pokud bychom měli jednoduše exponenciální mez na nejkratší možnou velikost nejkratší proměnné v řešení pro 2-soustavy o jedné rovnici, měli bychom na základě opětovného využití tvrzení 2.66 i jednoduše exponenciální mez pro všechny řešitelné kvadratické soustavy.

Kapitola 4

Složitost řešení kvadratických soustav

Tato kapitola ukazuje souvislost práce s výpočetní složitostí určení řešitelnosti kvadratických rovnic. Obě části této kapitoly vychází z článku [1].

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

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

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

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

Pokud mezi těmito vrcholy naopak hrana nevede, sestavíme dvojici rovnic

$$\begin{aligned} B &= b_{v,w} \star, \\ B &= b_{w,v} \star. \end{aligned} \tag{3}$$

A nakonec sestavíme rovnici

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

Pozorování 4.1. Sestavená soustava 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). Všechny proměnné značené hvězdičkou se v soustavě vyskytují jen jednou.

Tvrzení 4.2. Existovala-li v grafu nezávislá množina N velikosti k, pak je sestavená soustava řešitelná.

Důkaz: Stačí volit $a_v = A$ pro všechna $v \in N$ a $b_{v,w} = B$ 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 4.3. Je-li sestavená soustava řešitelná, pak v grafu existuje nezávislá množina velikosti k.

Důkaz: Označme σ řešení soustavy. Na základě rovnic (1) může každé $\sigma(a_v)$ obsahovat nejvýše jedno A. Dále podle rovnice (4) se $\sigma(a_v)$ mohou skládat pouze z konstant A, proto $\sigma(a_v) = A$ pro jistou množinu $N \subseteq V$, zbylá $\sigma(a_v)$ 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 $\sigma(b_{v,v_1} b_{v,v_2} \dots b_{v,v_n})$ délku alespoň n. Navíc podle rovnic (2) resp. (3) je každá proměnná $b_{x,y}$ nejvýše jednoprvková, proto pro všechna $x \in V$ je $\sigma(b_{v,x}) = B$ a podobně i $\sigma(b_{w,x}) = B$. Tedy $\sigma(b_{v,w} b_{w,v}) = BB$ a proto díky rovnicím (2) mezi těmito vrcholy nevede hrana. ∎

Poznámka 4.4. Bylo by možné sestavením ještě další rovnice soustavu doplnit, aby se v ní každá proměnná vyskytovala právě dvakrát a přesto aby splňovala požadované vlastnosti. Na druhou stranu předvedený postup značně využívá skutečnosti, že v soustavě nejsou rozrůzněné konstanty, tedy není jasné, zda i řešitelnost 2-soustav je NP-těžký problém.

4.2 Algoritmus pro ověření řešitelnosti rovnice s předepsanými délkami

Pokud bychom dostali kvadratickou soustavu a v ní nějaký typ, mohli bychom v lineárním čase vzhledem k délce tohoto typu ověřit, zda existuje řešení, ve kterém mají proměnné onu předepsanou délku. Jednoduše najdeme protějšek každé konstanty a ověříme, zda si stejné konstanty odpovídají. Takový postup by byl lineární vzhledem k délce tohoto typu. Ukážeme ale, že je možné postupovat ještě efektivněji – totiž polynomiálně vzhledem k součtu délky soustavy a logaritmu délky typu.

Pozorování 4.5. Pokud provedeme makro-operaci na soustavu $\mathcal{S}$ se smysluplným typovým homomorfismem t, čímž vznikne soustava $\mathcal{T}$ s typovým homomorfismem t_2, bude existovat řešení soustavy $\mathcal{S}$ typu t, právě když bude existovat řešení soustavy $\mathcal{T}$ typu t_2.

Tvrzení 4.6. Existuje algoritmus, který v polynomiálním čase vzhledem k velikosti vstupu ověří, zda existuje řešení dané soustavy $\mathcal{S}$ předepsaného typového homomorfismu t. Tento algoritmus na vstupu dostane soustavu $\mathcal{S}$ a pro každou proměnnou x její délku $t(x)$ zapsanou ve dvojkové soustavě. Na výstupu odpoví zda příslušné řešení existuje nebo ne.

Důkaz: Popíšeme algoritmus.

Nejprve ověříme, zda je typ t smysluplný. Pokud ne, rovnou známe negativní odpověď. Dále tedy budeme předpokládat tuto vlastnost.

S délkami proměnných je možné v polynomiálním čase provádět aritmetické operace, konkrétně sčítání, odčítání, zbytek po dělení. Tedy i makro-operace je možné provádět v polynomiálním čase vzhledem k aktuální délce soustavy.

Na začátku ze soustavy na základě typu t smažeme všechny prázdné proměnné. Dále budeme soustavu měnit pomocí makro-operací, tedy prázdné proměnné se v ní již neobjeví a hodnota výroku „existuje řešení $\mathcal{S}$ typu t“ se nezmění. Souběžně se soustavou upravujeme i typový homomorfismus t (přesněji algoritmus upravuje pouze typ, na pojmenovávání pozic nemá čas).

Algoritmus postupuje podle následujících bodů

1) Dokud existují dosaditelné rovnice, dosazuj je. Dokud existují eliminovatelné triviální rovnice, eliminuj je.

2) Pokud je v rovnici triviální rovnice $A = B$, kde A, B jsou různé konstanty, odpověz, že je soustava neřešitelná a ukonči činnost. Pokud v soustavě nezbyly žádné rovnice, odpověz že je soustava řešitelná a ukonči činnost.
3) Jinak najdi v soustavě rovnici $R_r = (S_{r,0}, S_{r,1})$, pro níž je $|t(S_{r,0})|$ nejvyšší možné.
4) Uvaž index $i = \lfloor |t(S_{r,0})|/2 \rfloor$ a výskyt $\boldsymbol{v} = Z_t^{-1}\left(\mathscr{P}_{r,0,i}^{\mathcal{S},t}\right)$.
5) Je-li $\Phi(\boldsymbol{v})$ proměnná a má-li v t netriviální překryv, pokraťjej.
6) Vylom výskyt $\boldsymbol{v}$.
7) Vraťse na bod (1).

Algoritmus zřejmě skončí a odpoví správně, vzhledem k tomu, že snižuje délku typového homomorfismu a provádí se soustavou, co do existence řešení příslušného typu, ekvivalentní úpravy.

Zbývá si rozmyslet, že tento algoritmus skutečně běží v polynomiálním čase. Snadno nahlédneme, že každý krok skutečně probíhá v polynomiálním čase vzhledem k aktuální délce soustavy a součtu logaritmů délek proměnných. 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 (7). Zbývá ukázat, že počet cyklů algoritmu je možné polynomiálně omezit.

Pro soustavu $\mathcal{S}$ s typem t a přirozené číslo n označme $\langle\langle t(\mathcal{S})\rangle\rangle_n$ jako počet rovnic $S_{r,0} = S_{r,1}$, pro které platí $t(S_{r,0}) \leq n$.

Nakonec při každém provádění kroku (3) cyklu najdeme nejvyšší k, pro které platí $\langle\langle t(\mathcal{S})\rangle\rangle_{2^k} > 0$ a nazvěme jej *složitostí* typu.

Pro dokončení důkazu stačí již jen několik pozorování:

- Složitost typu na začátku je lineárně omezená délkou vstupu.
- V každém kroku (3) je $\langle\langle\mathcal{S}\rangle\rangle_{2^k}$ menší nebo rovno počtu zlomů vstupní soustavy.
- Rovnice vznikající v kroku (6), které nejsou dosaditelné, budou mít v typu t nejvýše poloviční délku oproti R_r, ze které vznikly.
- Složitost k aktuálního typu tak neroste, a za předpokladu, že zůstane stejná, se $\langle\langle\mathcal{S}\rangle\rangle_{2^k}$ sníží 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 času běhu algoritmu. ∎

Důsledek 4.7. Pokud platí hypotéza 1.22 o jednoduše exponenciální mezi, je řešitelnost kvadratických soustav NP-úplný problém. Nedeterministický algoritmus bude pracovat tak, že uhodne typ nejkratšího řešení a následně v polynomiálním čase ověří, že opravdu existuje řešení daného typu.

Poznámka 4.8. Uvedený algoritmus má při důkladnějším zkoumání kvadratickou složitost, zdroj [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 důsledku zcela postačující.

