Palindromy v nekonečných slovech modelujících kvazikrystaly

Ľubomíra Balková

ČVUT, Fakulta jaderná a fyzikálně inženýrská, Katedra matematiky

Palindromy známé neznámé

Palindromická slova a palindromické věty, tedy slova a věty, které zůstávají stejné, ať je čteme odpředu či odzadu, jsou oblíbenou hrátkou. Čeština nás může pobavit palindromickými větami jako: *Kobyla má malý bok. Jelenovi pivo nelej. Nevypusť supy ven.* Nejdelším palindromickým slovem nalezeným v Oxfordském slovníku je *tattarrattat*, což je slovo vzniklé ve fantazii Jamese Joyce a použité v románu Odysseus, které znamená "klepat na dveře". V Guinnessově knize rekordů je uvedeno slovo *detartrated*, minulý čas slovesa detartrate, chemiky vymyšelný termín, který znamená odstranit vínan (skupina organických látek). Ve slovnících je často uváděný *rotavator*, obchodní značka zemědělského stroje. *Malayalam*, indiánský jazyk, je palindromické slovo stejné délky. Zdá se, že nejdelší v běžné mluvě užívaný palindrom je finské slovo *saippuakauppias*, které označuje prodavače mýdla. V roce 1991 sestavil Gordon Dow palindromickou větu obsahující 306 slov s titulem *Dog Sees Ada.* Tento palindrom si zaslouží pozornost, protože obsahuje jen málo 'umělých' slov. Ovšem palindromy se objevují i v seriózních disciplínách.

Ve většině genomů (soubory genů obsažené v chromozómech buněčného jádra organismů) se objevují palindromické motivy. Pod palindromem rozumějí biologové ovšem trochu odlišný pojem. DNA je tvořena dvěma vlákny nukleotidů a tyto báze jsou vzájemně komplementární, tj. vždy se párují stejným způsobem (adenin s tyminem a cytosin s guaninem). Posloupnost nukleotidů ve vláknu DNA se pak nazývá palindromem, pokud je rovna své komplementární posloupnosti čtené pozpátku. Například posloupnost ACCTAGGT je palindromem, protože přečteme-li její komplement TGGATCCA pozpátku, dostaneme původní posloupnost. Motivy formované řazením nukleotidů specifikují, jaké bílkoviny mají být buňkou produkovány. Nedávné výzkumy ukázaly, že mnoho bází chromozómu Y je uspořádáno palindromicky. Toto uspořádání umožňuje totiž chromozómu Y, dojde-li k jeho poškození, jednoduchý samoopravný proces pomocí překopírování genetické informace své 'zdravé' poloviny.

V tomto článku se zaměříme na jiné odvětví, kde se palindromy intenzivně studují, a tím je teorie kvazikrystalů. Objev kvazikrystalů v roce 1982 [6] způsobil poprask. Kvazikrystaly jsou materiály s ostrými body na difrakčním obrázku, ale zároveň s nekrystalografickými symetriemi. Tyto materiály musí tedy být uspořádány na dlouhou vzdálenost, toto uspořádání ovšem není periodické. Až do doby objevu kvazikrystalů se věřilo, že pravidelnost ve struktuře látek je synomem periodicity. K modelování jednorozměrných kvazikrystalů slouží některá nekonečná neperiodická slova. Zkoumáním jejich kombinatorických a algebraických vlastností tak nepřímo studujeme fyzikální vlastnosti kvazikrystalů. *Faktorová komplexita* odpovídá počtu lokálních konfigurací atomů v materiálu. *Palindromická komplexita* popisuje lokální symetrie materiálu. My se zaměříme na zkoumání palindromů jisté skupiny nekonečných slov, která jsou vhodným modelem kvazikrystalů.

