Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
Уральский федеральный университет
имени первого Президента России Б.Н. Ельцина

На правах рукописи

Петрова Елена Александровна

О КОМБИНАТОРНЫХ СВОЙСТВАХ БЕСПОВТОРНЫХ ЯЗЫКОВ

01.01.09 - Дискретная математика и математическая кибернетика

Диссертация на соискание учёной степени кандидата
физико-математических наук

Научный руководитель:
профессор, доктор физ.-мат. наук
А. М. Шур

Екатеринбург
2016

Оглавление

Введение

Комбинаторика слов, относительно молодая и активно развивающаяся дисциплина, возникшая на рубеже XIX и XX веков, представляет собой важное направление на стыке современной дискретной математики и теоретических компьютерных наук, изучающее свойства символьных последовательностей (слов). Методы и результаты, разработанные и полученные в комбинаторике слов, широко применимы во многих областях, например, в теории групп, теории кодирования, теории игр, биоинформатике, сжатии данных.

Одним из основных объектов исследований в комбинаторике слов являются *бесповторные языки* — множества слов, не содержащих внутри себя повторяющихся фрагментов, или, как говорят, *избегающих* определённые структурные элементы. Норвежский математик Аксель Туэ в начале XX века занимался конструированием и изучением нескольких конкретных бесповторных языков, и принято считать, что именно с этих работ комбинаторика слов берёт своё начало как самостоятельная дисциплина. Туэ рассматривал, например, бескубный язык над двухбуквенным (бинарным) алфавитом и бесквадратный язык над трёхбуквенным (тернарным) алфавитом — эти языки состоят из слов, не содержащих трёх и двух идущих подряд одинаковых фрагментов, соответственно, — и получил ряд нетривиальных результатов, касающихся, в первую очередь, вопроса о конечности или бесконечности того или иного языка. Со времён Туэ тема бесповторных языков получила весьма широкое развитие во многих направлениях: было обобщено понятие степени, изучены алфавиты большей мощности, обобщено понятие избегаемости с простых повторов до шаблонов, абелевых степеней и т.д., получены результаты о комбинаторной сложности бесповторных языков, рассмотрены бесповторные слова с дополнительными ограничениями. Тем не менее, в отношении языков, рассмотренных Туэ, до сих пор остаётся множество открытых проблем. Некоторые из них решены в данной диссертации.

Любой бесповторный язык является факториальным, то есть замкнутым относительно операции взятия подслова. На множестве слов факториального языка можно ввести три естественных отношения: «быть префиксом», «быть суффиксом», «быть подсловом», относительно которых это множество частично упорядочено, и диаграммы этих частично упорядоченных множеств (ч.у.м.) в первых двух случаях представляют собой деревья, в третьем — ациклический граф более общего вида. Изучение структуры таких ч.у.м. оказывается очень интересной темой в исследовании бесповторных языков, проливающей свет на их внутреннее устройство. Среди задач в этом направлении можно назвать вопросы об изоморфизме поддеревьев, порождённых двумя данными словами в дереве префиксного (суффиксного) порядка, о существовании поддеревьев или подграфов произвольной конечной высоты, о частоте и регулярности ветвления деревьев префиксного (суффиксного) порядка и другие. Основная часть диссертации посвящена проблемам, связанным со структурой ч.у.м. слов бинарного бескубного языка и тернарного бесквадратного языка.

Предварительные сведения

В данном разделе приведены определения, понятия и утверждения, используемые внутри диссертации. Систематическое изложение основ комбинаторики слов см., например, в [40, 41].

Алфавитом называется непустое конечное или счётное множество, элементы которого — *буквы* или *символы*. Произвольная последовательность символов над алфавитом называется *словом*. Для обозначения слов в диссертации используются строчные буквы из конца латинского алфавита — s, t, v, w, u с индексами или без, для обозначения букв в качестве переменных — x, y, z. Некоторые конкретные слова имеют собственные обозначения.

Мы рассматриваем конечные и бесконечные слова и языки над основными алфавитами $\Sigma_2 = \{a, b\}$ (бинарный), $\Sigma_3 = \{a, b, c\}$ (тернарный) и некоторыми вспомогательными алфавитами. Если одна из букв бинарного алфавита обозначена как x, то под $\bar{x}$ будем подразумевать другую букву. Бесконечные вправо слова называются *ω-словами*, бесконечные влево — *ω^*-словами*. Слово, не содержащее букв, или *пустое* слово, обозначается как λ. Множество всех слов над алфавитом Σ вместе с λ образует свободный моноид, обозначаемый как Σ^*, множество всех непустых слов образует свободную полугруппу Σ^+. Произвольное подмножество Σ^* называется *языком* над Σ.

Длина слова w — это количество содержащихся в нём букв, обозначается как $|w|$. Слово w также рассматривается как функция из множества его позиций во множество букв алфавита, $w[i]$, где $i \in \{1, \ldots, |w|\}$ — буква, находящаяся в i-й позиции слова w. *Реверс* w – это слово $\overleftarrow{w} = w[|w|] \cdots w[1]$. Слово w называется *палиндромом*, если оно совпадает со своим реверсом.

Слово v называется *подсловом* w (или, говорят, w *содержит* v, обозначается $v \leqslant w$), если $w = pvs$ для некоторых слов p, s. Причём v является:

- *собственным подсловом* w (обозначается $v < w$), если $p, s \neq \lambda$;

- *префиксом* w, если $p = \lambda$;
- *собственным префиксом* w, если $p = \lambda, s \neq \lambda$;
- *суффиксом* w, если $s = \lambda$;
- *собственным суффиксом* w, если $s = \lambda, p \neq \lambda$.

Для удобства подслово w, начинающееся в позиции i и заканчивающееся в позиции j, будем обозначать как $w[i..j]$.

Слова u и v называются *сопряжёнными*, если найдутся слова p и s такие, что $u = ps$ и $v = sp$. Сопряжённость — отношение эквивалентности, то есть все слова разбиваются на классы сопряжённости. Если линейный порядок на позициях слова u заменить циклическим так, что последняя буква предшествует первой (иначе говоря, соединить концы слова u), то получится циклическая последовательность букв, называемая *циклическим словом* и обозначаемая как (u). Циклические слова находятся в биективном соответствии с классами сопряжённости. Подсловами циклического слова (u) являются обычные слова длины $\leqslant |u|$, включая слово u и сопряжённые с ним.