Kapitola 5

O jisté konkrétní palindromické rovnici

Účelem této kapitoly je poskytnout sadu (proti)příkladů a na konkrétní rovnici demonstrovat, jak se kvadratické rovnice mohou chovat. Předvedeme soustavu a její exponenciálně velké řešení, které je c-minimální i l-minimální. Kdyby bylo nejkratší, měli bychom vyvrácenou hypotézu 1.23, nicméně takové není.

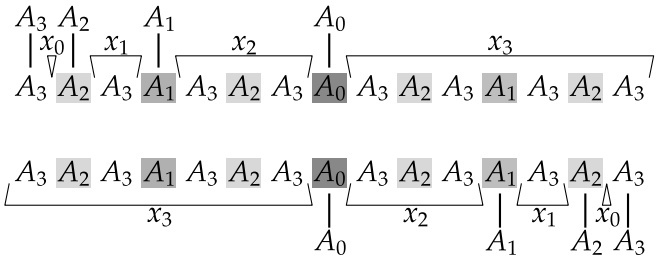
Definice 5.1. Pro dané $n \in \mathbb{N}$ definujeme 2-soustavu Pal_n o jedné rovnici jako

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

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

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

Pozorování 5.2. Zobrazení PalSol_n je opravdu řešením rovnice Pal_n a má délku $2(2^n - 1)$.



Obrázek 5.1. Rovnice Pal_4 s řešením PalSol_4

Pozorování 5.3. Pro $n \geq 3$ není PalSol_n nejkratším řešením, můžeme totiž definovat řešení σ délky $4n$ jako

$$\sigma(x_i) = A_i.$$

5.1 Mikro-chování Pal_n

Tvrzení 5.4. Pro libovolné n je řešení PalSol_n l-minimální. Dokonce, uspořádáme-li řešení σ abecedním uspořádáním podle n-tic

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

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 $\sigma_{\min}$, které je nejmenší v l-abecedním uspořádání. Vzhledem k tomu, že $\mathrm{PalSol}_n(x_0)$ je prázdné musí i $\sigma_{\min}(x_0)$ být prázdné. Uvažujme tedy

soustavu $\mathcal{R}$, která vznikne z Pal_n smazáním proměnné x_0. Řešení $\mathrm{PalSol}_n(x_0)$ i $\sigma_{\min}$ je tak možné vnímat jako řešení $\mathcal{R}$.

Dále uvažujme obecné řešení σ soustavy $\mathcal{R}$. Dokážeme (zanořenou) indukcí, že pro všechna $i \in \{1, \dots n-1\}$ je $\sigma(x_n)$ neprázdné a navíc $\sigma(x_n)[0] = A_{n-1}$.

(i) Uvažme (jednoznačně určený, na pravé straně rovnice) faktor pozic $\boldsymbol{F}$ délky 2, pro který $\Phi Z_\sigma^{-1}(\boldsymbol{F}) = A_{n-2}A_{n-1}$. Jeho protějšek je na levé straně a musí v něm ležet nějaký zlom. Jediný možný zlom je ten mezi výskytem konstanty A_{n-2} a proměnné x_1, protože levý výskyt proměnné A_{n-1} je v levé straně první. Proto $x_1[0] = A_{n-1}$.

(ii) Předpokládejme $x_i[0] = A_{n-1}$ a uvažme (jednoznačně určený, na pravé straně rovnice) faktor pozic $\boldsymbol{F}$ délky 2, pro který $\Phi Z_\sigma^{-1}(\boldsymbol{F}) = A_{n-2-i}x_i$. Jeho protějšek je na levé straně a musí v něm ležet nějaký zlom. Jediný možný zlom je ten mezi výskytem konstanty A_{n-2-i} a proměnné x_{i+1}, protože $\tilde{\sigma}\Psi_\sigma(\boldsymbol{F}) = A_{n-2-i}A_{n-1}$ a levý výskyt proměnné A_{n-1} je v levé straně první. Máme tak i $x_{i+1}[0] = A_{n-1}$.

Obdobně shledáme, že každá proměnná musí konstantou A_{n-1} i končit.

Nyní, kdykoli vezmeme libovolný faktor pozic $\boldsymbol{F}$ délky dva, musí být jeho protějškem nějaká dvojice pozic $\boldsymbol{pq}$, o které jsme již zjistili, že právě jedna z konstant $\tilde{\sigma}\Psi_\sigma(\boldsymbol{p}), \tilde{\sigma}\Psi_\sigma(\boldsymbol{q})$ je A_{n-1}. Tedy i právě jedna z konstant ve slově $\tilde{\sigma}\Psi_\sigma(\boldsymbol{F})$ je A_{n-1}. Z toho plyne, že konstanta A_{n-1} je v řešení „na střídačku“. Přesněji pro každé $i \in \{1, \dots, n-1\}$ je $|\sigma(x_i)|$ liché a navíc pro každé $k \in \{0, \dots |\sigma(x_i)| - 1\}$ je $\sigma(x_i)[k] = A_n$ právě když k je sudé.

Popíšeme bijekci zachovávající l-abecední uspořádání mezi řešeními rovnice $\mathcal{R}$ a řešeními rovnice Pal_{n-1}. Tím z indukčního předpokladu přejde řešení $\sigma_{\min}$ 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 určena). K vítězství pak stačí, že i PalSol_n tato bijekce zobrazí na PalSol_{n-1}.

Postup je jednoduchý. Z řešení σ odstraníme všechny pozice konstanty A_{n-1} a taktéž ze soustavy $\mathcal{R}$ odstraníme oba výskyty této konstanty. Tím dostaneme řešení σ' soustavy $\mathcal{R}'$, 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 uspořádání, 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-1} je možné snadno ověřit porovnáním délek proměnných v příslušných řešeních. ■

Tvrzení 5.5. Existuje reálná konstanta $c > 0$ a nekonečně mnoho neisomorfních řešitelných 2-soustav $\mathcal{S}$ o jedné rovnici, pro které platí $\min \mathcal{S} > c|S|^2$.

Důkaz: Za soustavy $\mathcal{S}$ volíme rovnice Pal_n, s následujícími dvěma úpravami:

- Smažeme proměnnou x_0.
- Konstantu A_{n-1} nahradíme posloupností konstant $B_1B_2 \dots B_n$.

Tyto soustavy jsou řešitelné – stačí v řešeních PalSol_n provést patřičné nahrazení konstant. Dále díky tvrzení 2.46 existuje nejkratší neštěpící řešení, označme jej σ. Jak bylo ukázáno v důkazu předchozího tvrzení, musí každá proměnná $x_1, \dots, x_n$ začínat na $B_1 \dots B_n$. Toto řešení má tedy velikost alespoň n^2, zatímco délka soustavy je $3n - 2$. ■

5.2 Makro-chování Pal_n

Tvrzení 5.6. Pro libovolné n je řešení PalSol_n c-minimální. Dokonce, uspořádáme-li řešení σ abecedním uspořádáním podle n-tic

$$\left(|\sigma(\mathrm{Pal}_n)|_{A_0}, |\sigma(\mathrm{Pal}_n)|_{A_1}, \dots, |\sigma(\mathrm{Pal}_n)|_{A_{n-1}}\right)$$

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 $n \geq 2$ uvažme řešení $\sigma_{\min}$, které je nejmenší v c-abecedním uspořádání. Vzhledem k tomu, že $|\mathrm{PalSol}_n(\mathrm{Pal}_n)|_{A_0} = 2$, musí i $|\sigma_{\min}(\mathrm{Pal}_n)|_{A_0} = 2$. Obě pozice konstanty A_0 tak mají v řešení $\sigma_{\min}$ stejný index. Uvažujme tedy soustavu tří rovnic $\mathcal{S}$, která vznikne vyříznutím konstanty A_0 (je jedno, kterého výskytu). Triviální rovnici $A_0 = A_0$ můžeme eliminovat bez vlivu na c-abecední uspořádání.