Značení

Abeceda $\mathcal{A}$ je konečná množina symbolů zvaných *písmena*. Zřetězením písmen získáme *slovo*. $\mathcal{A}^*$ je množina všech konečných slov nad abecedou $\mathcal{A}$. Pokračujeme-li v řetězení písmen bez omezení, získáme doprava nekonečné slovo $u = u_0u_1u_2....$ Analogicky získáme doleva nekonečné slovo. *Faktor* (*podslovo*) slova u je každé slovo (konečné či nekonečné), které se v u vyskytuje. Písmeno $d \in \mathcal{A}$ je *levým prodloužením* faktoru v, pokud dv je faktor u. Slovo v je *levý speciální faktor*, pokud v má více levých prodloužení. Slovo v je předpona u, pokud $u = vw$ pro nějaký faktor w slova u. *Jazyk* $L(u)$ je množina všech faktorů slova u. Symbolem $L_n(u)$ značíme množinu všech faktorů délky n slova u. Symbolem ε značíme prázdné slovo. Substituce φ je homomorfismus : $\mathcal{A}^* \to \mathcal{A}^*$ takový, že existuje $d \in \mathcal{A}$ splňující $\varphi(d) = dw$ pro nějaké neprázdné slovo $w \in \mathcal{A}^*$ a $\varphi(d) \neq \varepsilon$ pro každé $d \in \mathcal{A}$. Substituce může být definována pro doprava nekonečné slovo $u = u_0u_1u_2 \ldots$ předpisem: $\varphi(u) := \varphi(u_0)\varphi(u_1)\varphi(u_2) \ldots$ Poznamenejme, že každá substituce má pevný bod, tj. existuje nekonečné slovo u takové, že $\varphi(u) = u$. Substituce φ se nazývá primitivní, pokud existuje $k \in \mathbb{N}$ takové, že pro každou dvojici písmen $c, d \in \mathcal{A}$ se písmeno c vyskytuje ve slově $\varphi^k(d)$. Stejnoměrně rekurentní je takové nekonečné slovo u, které má následující vlastnost: pro každé $n \in \mathbb{N}$ existuje $R(n) > 0$ takové, že každé podslovo slova u délky $R(n)$ obsahuje všechny faktory slova u délky n.

Nekonečná slova jako model kvazikrystalů

Cílem této kapitoly je charakterizovat skupinu nekonečných slov, která jsou vhodným modelem kvazikrystalů. K tomutu účelu je nejprve nutno zavést několik pojmů.

Rényiho rozvoj jedničky a Parryho čísla

Nechť $\beta > 1$, potom T_β je zobrazení definované pro každé $x \in [0,1]$ předpisem $T_\beta(x) := \beta x - \lfloor \beta x \rfloor$. Nechť $\beta > 1$ a $x \in [0,1)$. Reprezentace $x = \sum_{i \geq 1} t_i \beta^{-i}$, kde $t_i \in \mathbb{N}$, se nazývá *Rényiho rozvoj* x v bázi β, pokud pro koeficienty platí: $t_i := \lfloor \beta T_\beta^{i-1}(x) \rfloor$. Každé $\beta > 1$ je charakterizováno svým Rényiho rozvojem jedničky, značeným $d_\beta(1) = t_1t_2 \ldots$, kde $t_i := \lfloor \beta T_\beta^{i-1}(1) \rfloor$. Pokud $d_\beta(1)$ je posloupnost periodická od určitého členu, β se nazývá *Parryho číslo*. Pokud $d_\beta(1)$ je konečná posloupnost, β se nazývá *jednoduché Parryho číslo*.

Nechť $d_\beta(1)$ je periodická od určitého členu, speciálně nechť $m \in \mathbb{N}$ a $p \in \mathbb{N}$ jsou minimální taková, že $d_\beta(1) = t_1...t_m(t_{m+1}...t_{m+p})^\omega$. Substituce $\varphi = \varphi_\beta$ přidružená β je definovaná na abecedě $\{0, 1, ..., m+p-1\}$ předpisem

$$\varphi(0) = 0^{t_1}1, \ldots, \ \varphi(m+p-2) = 0^{t_{m+p-1}}(m+p-1), \ \varphi(m+p-1) = 0^{t_{m+p}}m. \quad (1)$$

Nekonečné slovo přidružené β je pevný bod $u_\beta = \lim_{n\to\infty} \varphi^n(0)$ substituce φ. (Jelikož $\varphi^n(0)$ je vlastní předponou $\varphi^{n+1}(0)$, nové a nové aplikování substitučního předpisu skutečně vytvoří nekonečné slovo.)

β-rozvoj a β-celá čísla

Podívejme se nyní na slovo u_β z algebraického pohledu. Každý z nás se běžně setkává s rozvojem čísel v desítkové soustavě. Často je však výhodné rozvíjet čísla i v jiných bázích. Nechť $\beta > 1$ a $x \geq 1$ a $x/\beta^k < 1$, β-rozvojem čísla x nazýváme Rényiho rozvoj x/β^k násobený β^k. Samozřejmě pro $x \in [0, 1)$ definujeme β-rozvoj x jako jeho Rényiho rozvoj. Nezáporná β-celá čísla jsou pak definována jako: $\mathbb{Z}_\beta^+ := \{x \geq 0 \mid x = \sum_{i=0}^k x_i\beta^i \text{ je } \beta\text{-rozvoj čísla } x\}$. Je přirozené definovat pak β-celá čísla jako: $\mathbb{Z}_\beta = -\mathbb{Z}_\beta^+ \cup \mathbb{Z}_\beta^+$. β-celá čísla tedy hrají v oblasti nestandardních číselných systémů stejně významnou roli jako celá čísla při rozvoji v desítkové soustavě.

