Neohraničená slova a (lokální) periody

Řekneme, že slovo w je *ohraničené*, pokud existuje neprázdné slovo $u \neq w$, které je současně prefixem a sufixem w. Takové slovo u nazveme *hranicí* slova w. Je-li slovo u samo ohraničené a má hranici z, je z také hranice w. Z toho snadno odvodíme, že nejkratší hranice ohraničeného slova není ohraničená. Je-li hranice u slova w delší než $|w|/2$, pak se u překrývá samo se sebou a je tedy ohraničené. Z toho vidíme, že každé ohraničené slovo je tvaru uvu, $u \neq \varepsilon$. Dále zřejmě platí, že $|w| - |u|$ je perioda slova w. Označíme-li nejkratší periodu slova w symbolem $\pi(w)$, dostáváme, že pro ohraničené slovo w platí $\pi(w) = |w| - |u|$, kde u je jeho nejdelší hranice. Slovo je tedy *neohraničené* (tedy není ohraničené), právě když $\pi(w)$ je maximální možné, totiž $|w|$.

Lyndonova slova. Primitivní slovo je takové, které není celočíselnou mocninou žádného kratšího slova. Každé ohraničené slovo je však racionální mocninou nějakého kratšího slova, např. $aabaaba = (aab)^{\frac{7}{3}}$. Ukážeme ale, že každé primitivní slovo je konjugované s nějakým neohraničeným slovem. Takových slov může být více, ale jedním z nich je vždy tzv. *Lyndonovo slovo*, které definujeme jako primitivní slovo, které je minimální ve své konjugační třídě (tj. ekvivalenční třídě vzájemně konjugovaných slov) vzhledem k nějakému lexikografickému uspořádání $\lhd$.

Připomeňme, že je-li slovo primitivní, jsou primitivní i všechna slova s ním konjugovaná. Z primitivity plyne, že Lyndonovo slovo odpovídá jediné konjugaci, stejný výsledek dvou různých konjugací by znamenal, že slovo není primitivní (je-li uv primitivní, pak $uv \neq vu$).

Poznamenejme několik vlastností lexikografických uspořádání. Každé takové uspořádání je určeno nějakým lineárním uspořádáním písmen. Jsou-li $\lhd$ a $\blacktriangleleft$ lexikografická uspořádání daná opačnými uspořádáními písmen (tj. $a \lhd b \Leftrightarrow b \blacktriangleleft a$), pak pro dvě prefixově nesrovnatelná slova u a v platí rovněž $u \lhd v \Leftrightarrow v \blacktriangleleft u$. Pokud je ovšem u prefix v, chovají se uspořádání stejně: $u \lhd v$ i $u \blacktriangleleft v$.

Dále platí, že pro každé z je $u \lhd v \Leftrightarrow zu \lhd zv$. Pokud jsou u a v prefixově nesrovnatelná (speciálně pokud jsou stejně dlouháa různá), máme také $u \lhd v \Leftrightarrow uz \lhd vz'$ pro každé z a z'.

Věta. Každé Lyndonovo slovo je neohraničené.

Důkaz. Sporem. Nechť je Lyndonovo slovo w tvaru uvu. Z minimality plyne $uvu \lhd vuu$, tedy $uv \lhd vu$, a proto také $uuv \lhd uvu$, což je spor s minimalitou w. □

Lyndonova slova lze charakterizovat jako „samominimální“:

Věta. Slovo w je Lyndonovo, právě když je nejmenší ze svých sufixů (tj. je vzhledem k $\lhd$ striktně menší než jakýkoli vlastní sufix).

Důkaz. Nechť je w Lyndonovo slovo a nechť p je prefix w a s je sufix w tak, že $|p| = |s|$. Protože je w Lyndonovo (a tedy neohraničené), platí $p \neq s$ a $p \lhd s$, a tedy také $w \lhd s$.

Nechť naopak w Lyndonovo není. Pokud w není primitivní, je lexikograficky větší než vlastní primitivní kořen. Pokud je $w = uv$ primitivní a vu je Lyndonovo, platí $vu \lhd uv$, a tedy také $v \lhd uv$. □

Věta. Nechť u i v jsou Lyndonova slova (vzhledem k $\lhd$). Pokud $u \lhd v$ a $u \neq v$, pak je i uv Lyndonovo slovo.