Zbývá soustava

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

Dosazením proměnné x_{n-1} zbude soustava $\mathcal{R}$, která je až na posunutí indexů u konstant shodná se soustavou Pal_{n-1}. Navíc, máme-li obecné řešení σ soustavy S a označíme-li σ' odvozené řešení rovnice $\mathcal{R}$, bude pro libovolné $i \in \{1, \dots, n-1\}$

$$|\sigma(\mathrm{Pal}_n)|_{A_i} = 2|\sigma'(\mathrm{Pal}_n)|_{A_i}.$$

Proto $\sigma_{\min}$ po dosazení odpovídá nejmenšímu řešení Pal_{n-1} v c-abecedním uspořádání, tedy PalSol_{n-1}. Zpětným krokem dostaneme PalSol_n, což jsme chtěli dokázat. ∎

Tvrzení 5.7. Pro každé řešení σ rovnice Pal_n, ze kterého nelze odstranit neprázdnou množinu pozic, platí:

- Existuje právě jedna konstanta A_i, pro kterou $|\sigma(\mathrm{Pal}_n)|_{A_i} = 2$.
- Pro všechny ostatní konstanty A_i je $|\sigma(\mathrm{Pal}_n)|_{A_i}$ dělitelná čtyřmi.

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

Díky palindromicitě rovnice Pal_n musí v řešení σ existovat překryv proměnných nebo dvojice výskytů $\boldsymbol{v}$, $\boldsymbol{w}$ jedné konstanty A_i, pro které platí

$$Z_\sigma(\boldsymbol{v})\beta = Z_\sigma(\boldsymbol{w}).$$

První možnost nenastane díky tvrzení 3.26, nastane druhá – můžeme tedy na základě σ vylomit výskyt $\boldsymbol{v}$. Eliminujeme triviální rovnici $A_i = A_i$ a zbudou dvě rovnice R_0, R_1.

Zbývá si všimnout, že množina všech konstant a proměnných levé strany rovnice R_0 je totožná s množinou všech konstant a proměnných pravé strany rovnice R_1. Z toho pro libovolnou konstantu $A \neq A_i$ je $|\sigma(R_0)|_A = |\sigma(R_1)|_A$, tedy celkem je $|\sigma(\mathrm{Pal}_n)|_A$ dělitelné čtyřmi. ∎

5.3 Možné vylepšení – přidání podmínek

Samotná rovnice Pal_n jako kandidát na „protipříklad na polynomiální mez pro řešení kvadratických rovnic“ neobstojí. Jakkoli je 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 5.8. *Soustavou s podmínkami* rozumíme soustavu rovnic 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 soustavy s podmínkami pak rozumíme takové řešení, ve kterém navíc proměnné splňují zadané podmínky.

Tvrzení 5.9. Pro danou kvadratickou soustavu s p podmínkami $\mathcal{S}$ existuje kvadratická soustava $\mathcal{S}_2$ (bez podmínek), která splňuje

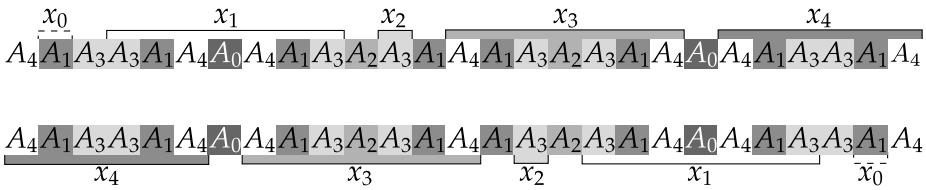
- $\min(\mathcal{S}_2) = \min(\mathcal{S})$,
- $|\mathcal{S}_2| \leq |\mathcal{S}| + 4p$.

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 proměnné nově založené pro každou podmínku. ∎

Pozorování 5.10. Řešení PalSol_n je řešením soustavy Pal_n s podmínkami „proměnná x_i musí obsahovat konstantu A_{n-i}“ pro $i \in \{1, \ldots, n-1\}$.

Hypotéza 5.11. Velikost nejkratšího řešení soustavy s podmínkami z předchozího pozorování není možné polynomiálně odhadnout.

Poznámka 5.12. Ani s takovými podmínkami není nutně řešení PalSol_n nejkratší. 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 5.2.



Obrázek 5.2. Obrázek demonstrující, že ani po přidání podmínek k rovnici Pal_5 nemusí být PalSol_5 nejkratší řešení.

Poznamenejme, že v zobrazeném řešení není žádná proměnná prázdná, ani žádná konstanta pouze dvakrát – to proto, že tu máme vedle sebe dvě konstanty A_3 a proměnné x_3 se překrývají. Avšak pokud bychom odstranili množinu pozic na základě tvrzení 2.69, proměnná x_2 by se stala prázdnou a nesplňovala by podmínku na obsahování konstanty A_3. Podobně, pokud bychom pokrátili proměnnou x_3, přestala by obsahovat konstantu A_2.

Kapitola 6

Dualita mikro a makro přístupu

Poslední kapitola je mírně poetičtějšího ražení nežli předchozí. Nabízí spíše pohled na věc než konkrétní řešení konkrétních problémů. Čtenář, který si přeje matematický text v zaběhaných kolejích, by jí mohl být mírně zaskočen. Nicméně kapitola objasňuje podobnost mikro-přístupu a makro-přístupu pro balancované 2-soustavy.

6.1 Motivace a intuice

Jak je patrno z předchozích kapitol 2, 3 a 5, postupy v mikro-přístupu jsou v jistém smyslu analogické k postupům v makro-přístupu, jakkoli oba přístupy vzešly ze značně odlišných úvah. Tabulka 6.1 zobrazuje pojmy a tvrzení z obou kapitol, které si odpovídají.

mikro-přístup	makro-přístup
délka proměnné $\lvert\sigma(x)\rvert$	počet pozic konstanty $\lvert\sigma(\mathcal{S})\rvert_A$
l-uspořádání	c-uspořádání
c-uspořádání	l-uspořádání
prázdná proměnná	konstanta s pouze dvěma pozicemi
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 2.66	tvrzení 3.34

Tabulka 6.1. Intuitivně duální pojmy na základě mikro/makro analogie

Tato kapitola vysvětluje, kde se zde tato dualita bere pro balancované 2-soustavy zavedením pojmu 2D rovnice – objektu podobnému balancovaným 2-soustavám, avšak se zaměnitelnými rolemi konstant a proměnných.

Poznámka 6.1. Řešení σ balancované 2-soustavy $\mathcal{S}$ je monoidový homomorfismus $(\mathcal{V}_\mathcal{S} \cup \mathcal{C}_\mathcal{S})^* \to \mathcal{C}_\mathcal{S}^*$ zachovávající konstanty. Na základě něj můžeme zavést „duální řešení“ σ^T jakožto homomorfismus $(\mathcal{V}_\mathcal{S} \cup \mathcal{C}_\mathcal{S})^* \to \mathcal{V}_\mathcal{S}^*$ zachovávající proměnné. Přitom pro libovolnou konstantu A a její levý výskyt $\boldsymbol{v}$ definujeme slovo z proměnných

$$\sigma^T(A) = \Phi Z_\sigma^{-1}\left(Z_\sigma(\boldsymbol{v})\beta\right)\Phi Z_\sigma^{-1}\left(Z_\sigma(\boldsymbol{v})\beta\gamma\beta\right)\ldots\Phi Z_\sigma^{-1}\left(Z_\sigma(\boldsymbol{v})\beta(\gamma\beta)^{\mathrm{MAX}}\beta\gamma\right).$$

Zajisté tu existují podmínky, které duální řešení musí splňovat. Zjistíme, že jsou v podstatě analogické opět balancovaným soustavám.