Nechť $d_\beta(1)$ je periodická od určitého členu, speciálně nechť $m \in \mathbb{N}$ a $p \in \mathbb{N}$ jsou minimální taková, že $d_\beta(1) = t_1...t_m(t_{m+1}...t_{m+p})^\omega$. Potom existuje právě $m + p$ mezer mezi po sobě jdoucími β-celými čísly. Přiřadíme-li mezerám písmena $0, 1, \ldots, m + p - 1$ podle velikosti (největší mezeře číslo 0, až nejmenší $m + p - 1$), získáme právě nekonečné slovo u_β, tedy pevný bod substituce φ.

Zpět ke kvazikrystalům

Nyní, když se na nekonečné slovo u_β můžeme dívat jako na kódování algebraické struktury $\mathbb{Z}_\beta$, lze hledat souvislost s kvazikrystalickou strukturou. Thurston [7] ukázal, že pro Parryho číslo β je $\mathbb{Z}_\beta$ stejnoměrně diskrétní a relativně hustá množina, tj. *delonovská množina* [5]. Pro β, které je navíc Pisotovo, vyhovuje $\mathbb{Z}_\beta$ *meyerovské podmínce*: $\mathbb{Z}_\beta - \mathbb{Z}_\beta \subset \mathbb{Z}_\beta + F$ pro konečnou množinu F. Připomeňme, že Pisotovo číslo je algebraické celé číslo, jehož všechny sdružené kořeny jsou v absolutní hodnotě ostře menší než jedna. Delonovské množiny splňující meyerovskou podmínku jsou vhodnými modely kvazikrystalů, a tedy β-celá čísla pro β Pisotovo číslo slouží jako 1-dimenzionální model kvazikrystalů. V případě kvadratických algebraických čísel splývá pojem Parryho a Pisotova čísla. Tedy zkoumáme-li nekonečná slova, která odpovídají kvadratickému Parryho číslu (a to je právě případ, který nás bude zajímat), studujeme nepřímo jednodimenzionální kvazikrystaly.

Jazyky uzavřené na reverzi

V této kapitole vysvětlíme, proč se při zkoumání palindromické komplexity stačí zaměřit na nekonečná slova přidružená kvadratickému Parryho číslu. V článku [3] je možno nalézt důkaz následujících dvou tvrzení.

1. Nechť u je nekonečné stejnoměrně rekurentní slovo (a to pevné body primitivní substituce jsou). Pak pokud v jazyce $L(u)$ existuje palindrom délky $> n$ pro každé $n \in \mathbb{N}$, potom je $L(u)$ uzavřený na reverzi, tj. s každým slovem obsahuje i jeho zrcadlový obraz.

2. Nechť u_β je pevným bodem substituce (1). Jazyk $L(u_\beta)$ je uzavřený na reverzi tehdy a jen tehdy, když $m = p = 1$, tj. $\varphi(0) = 0^{t_1}1$, $\varphi(1) = 0^{t_2}1$.

Budeme zkoumat nekonečné slovo u_β, které je pevným bodem substituce $\varphi(0) = 0^a1, \varphi(1) = 0^b1$, kde $1 \leq b < a$, tj. substituce odpovídající $d_\beta(1) = ab^\omega$, protože z předchozích tvrzení vyplývá, že v jiných případech Parryho čísel obsahuje nekonečné slovo u_β jen konečně mnoho palindromů.

Komplexita

Poznamenejme, že pro jednoduchá Parryho čísla byla komplexita vyšetřena v [4] a důkazy tvrzení této kapitoly je možno nalézt v [2]. Prvním krokem k určení palindromické komplexity u_β je nalezení jeho komplexity. Faktorová komplexita (nebo jednoduše komplexita) je funkce $C : \mathbb{N} \Rightarrow \mathbb{N}$ definovaná předpisem $C(n) =$ počet různých faktorů délky n. Jednoduchým pozorováním je následující fakt. Nechť M_n je množina levých speciálních faktorů jazyka $L(u_\beta)$ délky n. Potom $C(n+1) - C(n) = \#M_n$. Zaměřme se tedy na levé speciální faktory. Je jich několik druhů.