Бесповторные слова и языки, избегаемость

Натуральное число $p \leqslant |w|$ называется *периодом* слова w, если $w[i] = w[i+p]$ для всех $i \in \{1, \ldots |w| - p\}$. Минимальный период слова w обозначается как $\operatorname{per}(w)$. Слово называется *периодическим*, если оно имеет нетривиальные периоды (не равные длине слова). *Экспонента* слова — это отношение длины этого слова к его минимальному периоду: $\exp(w) = |w|/\operatorname{per}(w)$. Из определения видно, что целую экспоненту имеют в точности слова, являющиеся целыми степенями других слов. Таким образом, экспонента является обобщением понятия степени на нецелые показатели. *Локальной экспонентой* слова будем называть число $\operatorname{lexp}(w) = \sup\{\exp(v) | v \leqslant w\}$. Локальная экспонента характеризует степень повторяемости фрагментов внутри слова. Локальные экспоненты бесконечных слов также называют *критическими экспонентами*; в этом случае супремум в определении может не достигаться. Слова экспоненты 2 называются *квадратами*, экспоненты 3 — *кубами*. Квадраты с периодом длины p будем называть *p-квадратами*. *Минимальный квадрат* — квадрат, не содержащий более коротких квадратов в качестве подслов. Для слов над алфавитом Σ_3 справедлива следующая

Теорема 1 ([19, 75]) *Над тернарным алфавитом существуют минимальные квадраты с любым периодом, кроме 5, 7, 9, 10, 14, 17.*

Периодические слова обладают *свойством взаимодействия периодов*:

Теорема 2 ([27]) *Если слово u имеет периоды p и q, и $|u| \geqslant p + q - (p, q)$, то u имеет период (p, q).*

Слово w *β-свободно* [β^+-свободно], если $\operatorname{lexp}(w) < \beta$ [соответственно, $\operatorname{lexp}(w) \leqslant \beta$]. Говорят также, что слово *избегает* степень β [соответственно, β^+]. Язык называется *бесповторным*, если он состоит из слов, избегающих некоторую фиксированную экспоненту β или β^+. Экспонента β называется k-избегаемой [k-неизбежной], если существует ω-слово над k-буквенным алфавитом, избегающее β [соответственно, множество слов над k-буквенным алфавитом, избегающих β, конечно]. Основные бесповторные языки, рассматриваемые в данной диссертации:

- избегающий экспоненту 2 над тернарным алфавитом, или тернарный бесквадратный язык (обозначается как SF, от англ. “square-free“);

- избегающий экспоненту 3 над бинарным алфавитом, или бинарный бескубный язык (обозначается как CF, от англ. “cube-free“);

- избегающий экспоненту 2^+ над бинарным алфавитом, или бинарный сильно бескубный язык (обозначается как OF, от англ. “overlap-free“).

Морфизм называется *равноблочным*, если образы всех букв под его действием имеют одинаковую длину. Морфизм $h : \Sigma^* \to \Sigma^*$ называется *нестирающим*, если $h(x) \neq \lambda$ для любого $x \in \Sigma$. Если слово w не содержит подслов вида $h(u)$, где h — нестирающий морфизм, то говорят, что w *избегает шаблон u*. Шаблон u называется k-избегаемым [k-неизбежным], если существует бесконечное слово над k-буквенным алфавитом, избегающее u [существует лишь конечное множество слов над k-буквенным алфавитом, избегающих u, соответственно][1]. Мы также будем рассматривать шаблоны специального вида. Будем говорить, что слово w над k-арным алфавитом избегает *буквенный шаблон u*, если u — слово над k-арным *алфавитом переменных* Σ_k и w не содержит подслов вида $h(u)$, где для любого $x \in \Sigma_k \quad |h(x)| = 1$ и для любых $x, y \in \Sigma_k \quad (x \neq y) \Rightarrow (h(x) \neq h(y))$. Например для алфавита переменных $\{x, y, z\}$ слово $w \in \Sigma_3$ избегает шаблон xyz, если оно не содержит следующих подслов: $abc, bca, cab, cba, acb, bac$.

Некоторые известные последовательности

Рассмотрим определения нескольких известных последовательностей, которые активно используются в конструкциях для доказательства основных результатов диссертации.

[1]Легко проверить, что условия «существует бесконечное слово» и «существует бесконечно много слов» эквивалентны ввиду замкнутости относительно подслов и конечности алфавита.

1. Слова Туэ–Морса

Морфизм Туэ–Морса θ над Σ_2^+ определяется следующими формулами: $\theta(a) = ab$, $\theta(b) = ba$. Слова

$$\mathsf{T}_n^a = \theta^n(a),\ \mathsf{T}_n^b = \theta^n(b)\ (n \geqslant 0)$$

называются *блоками Туэ–Морса* или просто n-блоками. Из определения следует, что $\mathsf{T}_{n+1}^x = \mathsf{T}_n^x \mathsf{T}_n^{\bar{x}}$. Значит, последовательности $\{\mathsf{T}_n^a\}$ и $\{\mathsf{T}_n^b\}$ имеют «пределы» — бесконечные вправо *слова Туэ–Морса* T_∞^a и T_∞^b, соответственно. Также мы рассматриваем их реверсы (будем обозначать как ${}_\infty^a\mathsf{T}$ и ${}_\infty^b\mathsf{T}$). *Язык Туэ–Морса* TM состоит из слов Туэ–Морса и всех их подслов. Заметим, что любое слово из TM может быть записано в виде $w = xu_1 \cdots u_n y$, где $x, y \in \Sigma_2 \cup \{\lambda\}$, $u_1, \ldots, u_n \in \{ba, ab\}$. Известно [80], что $\mathsf{TM} \subset \mathsf{OF}$.

2. Слова Аршона

Рассмотрим подстановки α_o («нечётная») и α_e («чётная») над алфавитом Σ_3:

$$\alpha_o:\ a \to abc, b \to bca, c \to cab, \quad \alpha_e:\ a \to cba, b \to acb, c \to bac \qquad (1)$$

Эти подстановки могут быть продолжены до функции $\alpha : \Sigma_3^* \to \Sigma_3^*$ следующим образом. Чтобы получить образ слова w, мы применяем подстановки (1) к каждой букве $w[i]$; если i нечётно [чётно], то используется нечётная [соответственно, чётная] подстановка. Из определения следует, что слово $\alpha^k(a)$ является префиксом слова $\alpha^{k+1}(a)$ для любого $k \geqslant 0$. Следовательно, можно говорить о «предельном» ω-слове

$$\mathsf{A} = \alpha^\infty(a) = abc\,acb\,cab\,cba\,cab\,acb\,cab\,cba\,bca \ldots.$$

Это слово называется *словом Аршона* (впервые было определено в [1]). Будем также рассматривать реверс $\overleftarrow{\mathsf{A}}$ слова A. Слово A бесквадратно [1] и, более того, $(7/4)^+$-свободно [2]. Подслова слова A называются *подсловами Аршона*.