Důkaz. Nejprve ukažme, že $uv \lhd v$. Pokud je $v = uv'$, pak dokazované $uv \lhd uv' = v$ plyne z $v \lhd v'$. Pokud u není prefix v, pak $uv \lhd v$ plyne z $u \lhd v$.

Nechť je nyní $z \neq v$ vlastní sufix uv. Pokud $uv = uv'z$, pak $uv \lhd v \lhd z$. Pokud $z = z'v$, pak $u \lhd z'$, a tedy $uv \lhd z'v$.

Slovo uv je tedy menší než všechny jeho vlastní sufixy a důkaz je hotov. □

Faktorizaci $w = uv$ nazýváme *standardní* (při pevně zvoleném lexikografickém uspořádání $\lhd$), pokud v je nejdelší sufix w, který je Lyndonovým slovem (vzhledem k $\lhd$). Vzhledem k tomu, že slovo délky jedna je vždy Lyndonovo, standardní faktorizace vždy existuje.

Věta. Nechť $\lhd$ je lexikografické uspořádání. Každé slovo w lze jednoznačně napsat jako součin nerostoucí posloupnosti Lyndonových slov (vzhledem k $\lhd$).

Důkaz. Uvažujme faktorizaci $w = u_1u_2 \cdots u_k$ definovanou tak, že $(u_1u_2 \cdots u_{j-1}) \cdot u_j$ je standardní faktorizace slova $u_1u_2 \cdots u_j$ pro každé $j = 2, 3, \ldots, k$ a u_1 je Lyndonovo slovo. Jinak řečeno, ze slova w postupně odebíráme jeho nejdelší Lyndonovy sufixy.

Je zřejmé, že faktorizace je dobře definovaná a všechna u_i jsou Lyndonova. Pokud $u_i \lhd u_{i+1}$ a $u_i \neq u_{i+1}$, pak z předchozí věty dostáváme spor s maximalitou u_{i+1}. Posloupnost je tedy nerostoucí.

Nechť nyní $w = v_1v_2 \ldots v_m = z_1z_2 \cdots z_\ell$ jsou dvě různé faktorizace splňující předpoklady věty. Bez újmy na obecnosti předpokládejme, že v_1 je vlastní prefix z_1 a $z_1 = v_1v_2 \cdots v_jr$, kde r je prefix v_{j+1}. Protože Lyndonovo slovo je striktně menší než jakýkoli jeho vlastní sufix, dostáváme

$$z_1 \lhd r \lhd v_{j+1} \lhd v_1 \lhd z_1,$$

a tedy $z_1 = v_1$, spor. □

Lokální perioda a kritická faktorizace.

Definice. Lokální periodou slova w na pozici k $\in \{0, 1, \ldots, |w|\}$ rozumíme délku nejkratšího slova x, které je sufixově srovnatelné s u a prefixově srovnatelné s v, kde $w = uv$ a $k = |u|$.

Jinak řečeno, lokální perioda je délka x u nejkratšího čtverce x^2, který je centrovaný na pozici k. Všimněme si, že takové x je vždy neohraničené: pokud by totiž u byla hranice x, byl by u kratší čtverec, rovněž centrovaný na pozici k.

Je snadné si všimnout, že lokální perioda je na každé pozici rovna nejvýše $\pi(w)$. Je-li lokální perioda na pozici $|u|$ rovna $\pi(w)$, nazýváme ji *kritickou pozicí* a faktorizaci $w = uv$, či přesněji dvojici (u, v), *kritickou faktorizací* slova w.

Věta (Věta o kritické faktorizaci – Critical Factorization Theorem). Každé slovo w připouští kritickou faktorizaci.

Důkaz. Zvolme dvě opačná lexikografická uspořádání $\lhd$ a $\blacktriangleleft$. Nechť α je maximální sufix w vzhledem k $\lhd$ a β maximální sufix w vzhledem k $\blacktriangleleft$. Pokud je w mocninou jednoho písmene, je tvrzení zřejmé, v opačném případě je jistě $\alpha \neq \beta$. Bez újmy na obecnosti předpokládejme $|\alpha| < |\beta|$.

Ukážeme, že $k = |w| - |\alpha|$ je kritický bod. Nechť je x^2 nejmenší čtverec centrovaný na pozici k. Předpokládejme pro spor, že $|x|$ není perioda slova w.

1. Je-li x prefix α a sufix $w\alpha^{-1}$, je $x\alpha$ sufixem w, a tedy $x\alpha \lhd \alpha$. Odtud dostáváme $\alpha \lhd x^{-1}\alpha$, což je spor s maximalitou α.