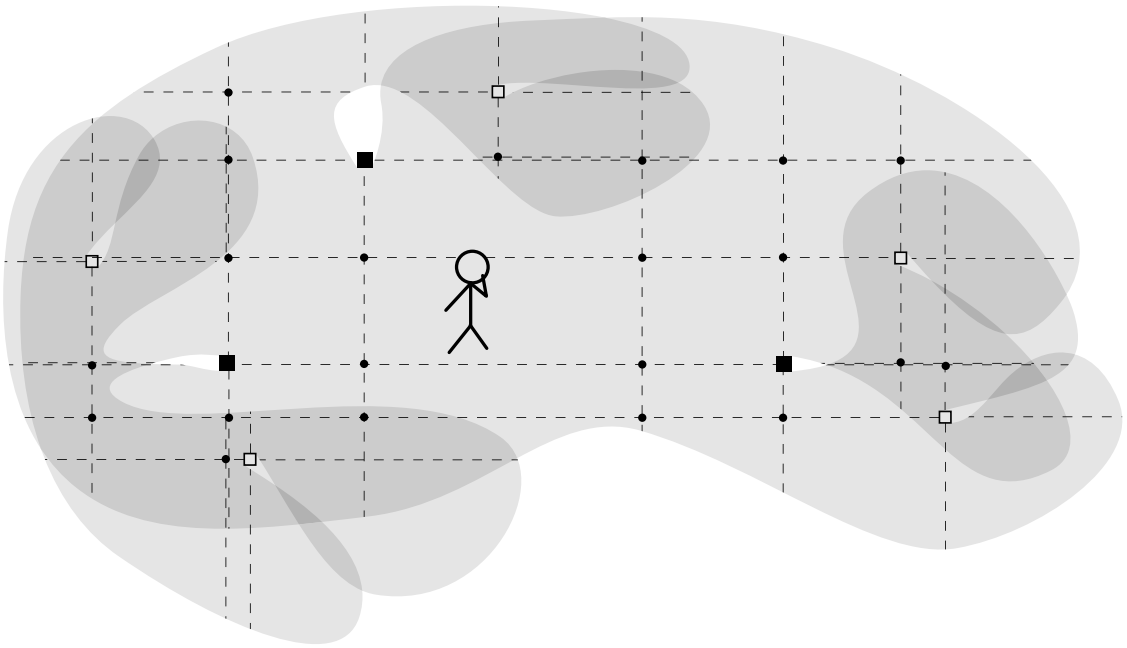
6.2 2D rovnice

V definici 6.2 a 6.4 je zavedena 2D rovnice a její řešení jako abstraktní algebraická struktura. Pro čtenáře může být nicméně užitečné tyto definice ilustrovat nejprve třeba násle-

dující představou. Množina X v řešení je konečná množina stejně velkých čtvercových dlaždic rozmístěných (víceméně) do čtvercové sítě. Virtuální chodec vždy na jedné dlaždici stojí a může udělat krok na jih ($\triangledown$), na východ ($\triangleright$), na sever ($\triangle$) nebo na západ ($\triangleleft$) a tím se dostane na sousední dlaždici. Jeho kroky jsou vratné, tj. například po kroku $\triangle$ následovaném $\triangledown$ se dostane tam, kde byl na začátku. Mříž dlaždic není ohraničena, tj. pokud chodec bude kráčet dostatečně dlouho třeba jen na sever a nenarazí na překážku (viz níže tzv. sloup), dostane se po konečně krocích tam, kde byl na začátku, protože množina dlaždic X je konečná.

Občas je v tomto světě místo některé dlaždice čtyřhranný sloup o základně shodné velikosti, jako dlaždice. Má-li chodec těsně před sebou sloup a udělá krok směrem ke sloupu, posadí se na stěnu, kterou před sebou vidí. Další krok stejným směrem nemá povolen. Může ale odbočit doprava, doleva nebo se vrátit. Kroky těmito směry realizuje v tomto případě stejně, jako kdyby stál na základně sloupu a žádné stěny mu nepřekážely. Představme si, že třeba těsně na sever od chodce je osamocený sloup. Krok $\triangle$ způsobí, že se chodec posadí na jižní stěnu sloupu. Následující krok $\triangle$ nemá chodec povolen. Ale třeba krok $\triangleright$ ho zavede na dlaždici těsně na východ od sloupu a pokud poté provede $\triangleleft$, usadí se na východní stěnu sloupu.

Pokud chodec provede třeba 6 kroků na sever, dále 6 kroků na východ, dále 6 kroků na jih a nakonec 6 kroků na západ, dostane se do stejného místa, jako byl, pokud jeho cesta neobešla nějaký sloup. Sloupy mají v sobě bludné kořeny, které zavedou chodce po obejití sloupu do jiného místa čtvercové sítě, než byl. Každý sloup má své magické číslo k, které značí, že nutné sloup k-krát obejít, aby se chodec vrátil tam, kde byl. Sloup tedy typicky nemá jen 4 stěny, ale $4k$ vzájemné různých stěn.



Obrázek 6.1. Umělecké ztvárnění světa řešení 2D rovnice, čtverečky značí sloupy, puntíky průsečíky cest od sloupů

V následujících definicích 2D soustavy a jejího řešení se používají pojmy, které lze ilustrovat výše uvedenou představou takto:

- $\mathscr{C}^{\triangledown}$, $\mathscr{V}^{\triangleright}$, $\mathscr{C}^{\triangle}$, $\mathscr{V}^{\triangleleft}$ jsou postupně množiny všech jižních stěn, východních stěn, severních stěn a západních stěn všech sloupů.
- Je-li $a \in \mathscr{C}^{\triangledown}$, pak pošleme chodce z místa a tak, že dělá jen kroky jižním směrem tak dlouho, až narazí na severní stěnu nějakého (třeba jiného) sloupu. Tato severní stěna je značena $\varphi(a)$. Podobně můžeme chodce poslat z prvku množiny $\mathscr{V}^{\triangleright}$ na východ,

z prvku množiny $\mathscr{C}^{\triangle}$ na sever a z prvku množiny $\mathscr{V}^{\triangleleft}$ na západ. Zobrazení φ najde vždy stěnu na konci této cesty, tedy přímo viditelnou stěnu daným směrem.
- Pro dvě dlaždice a, b platí, že $a \approx b$, právě když $a = b$. Pro dvě stěny sloupu a, b platí $a \approx b$, právě když se rovnají, nebo spolu sousedí na stejném sloupu (například jižní stěna sloupu sousedí s východní). Řetízek těchto sousedů má postupně různé prvky z množin $\mathscr{C}^{\triangledown}$, $\mathscr{V}^{\triangleright}$, $\mathscr{C}^{\triangle}$, $\mathscr{V}^{\triangleleft}$ a pro jeden sloup má délku $4k$.
- Zobrazení $\pi_{\mathscr{V}}$ (projekce k západu) zobrazí všechny dlaždice společného řádku do stěny sloupu, kterou na západě tento řádek končí.
- Zobrazení $\pi_{\mathscr{C}}$ je projekce k severu a je analogická k $\pi_{\mathscr{V}}$.

Souvislost této představy s balancovanými 2-soustavami je vysvětlena v tvrzení 6.11.

Definice 6.2. *2D rovnicí* rozumíme uspořádanou šestici

$$(\mathscr{C}^{\triangledown}, \mathscr{V}^{\triangleright}, \mathscr{C}^{\triangle}, \mathscr{V}^{\triangleleft}, \varphi, \approx),$$

kde první čtyři prvky jsou konečné neprázdné disjunktní množiny, φ je sama k sobě inverzní bijekce množiny $(\mathscr{C}^{\triangledown} \cup \mathscr{V}^{\triangleright} \cup \mathscr{C}^{\triangle} \cup \mathscr{V}^{\triangleleft})$ do sebe, $\approx$ je symetrická reflexivní relace na té samé množině, a navíc platí následující požadavky.

- Bijekce φ „posílá na opačný konec“, tedy
  - Pro $a \in \mathscr{C}^{\triangledown}$ je $\varphi(a) \in \mathscr{C}^{\triangle}$,
  - Pro $a \in \mathscr{V}^{\triangleright}$ je $\varphi(a) \in \mathscr{V}^{\triangleleft}$,
  - Pro $a \in \mathscr{C}^{\triangle}$ je $\varphi(a) \in \mathscr{C}^{\triangledown}$,
  - Pro $a \in \mathscr{V}^{\triangleleft}$ je $\varphi(a) \in \mathscr{V}^{\triangleright}$,
- Pro $a \in \mathscr{C}^{\triangledown}$
  - existuje právě jedno $b \in \mathscr{C}^{\triangledown}$ splňující $b \approx a$, konkrétně je $b = a$,
  - existuje právě jedno $b \in \mathscr{V}^{\triangleleft}$ splňující $b \approx a$,
  - existuje právě jedno $b \in \mathscr{V}^{\triangleright}$ splňující $b \approx a$,
  - neexistuje žádné $b \in \mathscr{C}^{\triangle}$, které by splňovalo $b \approx a$.
- Předchozí bod platí i pro zbylé 3 orientace šipek (tj. když se čtveřicí množin $(\mathscr{C}^{\triangle}, \mathscr{V}^{\triangleright}, \mathscr{C}^{\triangledown}, \mathscr{V}^{\triangleleft})$ provádíme cyklické záměny).

Z historických důvodů budeme prvkům množiny $\mathscr{C}^{\triangledown}$ říkat konstanty a prvkům množiny $\mathscr{V}^{\triangleright}$ budeme říkat proměnné. Tato terminologie vyplyne z podkapitoly 6.3.