3. Слова Фибоначчи

Морфизм Фибоначчи $\varphi : \Sigma_2 \to \Sigma_2$ определяется равенствами $\varphi(a) = ab, \varphi(b) = a$. Итерация этого морфизма по букве b порождает *слова Фибоначчи*: $\mathsf{F}_{-1} = b, \mathsf{F}_0 = a, \mathsf{F}_1 = ab, \mathsf{F}_2 = aba, \mathsf{F}_3 = abaab, \mathsf{F}_4 = abaababa$ и т.д. Так как F_n является префиксом F_{n+1} для любого $n \in \mathbb{N}$, можно рассмотреть бесконечное слово $\mathsf{F} = \lim_{n\to\infty}\{\mathsf{F}_n\}$, которое является неподвижной точкой морфизма φ. Мы называем F *ω-словом Фибоначчи*, чтобы отличать его от конечных слов Фибоначчи (заметим, что в монографии [41] термин «слово Фибоначчи» используется только для бесконечного слова). Термин объясняется тем, что слова Фибоначчи удовлетворяют рекуррентному соотношению $\mathsf{F}_n = \mathsf{F}_{n-1}\mathsf{F}_{n-2}$, в частности, их длины являются числами Фибоначчи. Слово F принадлежит к знаменитому классу слов Штурма, см. например [41, Ch. 2].

Порядки в бесповторных языках и предмаксимальные слова

На множестве слов бесповторного языка L естественным образом вводятся следующие отношения порядка:

1. $u \leqslant_p v$, если $u, v \in L$ и существует слово s такое, что $us = v$ (*префиксный порядок*);

2. $u \leqslant_s v$, если $u, v \in L$ и существует слово p такое, что $pu = v$ (*суффиксный порядок*);

3. $u \leqslant v$, если $u, v \in L$ и существуют слова p и s такие, что $pus = v$ (*подсловный порядок*).

Относительно введённых порядков множество слов L образует ч.у.м. В случае префиксного и суффиксного порядков диаграммы таких ч.у.м. представляют собой деревья (префиксное и суффиксное, соответственно), причём для каждого бесповторного языка префиксное и суффиксное деревья изоморфны ввиду замкнутости языка относительно реверса. В случае подсловного порядка диаграмма ч.у.м. является ациклическим графом более общего вида. Рёбра в нём можно разделить на два вида: $w \xrightarrow{x} wx$ (префиксные) и $w \xrightarrow{x} xw$ (суффиксные).

Пример 1

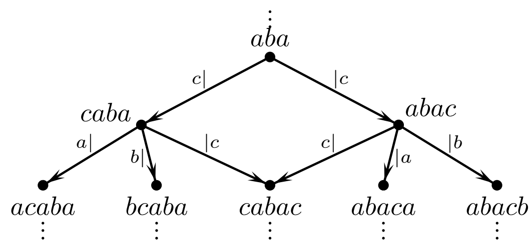


Рис. 1: Фрагмент диаграммы ч.у.м. слов языка SF в случае подсловного порядка

Поддерево префиксного или суффиксного дерева, *порождённое некоторой вершиной* v, — это дерево с корнем в v, содержащее всех потомков v из исходного дерева. Будем говорить, что поддерево префиксного или суффиксного дерева бесповторного языка имеет *индекс* n, если метка корня этого поддерева — слово длины n.

Пусть L — язык над Σ и $w \in L$. Слово $u \in \Sigma^*$ такое, что $uw \in L$, называется *левым контекстом* w в L. В случае, когда L — факториальный язык, *фиксированным левым контекстом* слова w называется такое слово v, что поддерево, порождённое узлом с меткой w в дереве суффиксного порядка, содержит узел vw и на пути от w к wv все узлы имеют ровно одного потомка. Слово w называется *максимальным слева* [*предмаксимальным слева*], если оно не имеет левых контекстов [соответственно, имеет конечное число левых контекстов]. *Уровнем* предмаксимального слева слова w называется длина его самого длинного левого контекста (или высота поддерева, порождённого словом w в дереве суффиксного порядка на L). Таким образом, у максимального слова уровень 0. Понятия предмаксимального справа и максимального справа слова определяются симметрично. Говорят, что слово *предмаксимально*, если оно предмаксимально и слева, и справа. *Уровнем* предмаксимального слова w называется пара $(n, k) \in \mathbb{N}_0^2$ такая, что n и k – длины самого длинного левого и самого длинного правого контекстов w.

Пример 2

- $abacab$ — слово с фиксированным левым контекстом bc в языке SF;
- $aabaabaa$ — максимальное слово в языке CF;
- $cabacabcbabcabacabcba$ — предмаксимальное слева слово уровня 2 в языке SF;
- $abaababaababa$ — предмаксимальное справа слово уровня 2 в языке CF.

Обзор исследований по теме диссертации

В серии из четырёх работ, опубликованных в период с 1906 по 1914 годы, Аксель Туэ рассматривал несколько комбинаторных задач, возникших в ходе изучения символьных последовательностей. Две из этих работ связаны с повторами в конечных и бесконечных словах. Из-за того, что эти статьи были опубликованы в малоизвестных норвежских журналах, недоступных для международного математического сообщества, результаты Туэ были никому не известны до 1950-х годов и были не один раз открыты заново. В [79] построено бесквадратное ω-слово над Σ_3. Там же приводится конструкция для бесквадратного ω-слова над четырёхбуквенным алфавитом, основанная на итерации морфизма. В [80] Туэ представил морфизм, ныне известный как морфизм Туэ–Морса (см. раздел Предварительные сведения) и доказал, что неподвижные точки этого морфизма — слова T_∞^a и T_∞^b — не только бескубны, но и сильно бескубны. Легко видеть, что над бинарным алфавитом квадраты неизбежны: самое длинное бинарное бесквадратное слово — это aba. Поэтому для бинарного алфавита слово Туэ–Морса является *граничным* словом в том смысле, что оно избегает минимально возможную степень. О таких граничных словах для бо́льших алфавитов будет сказано ниже. Последовательность Туэ–Морса позже появлялась в других работах, авторы которых были не знакомы с результатами Туэ, например, у Морса [46] и у Аршона [1]. В статье [1] также приводится конструкция для построения тернарного слова Аршона и доказывается, что это слово является бесквадратным. Позже Клепинин и Суханов [2] доказали, что слово Аршона $(7/4)^+$-свободно. Как и слово Туэ–Морса для бинарного алфавита, слово Аршона является граничным для тернарного алфавита.