2. Je-li α prefix x a x je sufix $w\alpha^{-1}$, je α prefixem $x\alpha$, a tedy $\alpha \lhd x\alpha$, což je opět spor.

3. Nechť je konečně x prefixem α a $w\alpha^{-1}$ sufixem x. Budiž $x^i x'$ nejdelší prefix α, který má periodu $|x|$. Pak existují písmena $c \neq d$ taková, že $x'c$ a $x^i x'd$ jsou prefixy α. Z maximality α plyne $d \lhd c$. Označme $v = \beta\alpha^{-1}$. Protože v je sufixem x, je $vx'd$ faktorem slova w. To je ovšem spor s maximalitou β, protože $vx'c$ je prefixem β a $vx'c \blacktriangleleft vx'd$. □

Nejdelší neohraničené faktory. Označme $\tau(w)$ délku nejdelšího neohraničeného faktoru slova w. Je-li v faktor slova w delší než $\pi(w)$, pak je jistě ohraničený, protože má rovněž periodu $\pi(w)$. Platí tedy $\tau(w) \leq \pi(w)$. Je-li $|w| \geq 2\pi(w)$, pak nutně platí rovnost, protože slovo w obsahuje nějaké Lyndonovo slovo délky $\pi(w)$; obsahuje totiž všechny prvky konjugační třídy svého periodického základu.

Neohraničené faktory délky $\pi(w)$ můžeme také hledat pomocí kritických pozic. Je-li totiž x^2 minimální čtverec centrovaný na nějaké kritické pozici slova w a alespoň jedno x leží celé ve w, je $\tau(w) = \pi(w)$, protože x je neohraničené, a z definice kritické pozice máme $|x| = \pi(w)$. To umožňuje výše uvedený odhad mírně zlepšit.

Věta. Je-li $|w| \geq 2\pi(w) - 3$, pak $\tau(w) = \pi(w)$.

Důkaz. Nechť $w = vc_1c_2c_3v$, kde $|vc_1c_2c_3| = \pi(w)$ (případy, kdy $\pi(w) < 3$ jsou snadné). Existuje-li kritická pozice $k \leq \pi(w) - 3$, pak x, pro které je x^2 centrované v k, leží celé ve slově w a jsme hotovi.

Uvažujme kritickou pozici ℓ popsanou v důkazu Věty o kritické faktorizaci. Předpokládejme nejprve, že je rovna $\pi(w) - 1$. Pak je c_3 maximální sufix $vc_1c_2c_3$ vzhledem k nějakému lexikografickému uspořádání. Písmeno c_3 se tedy nevyskytuje ve slově vc_1c_2, a $vc_1c_2c_3$ je neohraničené a délky $\pi(w)$.

Obsahuje-li slovo w tři různá písmena, můžeme zvolit $\lhd$ tak, že c_2 není ani nejmenší, ani největší písmeno, a tudíž příslušný kritický bod leží mimo $\pi(w) - 2$ a jsme hotovi.

Zbývá situace, kdy $\ell = \pi(w) - 2$ a w obsahuje právě dvě písmena. Je-li $c_2 \neq c_3$, zvolíme uspořádání $\lhd$, ve kterém je c_3 minimální a c_2 maximální. Opět není těžké ověřit, že c_2 se nevyskytuje ve v. Pak je ale $vc_1 = c_3v$ a w má v rozporu s předpokladem periodu $|vc_1c_2|$. Nechť tedy $c_2 = c_3 = b$ a $c_1 = a$. Protože bb je lexikograficky maximální sufix $vabb$, nemá bb žádný výskyt ve v. Pokud je $vabb$ neohraničené, jsme hotovi. V opačném případě je ba prefix v a není těžké ověřit, že $b^{-1}vabbb$ je neohraničený faktor w. □

Mez uvedená ve větě je optimální, jak ukazují slova $w = (aba)^k abba \cdot (aba)^k$, která jsou délky $2\pi(w)-4$ a nejdelší neohraničené faktory jsou $(aba)^k abb$ a $bba(aba)^k$ délky $\pi(w) - 1$.