Definice 6.3. *Velikostí 2D rovnice* rozumíme celkový počet proměnných a konstant, tedy $|\mathscr{C}^{\triangledown}| + |\mathscr{V}^{\triangleright}| = 2|\mathscr{V}^{\triangleright}|$.

Definice 6.4. *Řešením 2D rovnice* rozumíme sedmici

$$(X, \pi_{\mathscr{C}}, \pi_{\mathscr{V}}, \triangledown, \triangleright, \triangle, \triangleleft),$$

kde X je konečná množina, $\pi_{\mathscr{C}}, \pi_{\mathscr{V}}$ jsou zobrazení a $\triangledown, \triangleright, \triangle, \triangleleft$ 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 definičními obory a obory hodnot:
  - $\pi_{\mathscr{V}} : (X \cup \mathscr{V}^{\triangleright} \cup \mathscr{V}^{\triangleleft}) \to \mathscr{V}^{\triangleright}$,
  - $\pi_{\mathscr{C}} : (X \cup \mathscr{C}^{\triangledown} \cup \mathscr{C}^{\triangle}) \to \mathscr{C}^{\triangledown}$,
  - $\triangledown : (X \cup \mathscr{C}^{\triangledown} \cup \mathscr{V}^{\triangleleft} \cup \mathscr{V}^{\triangleright}) \to (X \cup \mathscr{C}^{\triangle})$,
  - $\triangleright : (X \cup \mathscr{V}^{\triangleright} \cup \mathscr{C}^{\triangledown} \cup \mathscr{C}^{\triangle}) \to (X \cup \mathscr{V}^{\triangleleft})$,
  - $\triangle : (X \cup \mathscr{C}^{\triangle} \cup \mathscr{V}^{\triangleright} \cup \mathscr{V}^{\triangleleft}) \to (X \cup \mathscr{C}^{\triangledown})$,
  - $\triangleleft : (X \cup \mathscr{V}^{\triangleleft} \cup \mathscr{C}^{\triangle} \cup \mathscr{C}^{\triangledown}) \to (X \cup \mathscr{V}^{\triangleright})$.

- Nechť $\boldsymbol{x} \in (X \cup \mathscr{C}^{\triangledown} \cup \mathscr{V}^{\triangleright} \cup \mathscr{C}^{\triangle} \cup \mathscr{V}^{\triangleleft})$. Kdykoli je definováno
  - $\boldsymbol{x}^{\triangledown}$, platí $\boldsymbol{x}^{\triangledown\triangle} \approx \boldsymbol{x}$,
  - $\boldsymbol{x}^{\triangleright}$, platí $\boldsymbol{x}^{\triangleright\triangleleft} \approx \boldsymbol{x}$,
  - $\boldsymbol{x}^{\triangle}$, platí $\boldsymbol{x}^{\triangle\triangledown} \approx \boldsymbol{x}$,
  - $\boldsymbol{x}^{\triangleleft}$, platí $\boldsymbol{x}^{\triangleleft\triangleright} \approx \boldsymbol{x}$,
  - $\boldsymbol{x}^{\triangledown}$, platí $\boldsymbol{x}^{\triangledown\triangleright\triangle\triangleleft\triangledown} = \boldsymbol{x}^{\triangledown}$.
- Zobrazení $\pi_{\mathscr{C}}$ splňuje podmínky:
  - pro všechna $a \in \mathscr{C}^{\triangledown}$ je $\pi_{\mathscr{C}}(a) = a$,
  - pro všechna $a \in \mathscr{C}^{\triangle}$ je $\pi_{\mathscr{C}}(a) = \varphi(a)$,
  - pro všechna $\boldsymbol{x} \in X \cup \mathscr{C}^{\triangle}$ je $\pi_{\mathscr{C}}(\boldsymbol{x}) = \pi_{\mathscr{C}}(\boldsymbol{x}^{\triangle})$.
- Zobrazení $\pi_{\mathscr{V}}$ splňuje podmínky:
  - pro všechna $a \in \mathscr{V}^{\triangleright}$ je $\pi_{\mathscr{V}}(a) = a$,
  - pro všechna $a \in \mathscr{V}^{\triangleleft}$ je $\pi_{\mathscr{V}}(a) = \varphi(a)$,
  - pro všechna $\boldsymbol{x} \in X \cup \mathscr{V}^{\triangleleft}$ je $\pi_{\mathscr{V}}(\boldsymbol{x}) = \pi_{\mathscr{V}}(\boldsymbol{x}^{\triangleleft})$.

Velikost řešení chápeme jako mohutnost množiny X.

Poznámka 6.5. Z hlediska našeho virtuálního chodce je 2D rovnice dána jako množina sloupů společně s tím, které dvojice stěn sloupů jsou vzájemně viditelné, což určuje zobrazení φ. Dlaždice mezi sloupy zatím mezi sloupy chybějí.

Řešení 2D rovnice je pak vydláždění prostoru mezi sloupy tak, aby se po dlaždicích mohl začít chodec procházet předepsaným způsobem. Zadané dvojice viditelných stěn musejí být spojeny konečnou posloupností dlaždic v jediném směru (severojižním nebo západovýchodním).

Definice 6.6. Je-li dána 2D rovnice $\mathcal{R}$ a její řešení σ, tak *duální rovnici* s *duálním řešením* značíme $\mathcal{R}^T$, σ^T a vyrobíme je

- výměnou množin $\mathscr{C}^{\triangledown}$ a $\mathscr{V}^{\triangleright}$,
- výměnou množin $\mathscr{V}^{\triangleleft}$ a $\mathscr{C}^{\triangle}$,
- výměnou zobrazení $\pi_{\mathscr{C}}$ a $\pi_{\mathscr{V}}$,
- výměnou zobrazení $\triangleleft$ a $\triangle$,
- výměnou zobrazení $\triangledown$ a $\triangleright$.

Poznámka 6.7. Definice duality využívá toho, že struktura řešení 2D rovnice je symetrická vzhledem k severojižnímu a západovýchodnímu směru. Tato myšlenka je použita v důkazu tvrzení 6.17.

Definice 6.8. Buď a proměnná nebo konstanta 2D rovnice $\mathcal{R}$, tj. $a \in \mathscr{C}^{\triangledown} \cup \mathscr{V}^{\triangleright}$. Uvažujme nějaké řešení této 2D rovnice

$$\sigma = (X, \pi_{\mathscr{C}}, \pi_{\mathscr{V}}, \triangledown, \triangleright, \triangle, \triangleleft).$$

Pak počet prvků $\boldsymbol{x} \in X$, pro které je $\pi_{\mathscr{C}}(\boldsymbol{x}) = a$ nebo $\pi_{\mathscr{V}}(\boldsymbol{x}) = a$ nazýváme *délkou a v řešení* σ a značíme $|\sigma|_a$.

Definice 6.9. *2D rovnicí s vynucenými nulami* rozumíme dvojici $(\mathcal{R}, \mathscr{E})$, kde

$$\mathcal{R} = (\mathscr{C}^{\triangledown}, \mathscr{V}^{\triangleright}, \mathscr{C}^{\triangle}, \mathscr{V}^{\triangleleft}, \varphi, \approx)$$

je 2D rovnice a $\mathscr{E} \subseteq \mathscr{C}^{\triangledown} \cup \mathscr{V}^{\triangleright}$. Řešením takové 2D rovnice s vynucenými nulami pak je takové řešení σ 2D rovnice $\mathcal{R}$, které splňuje $\forall a \in \mathscr{E} : |\sigma|_a = 0$. Podobně jako v případě rovnic na slovech definujeme $\min((\mathcal{R}, \mathscr{E}))$ jako nejmenší možnou velikost řešení příslušné 2D rovnice s vynucenými nulami, pokud řešení existuje.

Poznámka 6.10. Pokud bychom chtěli veličinu, která by odpovídala počtu zlomů v makro přístupu či počtu proměnných v mikro přístupu a vyjadřovala by, nakolik je zbývající 2D rovnice s vynucenými nulami rozsáhlá, je vhodné zvolit počet všech dvojic (x, A), kde $x \in \mathscr{V}^{\triangleright} \cup \mathscr{V}^{\triangleleft}$, $A \in \mathscr{C}^{\triangledown} \cup \mathscr{C}^{\triangle}$, platí $x \approx A$ a přitom ani x ani A neleží v $\mathscr{E}$.