В исследованиях, связанных с бесповторными языками, можно выделить несколько основных направлений:

- бесповторные морфизмы: построение и критерии;
- нахождение граничных экспонент (*границы повторяемости*) для алфавитов;
- изучение комбинаторной сложности языков;
- изучение бесповторных слов с дополнительными ограничениями;
- изучение структуры ч.у.м., образованных словами бесповторных языков.

Рассмотрим подробнее каждое из этих направлений.

Бесповторные морфизмы

Морфизм $f : \Sigma^+ \to \Sigma^+$ *избегает экспоненту* β, если из того, что $\mathrm{lexp}(u) < \beta$ следует, что $\mathrm{lexp}(f(u)) < \beta$ для любого слова u. Морфизмы, избегающие экспоненты, называют бесповторными. Такие морфизмы являются самым эффективным способом порождения бесконечных бесповторных последовательностей. Поэтому построение бесповторных морфизмов исторически является одной из самых первых задач в исследовании бесповторных языков. Ею, в частности, занимался и Туэ в своих работах. Он доказал, что морфизм Туэ–Морса является сильно бескубным и, более того, любой сильно бескубный морфизм построен из блоков Туэ–Морса:

Теорема 3 ([80]) *Для любого сильно бескубного морфизма h над бинарным алфавитом существует $k \in \mathbb{N}$ такое, что либо $h(a) = \theta^k(a), h(b) = \theta^k(b)$, либо $h(a) = \theta^k(b), h(b) = \theta^k(a)$.*

В [79] Туэ построил бесквадратный морфизм для четырёхбуквенного алфавита и затем нашёл условия, при которых произвольный морфизм для четырёхбуквенного алфавита является бесквадратным. В другой работе он сделал это для тернарного алфавита. Морфизм $h : \Sigma^* \to \Sigma^*$ называется *фактор-свободным*, если $\forall a, b \in \Sigma$ условие $h(a) \leqslant h(b)$ влечёт $a = b$.

Теорема 4 ([80]) *Пусть $h : \Sigma_3 \to \Sigma_3$ — нестирающий фактор-свободный морфизм. Если для любого трёхбуквенного бесквадратного слова w слово $h(w)$ бесквадратно, то h — бесквадратный морфизм.*

В [67] эта теорема доказана для случая алфавитов произвольной мощности.

Наилучшая с точки зрения вычислительной сложности характеризация бесквадратных морфизмов над произвольными алфавитами даёт следующая теорема, доказанная Крошмором:

Теорема 5 ([17]) *Пусть $h : \Sigma_n \to \Sigma_l$ — нестирающий морфизм, $M = \max_{a \in \Sigma_n} |h(a)|, m = \min_{a \in \Sigma_n} |h(a)|, k = \min(3, 1 + [(M-3)/m])$. Тогда h — бесквадратный морфизм тогда и только тогда, когда для любого слова w длины не более k слово $h(w)$ бесквадратно.*

Для автоматизированной проверки морфизмов на бесповторность удобно иметь некоторый конечный набор слов такой, что если образы всех слов набора не содержат подслов запрещённой степени, то можно заключить, что образ любого слова под действием морфизма не содержит таких подслов. Такие наборы называются *тестовыми множествами* морфизмов. Если удаётся найти множество из одного слова, то это слово называют *тестовым словом*.

Для случая сильно бескубных морфизмов тестовое слово было предъявлено Берстелем и Сиболдом в [11]: морфизм h сильно бескубен тогда и только тогда, когда $h(T_4^a)$ сильно бескубно. Этот результат был усилен в [63], где было доказано, что слово $bbabaa$ является тестовым для сильно бескубных морфизмов и никакое слово длины 5 и менее не может быть использовано в этом качестве, а также дана полная характеризация тестовых множеств для сильно бескубных морфизмов.

Кархумяки изучал бинарные бескубные морфизмы и показал [34], что для бинарного морфизма h со свойством $h(a) = au, u \neq \lambda$ множество $\{h^n(a) | n \leqslant 10\}$ является тестовым. Эти исследования были продолжены Ришомом и Влазинским в [64]. Ими дана характеризация тестовых множеств бинарных бескубных морфизмов: слова тестового множества должны содержать 12 определённых подслов. В частности, справедлива следующая

Теорема 6 ([64]) *Морфизм $f : \Sigma_2^+ \to \Sigma_2^+$ бескубен тогда и только тогда, когда слово $f(aabbababbabbaabaababaabb)$ бескубно.*

Длина тестового слова (24) является оптимальной. В этой же работе доказано, что для k-свободных морфизмов над алфавитами мощности 3 и более, где $k \geqslant 3$, не существует конечных тестовых множеств.

Граница повторяемости

Граница повторяемости $\mathsf{RT}(k)$ определяется как инфимум множества экспонент β таких, что существует бесконечное β^+-свободное слово над k-буквенным алфавитом. Ясно, что в однобуквенном алфавите никакая экспонента не является избегаемой. Туэ показал, что $\mathsf{RT}(2) = 2$. В 1972 году Ф. Дежан построила $(7/4)^+$-свободное слово над тернарным алфавитом, показала, что любое тернарное слово длины более 39 не является $(7/4)^+$-свободным (тем самым доказав, что $\mathsf{RT}(3) = 7/4$) и выдвинула гипотезу, что таблица значений функции $\mathsf{RT}(k)$ выглядит следующим образом [26]:

k	1	2	3	4	5	...	n	...
$\mathsf{RT}(k)$	∞	2	7/4	7/5	5/4	...	$n/(n-1)$	...

Наблюдение о том, что $k/(k-1)$ является нижней границей для $\mathsf{RT}(k)$, также принадлежит Дежан. На полное же доказательство гипотезы потребовались усилия многих исследователей в течение почти сорока лет. Подход, использованный Туэ и Дежан для бинарного и тернарного алфавитов, — порождение граничных слов с помощью итерации морфизмов — принципиально не годится для алфавитов большей мощности, как показал Бранденбург [12].

Выход из этой ситуации нашёл Пансьё, предложив идею бинарного кодирования слов над k-буквенным алфавитом, в которых любые $(k-1)$ подряд идущие буквы различны (таким свойством обладает любое $\mathsf{RT}(k)^+$-свободное слово) [50]. Это кодирование активно используется в доказательстве основных результатов диссертации и будет описано позднее в подразделе 1.3.1. Сам Пансьё доказал гипотезу Дежан для четырёхбуквенного алфавита, а кроме того, свёл задачу нахождения k-буквенного граничного ω-слова к нахождению бинарного кода, соответствующего $\mathsf{RT}(k)^+$-свободному ω-слову. Такие коды могут быть построены методом итерации морфизмов. Мулен-Оланье установил достаточные условия для морфизмов, порождающих эти коды, построил несколько таких морфизмов с помощью компьютерных вычислений и доказал гипотезу для $5 \leqslant k \leqslant 11$ [48], позже для $7 \leqslant k \leqslant 14$ соответствующие морфизмы были построены в совместной работе Карри и Мохаммад-Нури, которые опирались на подход, отличный от предложенного Мулен-Оланье [45].