Souvislost rovnosti $\pi(w) = \tau(w)$ s délkou w vyjádřenou pomocí $\pi(w)$ je tedy přesně popsána. Podobně se můžeme ptát na souvislost s délkou w vyjádřenou pomocí $\tau(w)$. Jinak řečeno, je-li v neohraničené slovo, jak moc ho můžeme prodloužit do slova w, aby $|v| = \tau(w) < \pi(w)$. V tomto případě platí, že $|w| \geq \frac{7}{3}\tau(w)$ implikuje $\pi(w) = \tau(w)$. Tato mez je asymptoticky optimální, jak ukazuje následující slovo:

$$w = a^n ba^{n+1} ba^n ba^{n+2} ba^n ba^{n+1} ba^n$$

kde $n \geq 0$, $\tau(w) = 3n + 6$ (dvě nejdelší neohraničená slova jsou $ba^{n+1}ba^nba^{n+2}$ a $a^{n+2}ba^nba^{n+1}b$), $\pi(w) = 4n + 7$ a $|w| = 7n + 10$. Tedy $\tau(w) < \pi(w)$ a $|w| = \frac{7}{3}\pi(w) - 4$.

Důkaz je obtížný. Dokážeme ale tvrzení týkající se speciálního případu, kdy dovolujeme neohraničené slovo prodlužovat jen jedním směrem (např. doprava). Řekněme, že vu je *Duvalovo prodloužení* neohraničeného slova v, pokud $|v| = \tau(vu) < \pi(vu)$. Jak ukazují slova $v = a^iba^jbb$ a $u = a^jba^i$, kde $1 \leq i < j$, je možné Duvalovo prodloužení o $|v| - 2$. To je optimální hranice. Ukážeme ale jen následující slabší tvrzení (případ $|v| - 1$ je jen o něco techničtější).

Věta. Nechť je v neohraničené slovo. Jeho Duvalovo prodloužení délky $|v|$ není možné.

Důkaz. Nechť $u \neq v$ je slovo délky $|v|$ takové, že $\pi(vu) > |v|$, tedy $u \neq v$. Najdeme neohraničený faktor vu delší než v.

Nechť $\lhd$ a $\blacktriangleleft$ jsou dvě opačná lexikografická uspořádání taková, že α je $\lhd$-maximální sufix v, β je $\blacktriangleleft$-maximální sufix v a $|\alpha| < |\beta|$. Z maximality α plyne, že má jen jeden výskyt ve v. Nechť má nějaký další výskyt ve vu a nechť je např $v'\alpha$ prefix vu, kde $|v'\alpha| > |v|$. Tvrdíme, že $v'\alpha$ je neohraničené. Případná nejkratší hranice r je totiž jistě kratší než v. Pokud by měla délku alespoň α, bylo by α sufixem r, a protože r je prefixem v, dostáváme výskyt α uvnitř v, což je spor. Pokud by bylo naopak $|r| < |\alpha|$, bylo by r hranicí v, což je opět spor.

Nechť má tedy α jen jeden výskyt ve vu. Nechť $u = u'u''$, kde $|u''| = |\alpha|$. Označme p nejkratší periodu slova $\alpha u'$. Slovo uv má nejkratší periodu ostře větší než p, jinak by se α v $\alpha u'$ opakovalo, což jsme vyloučili. Nechť je $\alpha u'zc$ nejkratší prefix slova αu s periodou ostře větší než p, kde c je písmeno a $z \in \Sigma^*$. Nechť $\alpha u'z = ws$, kde $|w| = p$; neboli w je periodický základ slov $\alpha u'$ a $\alpha u'z$, zatímco přidáním písmene c je perioda p porušena. Všimněme si, že $s \neq \alpha$ je prefix α, zatímco sc nikoli. Nechť je tedy sd prefix α, kde $d \neq c$ je písmeno. Pokud je $\alpha u'zc$ neohraničené, jsme hotovi. Nechť je tedy ohraničené a nechť je rc jeho nejdelší hranice. Z definice s plyne, že r je kratší než s, a je tedy hranicí s. Pak je rc prefix α a rd je faktorem α. Z maximality α dostáváme $d \lhd c$.

Uvažujme nyní slovo βwsc a předpokládejme, že tc je jeho nejdelší hranice. Je-li t sufix s, je tc prefix β a td faktor α, což je vzhledem k $tc \blacktriangleleft td$ spor s maximalitou β. Je-li naopak s sufix t, je sc (jakožto sufix tc) faktorem u, což je spor s maximalitou α, protože sd je prefix α a $sd \lhd sc$.

Slovo βwsc je tedy neohraničené a jsme hotovi. $\square$