6.3 Souvislost s balancovanými 2-soustavami

Tvrzení 6.11. Uvažujme balancovanou 2-soustavu $\mathcal{S}$ takovou, že v ní nejsou nikde po sobě dva výskyty proměnné. Pak existuje 2D rovnice

$$\mathcal{R} = ((\mathscr{C}^{\triangledown}, \mathscr{V}^{\triangleright}, \mathscr{C}^{\triangle}, \mathscr{V}^{\triangleleft}, \varphi, \approx), \mathscr{E})$$

a bijekce ξ z řešení soustavy $\mathcal{S}$ do řešení 2D rovnice $\mathcal{R}$ s vlastnostmi

- $\mathscr{C}^{\triangledown} \setminus \mathscr{E} = \mathscr{V}_{\mathcal{S}}$,
- $\mathscr{V}^{\triangleright} \setminus \mathscr{E} = \mathscr{C}_{\mathcal{S}}$,
- $|\mathscr{E} \cap \mathscr{C}^{\triangledown}| = \langle\langle \mathcal{S} \rangle\rangle$,
- $|\mathscr{E} \cap \mathscr{V}^{\triangleright}| = |\mathscr{C}_{\mathcal{S}}| - |\mathscr{V}_{\mathcal{S}}| + \langle\langle \mathcal{S} \rangle\rangle$,
- pro každé řešení σ soustavy $\mathcal{S}$ platí
  - $\forall x \in \mathscr{V}_{\mathcal{S}} : |\sigma(x)| = |\xi(\sigma)|_x$,
  - $\forall A \in \mathscr{C}_{\mathcal{S}} : |\sigma(\mathcal{S})|_A = 2(|\xi(\sigma)|_A + 1)$.

Důkaz: Založíme za každou rovnici novou konstantu a tuto konstantu přidáme na každý konec a začátek každé strany této rovnice (tedy takové konstanty se budou v soustavě vyskytovat čtyřikrát). Dále do soustavy $\mathcal{S}$ přidáme proměnnou mezi každé dva výskyty konstanty, které jsou těsně vedle sebe. Vložení nových proměnných můžeme provést jakkoli, abychom každou novou proměnnou přidali do soustavy dvakrát, na různé strany. Nově vzniklou soustavu nazveme $\mathcal{S}_2$. Množiny konstant resp. proměnných 2D rovnice $\mathcal{R}$ budou množinami konstant resp. proměnných soustavy $\mathcal{S}_2$, přitom množina $\mathscr{E}$ se bude skládat z nových proměnných a soustav.

Množiny $\mathscr{V}^{\triangleleft}, \mathscr{C}^{\triangle}$ budou disjunktní kopie množin $\mathscr{V}^{\triangleright}, \mathscr{C}^{\triangledown}$, spojené bijekcí φ. Nakonec relaci $\approx$ definujeme takto

- Pro $A \in \mathscr{C}^{\triangledown}, x \in \mathscr{V}^{\triangleright}$ definujeme $A \approx x$, právě když na levé straně nějaké rovnice soustavy $\mathcal{S}_2$ je výskyt konstanty A těsně před výskytem proměnné x.
- Pro $\varphi(x) \in \mathscr{V}^{\triangleleft}, A \in \mathscr{C}^{\triangledown}$ definujeme $x \approx A$, právě když na levé straně nějaké rovnice soustavy $\mathcal{S}_2$ je výskyt proměnné x těsně před výskytem konstanty A.
- Pro $\varphi(A) \in \mathscr{C}^{\triangle}, x \in \mathscr{V}^{\triangleright}$ definujeme $A \approx x$, právě když na pravé straně nějaké rovnice soustavy $\mathcal{S}_2$ je výskyt konstanty A těsně před výskytem proměnné x.
- Pro $\varphi(x) \in \mathscr{V}^{\triangleleft}, \varphi(A) \in \mathscr{C}^{\triangle}$ definujeme $x \approx A$, právě když na pravé straně nějaké rovnice soustavy $\mathcal{S}_2$ je výskyt proměnné x těsně před výskytem konstanty A.

Bijekci ξ definujeme následovně. Řešení σ soustavy S interpretujeme jako řešení σ_2 soustavy $\mathcal{S}_2$, ve kterém budou nové proměnné prázdné. Sestrojíme řešení $\xi(\sigma) = (X, \pi_{\mathscr{C}}, \pi_{\mathscr{V}}, \triangledown, \triangleright, \triangle, \triangleleft)$ takto:

- množina X se skládá z těch pozic $\boldsymbol{p} \in \mathscr{P}^{\mathcal{S}_2,\sigma_2}$, pro které je $\Phi Z^{-1}_{\sigma_2}(\boldsymbol{p}) \in \mathscr{V}_{\mathcal{S}_2}$.
- pro $\boldsymbol{x} \in X$ definujeme
  - $\pi_{\mathscr{C}}(\boldsymbol{x}) = \tilde{\sigma}\Psi_{\sigma}(\boldsymbol{x})$,
  - $\pi_{\mathscr{V}}(\boldsymbol{x}) = \Phi Z^{-1}_{\sigma}(\boldsymbol{x})$,
  - $\boldsymbol{x}^{\triangleright} = \boldsymbol{x} + 1$, je-li $\Phi Z^{-1}_{\sigma_2}(\boldsymbol{x}+1) \in \mathscr{V}_{\mathcal{S}_2}$, v opačném případě je $\boldsymbol{x}^{\triangleright} = \varphi(\pi_{\mathscr{V}}(\boldsymbol{x}))$,
  - $\boldsymbol{x}^{\triangleleft} = \boldsymbol{x} - 1$, je-li $\boldsymbol{x}^{\triangleright} = \pi_{\mathscr{V}}(\boldsymbol{x})$, v opačném případě je $\boldsymbol{x}^{\triangleleft} \in \mathscr{V}^{\triangleright}$,
  - $\boldsymbol{x}^{\triangledown} = \boldsymbol{x}\beta\gamma$, je-li $\Phi Z^{-1}_{\sigma_2}(\boldsymbol{x}\beta\gamma) \in \mathscr{V}_{\mathcal{S}_2}$, v opačném případě je $\boldsymbol{x}^{\triangledown} = \varphi(\pi_{\mathscr{C}}(\boldsymbol{x}))$,
  - $\boldsymbol{x}^{\triangle} = \boldsymbol{x}\gamma\beta$, je-li $\Phi Z^{-1}_{\sigma_2}(\boldsymbol{x}\gamma\beta) \in \mathscr{V}_{\mathcal{S}_2}$, v opačném případě je $\boldsymbol{x}^{\triangle} = \pi_{\mathscr{V}}(\boldsymbol{x})$,

Hodnoty zobrazení $\pi_{\mathscr{C}}, \pi_{\mathscr{V}}, \triangledown, \triangleright, \triangle, \triangleleft$ na množině $\mathscr{C}^{\triangledown} \cup \mathscr{C}^{\triangle} \cup \mathscr{V}^{\triangleright} \cup \mathscr{V}^{\triangleleft}$ jsou tímto již jednoznačně určeny, aby se jednalo o řešení 2D rovnice. Pečlivý čtenář snadno ověří, že takto vznikne řešení 2D rovnice $\mathcal{R}$, že definované zobrazení ξ splňuje podmínky, které jsou na něj kladeny a že se jedná o bijekci. ∎

Poznámka 6.12. Z obecné řešitelné balancované 2-soustavy můžeme vyrobit 2-soustavu splňující podmínky pro převedení na 2D soustavu tak, že na základě nejmenšího řešení vepíšeme konstantu na konec každé proměnné.

Balancované 2-soustavy tedy umíme převádět na 2D rovnice s vynucenými nulami. Nabízí se otázka, nakolik jsou 2D rovnice s vynucenými nulami obecnějším pojmem, tedy jestli je možné, aby 2D rovnice s vynucenými nulami měla mnohem větší minimální řešení než balancované kvadratické rovnice. Ve skutečnosti to možné není, tedy 2D rovnice a balancované rovnice jsou „skoro ekvivalentní", jak ukazuje následující tvrzení.