Прорыв в доказательстве в 2007 году совершил Карпи, предложив конструкцию для построения морфизмов в случаях $k \geqslant 33$ [14]. Сначала строится ω-слово над вспомогательным алфавитом мощности $m = \lfloor (k-3)/6 \rfloor$, которое затем отображается в бинарный код граничного слова для k-арного алфавита. Конструкция Карпи подходит для всех $m \geqslant 5$. Карри и Рамперсад смогли усилить результат Карпи на случай $m = 4$, доказав тем самым гипотезу для случаев $27 \leqslant k \leqslant 32$ [20]; им же удалось развить идеи Мулен-Оланье и подобрать на компьютере морфизмы для $15 \leqslant k \leqslant 26$ [23]. Примерно в то же время независимо от Карри и Рамперсада доказательство гипотезы завершил Рао, построив бинарные коды как образы слов Туэ–Морса под действием подходящих морфизмов [60].

За долгое время доказательства гипотеза Дежан породила множество задач, среди которых существуют усиления гипотезы (например, исследование конечной границы повторяемости [6–8] и экспоненциальная гипотеза [38, 49, 81]), обобщения гипотезы [28, 32], исследование границы повторяемости для других видов степеней [35, 43, 66] и других видов слов [3, 4, 76].

Комбинаторная сложность

Для оценки количества слов языка используется понятие *комбинаторной сложности* языка. Она определяется как функция $C_L(n) : \mathbb{N} \to \mathbb{N}$, значение которой равно количеству слов длины n в языке L. Впервые это понятие было введено в работе Морса и Хедлунда [47]. Комбинаторной сложности как естественной характеристике языка посвящено большое количество работ, особенно в области бесповторных языков; см. обзор [78].

Скорость роста комбинаторной сложности характеризуется *индексом роста* языка $\alpha(L)$, который определяется следующим образом:

$$\alpha(L) = \overline{\lim_{n\to\infty}} (C_L(n))^{1/n},$$

при этом в случае факториального языка верхний предел превращается просто в предел. Индекс роста, больший 1, говорит об экспоненциальном росте количества слов в языке, $\alpha(L) = 1$ соответствует субэкспоненциальному росту, а индекс роста конечного языка равен нулю.

Рестиво и Салеми в [61] установили, что язык OF имеет полиномиальную комбинаторную сложность, а также нашли верхнюю оценку для степени полинома: $C_{\mathsf{OF}}(n) \leqslant Cn^r$, где $C > 0$ и $r = \log_2 15 \approx 3.906$. Эта оценка была улучшена в [36]: $C_1 n^{1.155} \leqslant C_{\mathsf{OF}}(n) \leqslant C_2 n^{1.587}, C_1, C_2 > 0$. Позже Кассень [16] доказал, что функция C_{OF} не может быть выражена в виде полинома, т.е. что числа $\alpha = \sup\{r \mid \exists C > 0 \quad \forall n \quad C_{\mathsf{OF}} \geqslant Cn^r\}$ и $\beta = \inf\{r \mid \exists C > 0 \quad \forall n \quad C_{\mathsf{OF}} \leqslant Cn^r\}$ не равны друг другу, и построил верхнюю оценку для α и нижнюю для β, получив в итоге $1.155 < \alpha < 1.276 < 1.332 < \beta < 1.587$. Юнгерс, Протасов и Блондель [33] показали, что α и β являются совместными спектральными характеристиками пары неотрицательных матриц. Точные значения $\alpha = 1.27355\ldots$ и $\beta = 1.33224\ldots$ были недавно вычислены Гулиельми и Протасовым [31] в качестве приложения разработанного ими метода точного вычисления таких характеристик.

В работе, посвящённой бесповторным морфизмам [12], Бранденбург показал, в частности, что языки CF и SF имеют экспоненциальную комбинаторную сложность и получил следующие оценки: $2d_1^n \leqslant C_{\mathsf{CF}}(n) \leqslant 2d_2^n$, где $d_1 \geqslant 2^{1/9} \geqslant 1.08, d_2 \leqslant 1251^{1/17} \leqslant 1.552$; $6 \cdot 1.032^n \leqslant C_{\mathsf{SF}} \leqslant 6c_2^n$, где $c_2 = 1172^{1/22} \leqslant 1.38$. Для языка тернарных бесквадратных слов в [13] верхняя оценка индекса роста была улучшена до 1.316. Позднее с помощью метода, разработанного в [13], была получена нижняя оценка 1.1099 [30]. Наконец, в [71, 74] Шуром был разработан эффективный алгоритм вычисления верхней оценки индекса роста произвольного бесповторного языка над конечным алфавитом, основанный на последовательном приближении бесповторного языка регулярными языками; алгоритм вычисления нижних оценок был предложен Колпаковым [37, 39] и доработан Шуром [72]. Применение этих алгоритмов позволило получить для обсуждаемых языков следующие оценки: $1.457573 \leqslant \alpha(\mathsf{CF}) \leqslant 1.457577$, $1.301759 < \alpha(\mathsf{SF}) < 1.301762$.

В [73] Шуром получены асимптотические формулы для индексов роста $\alpha(k,\beta)$ β-свободных языков над произвольным k-буквенным алфавитом, где $\beta \geqslant 2$.

Теорема 7 ([73]) *Пусть $p \geqslant 2$ — натуральное число и $\beta \in [p^+, p+1]$. Тогда справедливо*

$$\alpha(k,\beta) = \begin{cases} k - \frac{1}{k^{p-1}} + \frac{1}{k^p} - \frac{1}{k^{2p-2}} + O(\frac{1}{k^{2p-1}}), & \text{если } \beta \in (p, p+\frac{1}{2}], \\ k - \frac{1}{k^{p-1}} + \frac{1}{k^p} + O(\frac{1}{k^{2p-1}}), & \text{если } \beta \in (p+\frac{1}{2}, p+1] \end{cases}$$

Теорема 8 ([73]) *Справедливы следующие равенства:*

$$\begin{array}{lcl} \alpha(k+1,2) & = & k - \frac{1}{k} - \frac{1}{k^3} + O(\frac{1}{k^5}), \\ \alpha(k,2^+) & = & k - \frac{1}{k} - \frac{1}{k^3} - \frac{1}{k^4} + O(\frac{1}{k^5}). \end{array}$$

Бесповторные слова с дополнительными ограничениями

Очень интересной и плодотворной темой в исследовании бесповторных языков является построение и изучение бесповторных слов с дополнительными ограничениями. Как правило, дополнительное ограничение заключается в том, что помимо повторов в конструируемом бесконечном слове требуется избежать ещё какие-то подслова определённого вида либо ограничить количество их вхождений в слово.