- Levý speciální faktor $v \in L(u_\beta)$ se nazývá maximální, pokud ani $v0$ ani $v1$ nejsou levé speciály.
- Nekonečné slovo u se nazývá nekonečný levý speciální faktor u_β, pokud každá předpona u je levým speciálním faktorem $L(u_\beta)$.
- Faktor v se nazývá totálně bispeciální, pokud jak $v0$ tak $v1$ jsou levé speciální faktory u_β.

Zřejmé je, že každý levý speciální faktor je předponou maximálního nebo nekonečného levého speciálního faktoru. Máme tedy návod, který říká, že je třeba vyšetřit maximální a nekonečné levé speciální faktory. Provedeme-li to, získáme následující výsledky pro případ, kdy $d_\beta(1) = ab^\omega$ a $b < a - 1$:

- Všechny totálně bispeciální faktory mají tvar:
  $V^{(1)} = 0^b$,
  $V^{(n)} = 0^b 1 \varphi(V^{n-1}) 0^b \quad \text{pro } n > 1$.

  Navíc, $V^{(n-1)}$ je předponou $V^{(n)}$ pro každé $n \in \mathbb{N}$.
- Existuje jediný nekonečný levý speciální faktor $\lim_{n\to\infty} V^{(n)}$.
- Všechny maximální levé speciální faktory mají tvar:
  $U^{(1)} = 0^{a-1}$,
  $U^{(n)} = 0^b 1 \varphi(U^{(n-1)}) 0^b \quad \text{pro } n > 1$.

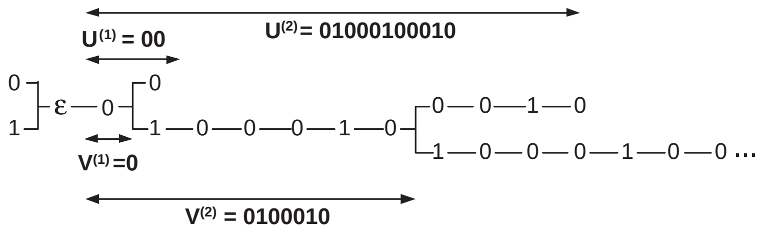
  Navíc, $V^{(n)}$ je předponou $U^{(n)}$ pro každé $n \in \mathbb{N}$.

Situaci můžeme znázornit pomocí stromu levých speciálních faktorů. Ten se skládá z jedné nekonečné větve ($\lim_{n\to\infty} V^{(n)}$) a z ní se odštěpujících větví, které odpovídají maximálním levým speciálním faktorům $U^{(n)}$. K větvení dochází v totálně bispeciálních faktorech $V^{(n)}$. Každý levý speciální faktor je předponou tohoto stromu. Z těchto úvah už snadno plyne vzorec pro komplexitu.

Věta 1. *Komplexita pro $d_\beta(1) = ab^\omega$, $b < a - 1$, splňuje pro každé $n \in \mathbb{N}$,*

$$\triangle C(n) = C(n+1) - C(n) = \begin{cases} 2 & |V^{(k)}| < n \le |U^{(k)}| \quad \text{pro } k \in \mathbb{N}, \\ 1 & \text{jinak.} \end{cases}$$

Uveďme pro ilustraci strom levých speciálních faktorů pro u_β, které je pevným bodem $\varphi(0) = 0001, \varphi(1) = 01$, tedy kde $d_\beta(1) = 31^\omega$. V tomto stromě vidíme všechny levé speciální faktory do délky 14.



Poznamenejme, že pro nekonečné slovo u_β, kde $d_\beta(1) = a(a-1)^\omega$, zjistíme, že neexistuje žádný maximální levý speciální faktor. Strom levých speciálních faktorů se v tomto případě skládá z jediné nekonečné větve. Dostáváme $C(n) = n + 1$. Jde o sturmovské slovo, tj. neperiodické slovo s nejmenší možnou komplexitou.

Palindromická komplexita

Poznamenejme, že palindromická komplexita pro slova přidružená jednoduchým Parryho číslům byla studována v [1] a důkazy a podrobnosti z této kapitoly naleznete v [3]. Palindromická komplexita je funkce $P : \mathbb{N} \Rightarrow \mathbb{N}$ definovaná předpisem $P(n)$ = počet různých palindromů délky n. Vyřešme hned na začátku případ $d_\beta(1) = a(a-1)^\omega$, kdy u_β je sturmovské slovo. Pro sturmovská slova bylo dokázáno, že $P(2n) = P(0) = 1$ a $P(2n+1) = P(1) = 2$ pro všechna $n \in \mathbb{N}$. Nikoho nepřekvapí, že k vyšetření hodnot palindromické komplexity je třeba rozlišit několik druhů palindromů.