Tvrzení 6.13. Buď $\mathcal{R} = ((\mathscr{C}^{\triangledown}, \mathscr{V}^{\triangleright}, \mathscr{C}^{\triangle}, \mathscr{V}^{\triangleleft}, \varphi, \approx), \mathscr{E})$ řešitelná 2D rovnice s vynucenými nulami a $\tau = (X, \pi_{\mathscr{C}}, \pi_{\mathscr{V}}, \triangledown, \triangleright, \triangle, \triangleleft)$ její řešení, které splňuje

$$\forall \boldsymbol{x} \in X : \exists i \in \mathbb{N} : x(\triangleright)^i \in \mathscr{V}^{\triangleleft}.$$

Pak existuje řešitelná balancovaná 2-soustava $\mathcal{S}$ a zobrazení ξ z množiny všech řešení soustavy $\mathcal{S}$ do množiny všech řešení 2D rovnice $\mathcal{R}$ splňující

- $\mathscr{C}^{\triangledown} \setminus \mathscr{E} = \mathscr{C}_{\mathcal{S}}$,
- $\mathscr{V}^{\triangleright} \setminus \mathscr{E} \subseteq \mathscr{V}_{\mathcal{S}}$,
- $|\mathscr{V}_{\mathcal{S}}| - |\mathscr{V}^{\triangleright} \setminus \mathscr{E}| \leq |\mathscr{C}^{\triangledown}|$,
- existuje řešení σ soustavy $\mathcal{S}$, pro které je $\xi\sigma = \tau$,
- pro každé řešení σ soustavy $\mathcal{S}$ platí
  - pro každou proměnnou $x \in \mathscr{V}^{\triangleright} \setminus \mathscr{E}$ je $|\sigma(x)| = |\xi(\sigma)|_x$,
  - pro každou konstantu $A \in \mathscr{C}^{\triangledown} \setminus \mathscr{E}$ je $|\sigma(\mathcal{S})|_A \leq 2(|\xi(\sigma)|_A + 1)$.

Důkaz: Na základě 2D rovnice $\mathcal{R}$ budeme budovat slova nad abecedou $(\mathscr{V}^{\triangleright} \cup \mathscr{C}^{\triangledown})^*$. Vždy začneme prvkem $A_1 \in \mathscr{C}^{\triangledown}$, následujeme jediným prvkem $x_1 \in \mathscr{V}^{\triangleright}$, pro který $A_1 \approx x_1$. Další znak, který do slova napíšeme, je jediný $A_2 \in \mathscr{C}^{\triangledown}$ splňující $\varphi(x_1) \approx A_2$, pokračujeme $x_2 \in \mathscr{V}^{\triangleright}$ splňující $A_2 \approx x_2$, atd. Skončíme v okamžiku, kdy bychom do slova měli opět napsat symbol A_1. Množinu všech slov, které takto mohou vzniknout, označme L_0 a dále označme jako L množinu L_0 vyfaktorizovanou podle cyklické záměny. Každý symbol množiny $\mathscr{V}^{\triangleright} \cup \mathscr{C}^{\triangledown}$ se vyskytuje v právě jednom prvku L.

Stejným způsobem sestrojíme množinu R_0 a R, pouze s tím rozdílem, že místo $A_i \approx x_i$ resp. $x_i \approx A_{i+1}$ budeme vyžadovat $\varphi(A_i) \approx x_i$ resp. $x_i \approx \varphi(A_{i+1})$.

Množina L reprezentuje levé strany v soustavě, množina R pravé. Je však třeba je spárovat a k tomu využijeme řešení τ. Prvek $l \in L$ spárujeme s prvkem $r \in R$, právě když existuje $\boldsymbol{x} \in l$, a $i \in \mathbb{N}_0$ takové, že v řešení τ platí $\boldsymbol{x}^{\triangledown}(\triangleright)^i \in r$. Takto zavedené spárování je bijekcí mezi prvky R a prvky L.

Nyní z každého tohoto páru (l, r) sestrojíme rovnici, nejprve zvolíme jejich reprezentanty $l_0 \in L_0, r_0 \in R_0$. Pokud se ve slovech l vyskytuje nějaký symbol A z množiny $\mathscr{E} \cap \mathscr{C}^{\triangledown}$, volíme l_0 tak, aby A byl první, stejně tak volíme r_0 tak, aby $\varphi(A)$ byl první a tak sestavíme rovnici $l_0 = r_0$. V opačném případě volíme reprezentanty jakkoli, založíme novou proměnnou x a sestavíme rovnici $xl_0 = r_0x$. V každé rovnici je na levé straně nějaká konstanta, proto počet nových proměnných nepřevyšuje původní počet konstant. Vzniklou balancovanou 2-soustavu označme $\mathcal{S}_0$.

Řešení σ_0 soustavy $\mathcal{S}_0$ volíme tak, že pro $x \in \mathscr{V}^{\triangleright}$ definujeme

$$\sigma_0(x) = \pi_{\mathscr{C}}(x\triangleright)\pi_{\mathscr{C}}(x(\triangleright)^2) \dots \pi_{\mathscr{C}}(x(\triangleright)^{|\tau|_x - 1}).$$

Tím bude platit $\sigma_0(l_0) = \sigma_0(r_0)$ v rovnicích s vynucenou nulovou konstantou. V ostatních rovnicích bude tato rovnost platit až na cyklickou záměnu, což ošetříme vhodnou volbou $\sigma_0(x)$ pro každou novou proměnnou x.

Zbývá se vypořádat s ostatními prvky $\mathscr{E}$. Pro každý prvek $x \in \mathscr{E} \cap \mathscr{V}^{\triangleright}$ odstraníme oba výskyty proměnné x z rovnice. Dále pro každý prvek $A \in \mathscr{E} \cap \mathscr{C}^{\triangledown}$ rovnici ve výskytech

této konstanty rozdělíme na dvě a samotnou konstantu smažeme. Vzniklou řešitelnou soustavu konečně prohlásíme za $\mathcal{S}$.

Zobrazení ξ sestrojíme tak, že pro dané řešení σ 2-soustavy $\mathcal{S}$ prohlásíme za množinu X řešení $\xi(\sigma)$ množinu všech levých pozic $\boldsymbol{p}$ soustavy $\mathcal{S}$ v řešení σ, pro které $Z_\sigma^{-1}(\boldsymbol{p}) \in \mathcal{V}^{\triangleright} \setminus \mathcal{E}$. Zbylé parametry řešení $\xi(\sigma)$ sestrojíme stejně jako v případě důkazu předchozího tvrzení, jen s tím rozdílem, že pokud při zavádění zobrazení $\triangledown$ resp. $\triangle$ dopadne $\Phi Z_\sigma^{-1}(\boldsymbol{x}\beta\gamma)$ resp. $\Phi Z_\sigma^{-1}(\boldsymbol{x}\gamma\beta)$ do nově zavedené proměnné, použijeme první $\Phi Z_\sigma^{-1}\left(\boldsymbol{x}(\beta\gamma)^i\right)$ resp. $\Phi Z_\sigma^{-1}\left(\boldsymbol{x}(\gamma\beta)^i\right)$, které nově zavedenou proměnnou není. ∎

Pozorování 6.14. Požadavek na řešení τ z předchozího tvrzení splňuje každé nejkratší řešení. V opačném případě totiž je možné z množiny X odebrat množinu $M = \{\boldsymbol{x}, \boldsymbol{x}\triangleright, \boldsymbol{x}(\triangleright)^2, \ldots, \boldsymbol{x}\triangleleft\}$ a prvky množiny X jižně od množiny M spojit s těmi severně od množiny M. Přesněji, pokud $\boldsymbol{y} \in X \setminus M$ a $\boldsymbol{y}\triangle \in M$, tak už nutně $\boldsymbol{y}\triangle \notin M$ a v novém kratším řešení tak tento prvek množiny X bude reprezentovat $\boldsymbol{y}\triangle$.

Důsledek 6.15. Polynomiální resp. jednoduše exponenciální omezení velikosti nejkratšího řešení 2D soustavy je ekvivalentní polynomiálnímu resp. jednoduše exponenciálnímu omezení nejkratšího řešení balancovaných kvadratických soustav.

Poznámka 6.16. Jak vyplývá z předchozích tvrzení a jejich důkazů, 2D rovnice a balancované 2-soustavy jsou (až na drobné technikálie) v podstatě totožné. Pokud se budeme dívat na pojmy 2-soustav z pohledu 2D rovnic, opravdu budou hesla z tabulky 6.1 na sebe přecházet při dualizaci 2D rovnice.