Первым, кто занимался этой темой в комбинаторике слов, опять же, по-видимому, следует считать Туэ. В [80] он не только доказал, что квадраты избегаемы над тернарным алфавитом, но и построил бесквадратные слова с дополнительными ограничениями. Над тернарным алфавитом существует 2 группы («правая» и «левая») бесквадратных слов длины 3 таких, что их первая и последняя буквы совпадают: aba, bcb, cac и bab, cbc, aca. Туэ показал, что в SF существуют ω-слова, избегающие любую пару этих трёхбуквенных слов такую, что одно слово — из правой группы, а другое — из левой. Этот результат имеет интересную интерпретацию в терминах кодов Пансьё и основанных на них маршрутных кодах бесквадратных слов, которые будут обсуждаться в подразделе 1.3.1. Также в [80] доказано, что, во-первых, три таких подслова одновременно избежать нельзя, а во-вторых, любое слово вида xyz, где x, y, z — различные буквы из Σ_3, неизбежно.

Много работ посвящено избегаемости шаблонов в бинарных словах. В [29] доказано, что для любого 2-избегаемого бинарного шаблона p язык, избегающий p, имеет экспоненциальную комбинаторную сложность. В [65] установлено, что все бинарные шаблоны длины 6 являются 2-избегаемыми, и это значение минимально (существуют 2-неизбежные шаблоны длины 5). Классификацию неизбежных бинарных шаблонов завершил Кассень в [15]. В [42] рассмотрены шаблоны, избегаемые в бинарных бескубных словах, и дано их полное описание: шаблон p является избегаемым в CF тогда и только тогда, когда содержит в себе хотя бы одно из слов $xyxyx$, $xxyxxy$, $xxyxyy$, $xxyyxx$, $xxyyxyx$, $xyxxyxy$, $xyxxyyxy$ с точностью до реверса; любой язык $L \subset$ CF, избегающий любой бинарный шаблон, кроме $xyxyx$, имеет экспоненциальную комбинаторную сложность.

В [59] производится построение бескубного слова, избегающего квадраты с периодами $\leqslant 4$. Позднее возник естественный вопрос: можно ли в $RT(k)^+$-свободных бесконечных словах над k-буквенным алфавитом обойтись конечным числом подслов экспоненты $RT(k)$? В серии работ [7–9] вводится и исследуется понятие *конечной границы повторяемости* $FRT(k)$ для k-буквенного алфавита. Это наименьшее рациональное число такое, что существует бесконечное слово w над k-буквенным алфавитом, локальная экспонента которого не превышает $FRT(k)$ и w содержит лишь конечное число подслов экспоненты $RT(k)$. В этих терминах из результата, полученного в [68], следует, что $FRT(2) = 7/3$. В [8] речь также идёт о конечной границе повторяемости для бинарного алфавита, там было построено $(7/3)^+$-свободное слово с минимально возможным количеством квадратов (12) и сделано предположение, что $FRT(3) = RT(3) = 7/4$. В [7] предположение было доказано, а также выдвинута гипотеза $FRT(4) = 7/5$. Наконец, в [9] было завершено исследование конечной границы повторяемости доказательством того, что $FRT(k) = RT(k), k \geqslant 4$. Также в [81] Тунев и Шур доказали, что для всех нечётных $k \leqslant 101$ число k-арных граничных слов, содержащих только однобуквенные повторы, эспоненциально.

В работах [21, 22] обсуждаются конструкции, решающие в некотором смысле противоположную задачу: в первой производится построение $(7/3)^+$–свободного слова, в каждой позиции которого начинаются сколь угодно большие квадраты и доказывается, что $7/3$ — оптимальное значение для таких слов; во второй построено бесконечное бинарное бескубное слово, содержащее экспоненциальное количество квадратов для каждой длины.

Структура частично упорядоченного множества слов бесповторного языка

Изучение структуры ч.у.м. слов бесповторных языков относительно префиксного, суффиксного и подсловного порядков формирует широкий круг задач, решение которых проливает свет на внутреннее устройство бесповторного языка. Наибольший интерес представляет исследование префиксного (или суффиксного) дерева языка, которому посвящён ряд работ. В [10] показано, что в дереве любого бесповторного языка каждое поддерево имеет хотя бы один лист. В [18] Карри доказал, что для некоторых k-свободных языков над алфавитом Σ разрешим вопрос о конечности поддерева, порождённого данным словом, а именно, при $k > 6$ и $|\Sigma| > 1$, $k > 3$ и $|\Sigma| > 2$, $k > 2$ и $|\Sigma| > 3$, $k > 1$ и $|\Sigma| > 4$, кроме того, деревья этих языков не содержат бесконечных изолированных ветвей, т.е. любая бесконечная ветвь ветвится бесконечно часто. Позднее Карри и Шелтон [25] распространили этот результат на все остальные k-свободные языки: если язык бесконечен, то его дерево обладает этими свойствами. Доказательство разрешимости вопроса о конечности порождённого поддерева, однако, осталось неконструктивным, то есть не был предъявлен алгоритм, решающий эту задачу.

Остальные известные результаты касаются конкретных языков. Самую простую структуру с точки зрения дерева имеет язык OF благодаря своей полиномиальной сложности и тесной связи с языком слов Туэ-Морса. Рестиво и Салеми в [62] предложили алгоритм проверки конечности поддерева, порождённого узлом в дереве языка OF, а также доказали существование в этом дереве поддеревьев произвольной конечной высоты. В [77] для этого же языка Шур показал разрешимость за линейное время вопроса об изоморфизме поддеревьев, порождённых двумя данными узлами и, кроме того, представил другой алгоритм проверки конечности поддерева, который к тому же вычисляет высоту поддерева в конечном случае.

В работах [69, 70] Шелтон и Сони занимались изучением структуры префиксного дерева языка тернарных бесквадратных слов. В частности, ими был доказан следующий результат:

Теорема 9 ([70]) *Если в префиксном дереве языка* SF *все бесконечные пути из узла* w *проходят через один и тот же узел* wv*, то* $|v| \leqslant C|w|^{2/3}$*, где*

C — некоторая абсолютная константа.

Иными словами, здесь получена верхняя субполиномиальная оценка на размер (глубину) конечного поддерева, порождённого узлом в дереве SF. Естественно выдвинуть гипотезу о том, что на самом деле справедлива логарифмическая оценка. В пользу этого предположения говорит, например, экспоненциальная сложность языка. Ручная проверка для поддеревьев небольшого размера показывает, что можно выдвинуть ещё более сильную гипотезу: конечное поддерево, порождённое словом длины n, имеет $O(\log n)$ узлов. Однако улучшением оценки Шелтона и Сони ввиду сложности задачи с тех пор никто не занимался.