- Palindrom $p \in L(u_\beta)$ se nazývá maximální, pokud $0p0 \notin L(u_\beta)$ a $1p1 \notin L(u_\beta)$.
- Nechť $v = ...v_3v_2v_1$ je doleva nekonečné slovo na abecedě $\{0, 1\}$. Označme $\overline{v} = v_1v_2v_3....$ Nechť $d \in \{\varepsilon, 0, 1\}$. Pokud pro každé $n \in \mathbb{N}$ je slovo $p = v_nv_{n-1}...v_1dv_1v_2...v_n \in L(u_\beta)$, potom $vd\overline{v}$ se nazývá nekonečná palindromická větev u_β se středem d.

Je jasné, že každý palindrom je centrálním faktorem maximálního palindromu nebo nekonečné palindromické větve.

- Palindrom p je maximální palindrom $\Leftrightarrow p = U^{(n)}$, tedy maximálnímu levému speciálnímu faktoru, pro nějaké přirozené číslo n. (Přesvědčte se na obrázku, že $U^{(1)}$ a $U^{(2)}$ jsou palindromy.)
- Pro každé $d \in \{\varepsilon, 0, 1\}$ existuje nejvýše jedna nekonečná palindromická větev se středem d.

Uveďme ještě vzorec, který umožňuje využít komplexity k výpočtu palindromické komplexity.

Věta 2. *Pro palindromickou komplexitu u_β, kde $d_\beta(1) = ab^\omega$, $b < a-1$, platí $P(n+1) + P(n) = \triangle C(n) + 2$ pro každé $n \in \mathbb{N}$. A tedy platí:*

$$P(2n+1) = \triangle C(2n) + 2 - P(2n).$$

Z posledního tvrzení vyplývá, že při zkoumání palindromické komplexity se stačí omezit na hledání její hodnoty například jen pro sudá čísla. Existují 4 různé vzorce pro palindromickou komplexitu v závislosti na sudosti či lichosti parametrů a a b. Pro ilustraci uvedeme výsledky pro případ, kdy a i b jsou lichá čísla.

- Existuje nekonečná palindromická větev se středem 0, oboustranná limita palindromů $V^{(n)}$, tedy nám známých totálně bispeciáních faktorů. Jiné nekonečné palindromické větve neexistují.
- Existuje jediný maximální palindrom sudé délky, jediný maximální palindrom se středem 1 a nekonečně mnoho maximálních palindromů se středem 0.

$$U^{(1)} = 0^{(a-1)} \text{ se středem } \varepsilon.$$

$$U^{(2)} \text{ se středem } 1.$$

$$U^{(n)} \text{ se středem 0 a s centrálním faktorem } V^{(n-2)}, \text{ pro } n \geq 3.$$

Připomeňme větu 2, která říká, že pro výpočet palindromické komplexity stačí vědět, že existuje jediný maximální palindrom sudé délky a že neexistuje nekonečná palindromická větev se středem ε.

Závěr

Cílem tohoto článku nebylo uvést přesné vzorce faktorové a palindromické komplexity, ale popsat nástroje, které umožní zkoumat tyto charakteristiky i pro jiná stejnoměrně rekurentní nekonečná slova, která mohou modelovat jiná fenomena než kvazikrystaly, třeba posloupnosti kódující genetickou informaci.

References

[1] P. Ambrož, Ch. Frougny, Z. Masáková, E. Pelantová, *Palindromic complexity of infinite words associated with simple Parry numbers*, to appear in Annales de l'Institut Fourier (2005)

[2] L. Balková, *Complexity for infinite words associated with quadratic non-simple Parry numbers*, to appear in J. Geom. Sym. Phys. WGMP Proceedings (2005)

[3] L. Balková, *Palindromic complexity of infinite words associated with quadratic non-simple Parry numbers*, to appear in J. Theor. Comp. Science (2006)

[4] Ch. Frougny, Z. Masáková, E. Pelantová, *Complexity of infinite words associated with beta-expansions*, RAIRO Theor. Inform. Appl. **38** (2004), 163-185

[5] J. Lagarias, *Geometric models for quasicrystals I. Delone sets of finite type*, Discrete Comput. Geom. **21** (1999), 161-191

[6] D. Shechtman, I. Blech, D. Gratias, and J. Cahn, *Metallic phase with long-range orientational order and no translational symmetry*, Phys. Rev. Lett. **53** (1984), 1951-1954

[7] W. P. Thurston, *Groups, tilings, and finite state automata*, Geom. supercomp. project research report GCG1, University of Minnesota (1989)