Jako ukázku využití 2D rovnic k duálním tvrzením dokážeme následující tvrzení.

Tvrzení 6.17. Existuje polynom p takový, že pro libovolnou balancovanou 2-soustavu $\mathcal{S}$ má každé c-minimální řešení délku menší než $2^{2^{p(|\mathcal{S}|)}}$.

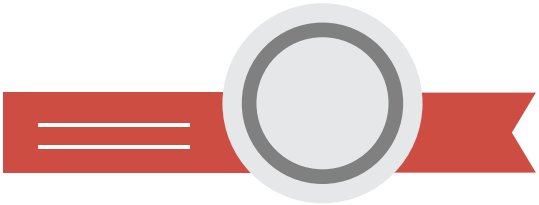
Důkaz: Zavedeme l-uspořádání a c-uspořádání i na 2D rovnicích: Pro dvě řešení τ_1, τ_2 píšeme $\tau_1 \leq_l \tau_2$, pokud pro každé $x \in \mathcal{V}^{\triangleright}$ platí $|\tau_1|_x \leq |\tau_2|_x$ a podobně píšeme $\tau_1 \leq_c \tau_2$, pokud pro každé $A \in \mathcal{C}^{\triangledown}$ platí $|\tau_1|_A \leq |\tau_2|_A$. Ve stejném smyslu budeme hovořit i o l-minimálních a c-minimálních řešeních 2D rovnic.

Uvažujme nyní c-minimální řešení σ 2-soustavy $\mathcal{S}$. Do této soustavy vepíšeme konstantu na konec každé neprázdné proměnné a rozrůzníme konstanty. Tím získáme soustavu $\mathcal{S}_2$ a její řešení σ_2, které je stále c-minimální a stejně velké jako σ. Navíc soustava $\mathcal{S}$ splňuje podmínku pro tvrzení 6.11, takže můžeme vyrobit 2D rovnici $\mathcal{R}$. Bijekcí ξ z tvrzení 6.11 převedeme řešení σ_2 na c-minimální řešení τ 2D rovnice $\mathcal{R}$.

Dualizací dostáváme l-minimální řešení τ^T soustavy $\mathcal{R}^T$ stejně velké jako τ.

Ze stejného argumentu jako v předchozím pozorování i každé l-minimální řešení splňuje požadavky pro tvrzení 6.13. Najdeme tedy na základě tohoto tvrzení soustavu $\mathcal{S}_3$ a řešení σ_3, které přejde na řešení τ^T 2D rovnice $\mathcal{R}^T$. Řešení σ_3 je l-minimální a proto můžeme jeho délku dvojitě exponenciálně omezit pomocí délky soustavy $\mathcal{S}_3$ podle tvrzení 3.33. Platí $\sigma_3(\mathcal{S}_3) \geq \sigma(\mathcal{S})$ a $|\mathcal{S}_3| \leq 8|\mathcal{S}|$, takže platí dvojitě exponenciální mez i pro původní řešení σ soustavy $\mathcal{S}$. ∎

Poznámka 6.18. Požadavek na balancovanost soustavy je ve 2D rovnicích proto, 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 dualizaci převést na obyčejnou kvadratickou rovnici na slovech, ale mohli bychom ji převést na takovou, kde se místo některých proměnných vyskytuje jejich „zrcadlový obraz“. Tedy tvrzení o dvojitě exponenciální mezi na c-minimální řešení by bylo možné s jistým množstvím péče dokázat i pro obecné rovnice.



Závěr

Práce ukazuje, že pokud by chtěl čtenář dokázat, že řešitelnost kvadratických rovnic je NP úplný problém, stačí mu dokázat jednoduše exponenciální mez pro nejkratší možnou délku nejkratší proměnné ve 2-soustavách.

Vedle toho jsou zde další cesty, kudy se může další výzkum ubírat. Za zmínku stojí hypotéza 5.11, jejíž dokázání by znamenalo popření polynomiální meze na velikost nejkratšího řešení. Dále tu máme poznámku 4.4, tázající se, zda nejen řešitelnost kvadratických soustav, ale i samotných 2-soustav (s rozrůzněnými konstantami) je NP těžkým problémem.

A nakonec i závěrečná poetická kapitola si zaslouží povšimnutí. Krom dodání puncu matematické exaktnosti je možné například zobecnit tvrzení 2.66 a 3.34 na dvoudimenzionální vzdálenosti, formulovat definici „neorientovaných“ 2D rovnic vzniklých z nebalancovaných 2-soustav, či zkoumat, jak vypadají množiny, které je možné z řešení 2D rovnice odstranit.

Práce tedy nabízí více cest, kudy pokračovat ve výzkumu, je jen na čtenáři, kterou se vydá, rozhodne-li se věnovat svůj um a čas kvadratickým rovnicím na slovech.

Literatura

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

[2] John Michael Robson, Volker Diekert, *Quadratic Word Equations*, Jewels are Forever 1999: 314–326

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

Příloha A

Značení

A.1 Symboly

Následuje seznam použitých symbolů spolu s odkazem na čísla definic, ve kterých jsou zavedeny.

$\mathscr{A}^*$ … množina slov nad abecedou $\mathscr{A}$, viz 1.1.
$|s|$ … délka slova s, viz 1.1.
$s[i]$ … $(i+1)$-ní písmeno slova s, viz 1.1.
$\mathrm{Pref}_k(a)$ … prefix délky k, viz 1.3.
$\mathrm{Suff}_k(a)$ … suffix délky k, viz 1.3.
$R_r = (S_{r,0}, S_{r,1})$ … r-tá rovnice soustavy, viz 1.4.
$\mathscr{V}_\mathcal{S}$ … množina proměnných soustavy $\mathcal{S}$, viz 1.5.
$\mathscr{C}_\mathcal{S}$ … množina konstant soustavy $\mathcal{S}$, viz 1.5.
$\langle\langle\mathcal{S}\rangle\rangle$ … počet rovnic soustavy $\mathcal{S}$, viz 1.5.
$|R|$ … délka rovnice R, viz 1.6.
$|\mathcal{S}|$ … délka soustavy $\mathcal{S}$, viz 1.6.
$\sigma(\mathcal{S})$ … dosazení řešení σ do soustavy $\mathcal{S}$, viz 1.9.
$\min(\mathcal{S})$ … délka nejkratšího řešení soustavy $\mathcal{S}$, viz 1.10.
$\sigma_1 <_l \sigma_2$ … řešení σ_1 je l-menší než σ_2, viz 1.12.
$|s|_a$ … počet písmen a ve slově s, viz 1.14.
$\sigma_1 <_c \sigma_2$ … řešení σ_1 je c-menší než σ_2, viz 1.16.
t_σ … typový homomorfismus řešení σ, viz 1.24.
$\tilde{\sigma}$ … zobrazení, které přiřazuje pozicím typového homomorfismu písmena, viz 1.24.
$\mathscr{Q}^\mathcal{S}$ … množina výskytů v soustavě $\mathcal{S}$, viz 2.1.
$\mathscr{Q}^\mathcal{S}_{r,b,i}$ … i-tý výskyt b-té strany r-té rovnice, viz 2.1.
Φ … zobrazení z výskytů do proměnných a konstant, viz 2.1.
$\mathscr{P}^{\mathcal{S},t}$ … množina všech pozic v typu t, viz 2.2.
$\mathscr{P}^{\mathcal{S},t}_{r,b,i}$ … i-tá pozice b-té strany r-té rovnice v typu t, viz 2.2.
Ψ_t … zobrazení z pozic do pozic typového homomorfismu t, viz 2.2.
Z_t je zobrazení z výskytů do slov z pozic, viz 2.3.
Z_t^{-1} je zobrazení z pozic zpátky do výskytů, viz 2.4.
$\boldsymbol{v} + k$ … výskyt / pozice o k napravo od $\boldsymbol{v}$, viz 2.6.
β … skok na druhou stranu rovnice, viz 2.7.
γ … skok na druhý výskyt proměnné, viz 2.7.
$(\gamma\beta)^{\mathrm{MAX}}$ … maximální možná iterace skoků $\gamma\beta$, viz 2.13.
Pal_n … jistá soustava délky $4n$, viz 5.1.
PalSol_n … jisté řešení soustavy Pal_n, viz 5.1.

A.2 Rejstřík pojmů

Rejstřík odkazuje na čísla definic, v závorce následují čísla stran.