Рестиво и Салеми в [62], помимо вопросов о конечности порождённого данным узлом поддерева и существовании конечных поддеревьев произвольной высоты, сформулировали следующую задачу: пусть даны два слова $u, v \in L$; существует ли такое слово w, что $uwv \in L$, и как, в случае положительного ответа, построить это слово? Эта задача, очевидно, тоже имеет связь со структурой ч.у.м. языка L. Авторы сформулировали все эти проблемы для языков OF и SF и привели решения для OF. Решение вопроса о конечности порождённого поддерева для SF следует из Теоремы 9: конечность поддерева проверяется за субэкспоненциальное время. Остальные проблемы, сформулированные Рестиво и Салеми, для языка SF до сих пор оставались открытыми. В книге Аллуша и Шаллита [5] также фигурирует открытая проблема о существовании конечных поддеревьев произвольной высоты в языке тернарных бесквадратных слов.

Обзор диссертации

Объект исследования и цели диссертации

Основные объекты исследования в диссертации — бинарный бескубный и тернарный бесквадратный языки CF и SF, а также ч.у.м., образованные на множествах слов этих языков отношениями префиксного, суффиксного и подсловного порядков. В качестве вспомогательных объектов и конструкций используются слова Туэ–Морса, слова Аршона, слова Фибоначчи и бинарные коды Пансьё.

Основная цель диссертации — развитие теории бесповторных языков в следующих направлениях: структурный анализ префиксных/суффиксных деревьев и подсловных ациклических графов, образованных множеством слов бесповторного языка; избегаемость буквенных шаблонов в бесповторных языках.

Основные задачи, рассматриваемые в диссертации:

1. конструктивно доказать существование поддеревьев произвольной конечной высоты в суффиксных (префиксных) деревьях и подсловных ациклических графах языков бинарных бескубных слов и тернарных бесквадратных слов;
2. доказать логарифмическую от длины слова оценку длины его фиксированного контекста и найти нижнюю границу частоты ветвления префиксного (суффиксного) дерева тернарного бесквадратного языка;
3. классифицировать буквенные шаблоны по их избегаемости в тернарном бесквадратном языке.

Основные методы

Для доказательства результатов, представленных в диссертации, в основном используются методы комбинаторики слов, основанные на свойствах периодических слов и граничных слов над бинарным и тернарным алфавитами (слова Туэ–Морса и Аршона, соответственно), слов Фибоначчи, кодировании бесповторных слов бинарными кодами Пансьё и «маршрутными кодами» (метод маршрутных кодов предложен Шуром в [75]). Также для доказательства одного из вспомогательных результатов используется метод решения одной из модификаций задачи составления расписаний.

Структура диссертации и основные результаты

Диссертация состоит из введения, трёх глав, заключения и списка литературы. Общий объём диссертации составляет 89 страниц. Библиографический список содержит 81 наименование.

Во введении определяются основные используемые понятия, приводится ряд известных утверждений, использующихся в доказательстве основных и вспомогательных результатов диссертации, а также дан обзор исследований в области бесповторных языков.

Первая глава посвящена построению серий односторонних и двусторонних предмаксимальных слов и состоит из трёх разделов. Основные результаты первой главы сформулированы в следующих двух теоремах:

Теорема 10 ([54]) *В языке* CF *существуют*
а) предмаксимальные слева слова любого уровня $n \in \mathbb{N}_0$*;*
б) предмаксимальные слова любого уровня $(n, k) \in \mathbb{N}_0^2$.

Теорема 11 ([55]) *В языке* SF *существуют*
а) предмаксимальные слева слова любого уровня $n \in \mathbb{N}_0$*;*
б) предмаксимальные слова любого уровня $(n, k) \in \mathbb{N}_0^2$.

Раздел 1.1 содержит общее описание основной конструкции для построения односторонних предмаксимальных слов в языках CF и SF. Построение производится в два этапа. Главная идея первого этапа заключается в итеративном построении префиксной последовательности слов $\{w_n\}$ таких, что слово w_n имеет фиксированный левый контекст длины не менее n, путём запрета появления «нежелательных» букв. Для того, чтобы каждое w_n было соответственно бескубным или бесквадратным, необходимо при его построении из слов w_{n-1} вставлять в нужные места определённые «буферные» слова. Второй этап достройки слова с фиксированным контекстом до предмаксимального слева слова нужного уровня идейно не отличается от первого.

Раздел 1.2 посвящён предмаксимальным словам в бинарном бескубном языке. В подразделе 1.2.1 описано построение последовательности буферных слов, основанное на свойствах слов Туэ–Морса и бескубных морфизмах. В подразделе 1.2.2 производится доказательство бескубности построенных односторонних слов, то есть утверждения а) Теоремы 10. Наконец, в подразделе 1.2.3 доказана вторая часть Теоремы 10.

В разделе 1.3 идёт речь о предмаксимальных словах в тернарном бесквадратном языке. Подраздел 1.3.1 отведён под определение кодов Пансьё, основанных на них маршрутных кодов тернарных бесквадратных слов и описание их свойств, на которых основана конструкция предмаксимальных слов для SF. На самом деле строятся не сами слова, а их маршрутные коды. В подразделе 1.3.2 описаны особенности основной конструкции в связи с применением этих маршрутных кодов. В подразделе 1.3.3 вычислен маршрутный код ω-слова Аршона, реверс которого используется в качестве левого контекста конструируемых слов. Свойства запрещаемых на каждом шаге построения левых контекстов обсуждаются в подразделе 1.3.4. В подразделе 1.3.5 описано построение буферных слов и доказана корректность конструкции для слов с фиксированным левым контекстом. Наконец, в подразделе 1.3.6 производится доказательство Теоремы 11.

Во второй главе идёт речь о тернарных бесквадратных буквенных шаблонах длины 5 и 6 в языке тернарных бесквадратных слов (из результатов Туэ [80] и свойств бесквадратных кодов Пансьё следует, что все шаблоны меньших длин неизбежны в SF). Основной результат этой главы —

Теорема 12 ([51]) *Следующие тернарные бесквадратные буквенные шаблоны избегаемы в языке* SF*:*
а) $xyxzx$, $xyzxy$*;*
б) $xyxzyz$ *и все шаблоны длины 6, подсловами которых являются шаблоны из а).*
Все другие шаблоны длины $\leqslant 6$ *неизбежны в* SF.

Доказательство того, что остальные 5- и 6-буквенные неизбежны, основано на свойствах кодов Пансьё тернарных бесквадратных слов. Доказательство Теоремы 12 опирается на маршрутные коды, описанные в подразделе 1.3.1. При ближайшем рассмотрении кодов Пансьё и маршрутных кодов шаблонов, фигурирующих в утверждении Теоремы 12, оказывается, что существование тернарных бесквадратных слов, избегающих эти шаблоны, равносильно существованию бесконечных бесквадратных маршрутных кодов с определённым свойством, а именно, кодов, построенных с использованием только двух букв из трёх возможных букв алфавита маршрутных кодов. Таким образом, для доказательства Теоремы 12 необходимо построить бинарные слова, удовлетворяющие некоторым условиям, причём условия одинаковы в случае всех трёх шаблонов. На роль такого бинарного слова подходит ω-слово Фибоначчи. Доказательство того, что на основе этого слова можно построить необходимые маршрутные коды, приведено в подразделе 2.1. Совмещая это с предварительными рассуждениями о неизбежных буквенных шаблонах, получаем утверждение Теоремы 12. В следующем подразделе 2.2 доказана теорема о локальных экспонентах тернарных бесквадратных слов, избегающих буквенные шаблоны длины $\leqslant 6$:

Теорема 13 ([51]) *Минимальная критическая экспонента тернарного бесквадратного слова, избегающего буквенный шаблон длины* $\leqslant 6$
(1) равна 15/8 для шаблона $xyxzx$,
(2) равна 11/6 для шаблона $xyzxy$, *и*
(3) не превосходит $1+\rho/2$ *для шаблона* $xyxzyz$, *где* ρ *— золотое сечение.*

В подразделе 2.3 обсуждается дальнейшая перспектива развития этой темы для буквенных шаблонов бóльших длин.

Третья глава посвящена исследованию структуры префиксного дерева тернарного бесквадратного языка и нахождению некоторых её комбинаторных характеристик. Основными результатами являются следующие утверждения:

Теорема 14 ([57]) *Любой фиксированный контекст тернарного бесквадратного слова* w *имеет длину* $O(\log|w|)$.

Теорема 15 ([57]) *Пусть* $w \in \mathsf{SF}$ *—* ω*-слово и* $b(n)$ *— количество ветвящихся узлов на пути* $(\lambda, w[1..n])$ *в дереве префиксного порядка языка* SF. *Тогда* $\liminf_{n\to\infty} \frac{b(n)}{n} \geqslant \frac{2}{9}$.

На основании этих теорем можно сделать вывод, что рассматриваемое префиксное дерево ветвится достаточно равномерно. Индекс роста языка SF говорит о том, что узел в префиксном дереве в среднем имеет приблизительно 1.3 сыновей; Теорема 15 утверждает, что вдоль любого бесконечного пути среднее количество сыновей не меньше 11/9, а Теорема 14 — что любые два соседних ветвящихся узла на пути расположены близко. При этом результат Теоремы 14 асимптотически неулучшаем: примеры слов из SF с логарифмической длиной фиксированного контекста построены в Главе 1.

Первый раздел отведён под доказательство Теоремы 14, которая является усилением Теоремы 9. Иными словами, утверждение Теоремы 14 гласит, что длина максимального начального неветвящегося сегмента в поддереве, порождённом узлом w, ограничена $C\log(|w|)$, где C — некоторая константа. Пусть $w \in \mathsf{SF}$ и v — некоторый фиксированный правый контекст w. Если заменить любую букву в суффиксе длины $|v|$ слова wv на отличную от неё и от предыдущей буквы, получится квадрат. Эти квадраты условно разделяются на две группы: короткие (с периодом $\leqslant 17$) и длинные (с периодом > 17), и для каждой группы по отдельности оценивается количество букв, фиксированных ими описанным выше образом. Подраздел 3.1.1 содержит доказательство оценки для коротких квадратов: не более $\frac{2}{3}l + O(1)$ позиций слова длины l фиксированы короткими квадратами. В основе доказательства лежит оптимизированный перебор всевозможных коротких квадратов.

В подразделе 3.1.2 рассмотрены длинные квадраты. Вначале доказывается свойство взаимодействия периодов в бесквадратных словах, которое в общих словах звучит следующим образом: если буква бесквадратного слова в некоторой позиции фиксирована некоторым длинным квадратом, то на длины квадратов, фиксирующих расположенные рядом позиции, наложены значительные ограничения. Если рассмотреть плоскость с системой координат, где ось абсцисс — позиции слова, а ось ординат — периоды фиксирующих эти позиции квадратов, то каждая точка (i, p), соответствующая квадрату с периодом p, фиксирующему позицию i, порождает справа от себя многоугольник запрещённых пар (позиция, период), границы которого зависят только от p. Для того, чтобы доказать верхнюю оценку доли позиций, зафиксированных длинными квадратами, необходимо в фиксированном контексте произвольного слова расположить позиции, зафиксированные длинными квадратами, как можно «плотнее». Оказывается, что эту задачу можно свести к варианту задачи составления расписаний, в котором жадный алгоритм даёт оптимальное решение. Используя это решение, удается доказать верхнюю оценку $l/9 + O(\log n)$ на количество позиций, фиксированных длинными квадратами в фиксированном контексте длины l слова длины n, причём показано, что $l \leqslant n$.

Объединяя результаты, полученные для коротких и длинных квадратов, получаем утверждение Теоремы 14.

В разделе 3.2 приведено доказательство Теоремы 15. Это следствие из доказательства Теоремы 14, где при суммировании долей букв, фиксированных короткими и длинными квадратами, остаётся «зазор»: грубо говоря, $2/3$ букв фиксировано короткими квадратами и $1/9$ — длинными. Отсюда имеем нижнюю оценку на количество ветвящихся узлов в дереве тернарного бесквадратного языка — $2/9$.

Наконец, в разделе 3.3 обсуждается возможное усиление Теоремы 14.

В заключении кратко подведятся итоги работы отдельно по каждой главе диссертации и обсуждаются пути дальнейших исследований.

Положения, выносимые на защиту

Основными результатами диссертации являются решения сформулированных выше задач 1-3.

1. Доказано существование конечных поддеревьев произвольной высоты в суффиксном и префиксном деревьях, а также конечных подграфов специального вида в ациклическом подсловном графе языков бинарных бескубных слов CF и тернарных бесквадратных слов SF. Доказательство произведено конструктивным образом путём построения соответствующих серий предмаксимальных слева (справа) слов для любого натурального уровня n и двусторонних предмаксимальных слов для любого уровня $(n, k) \in \mathbb{N}^2$. Тем самым решена открытая проблема, сформулированная Рестиво и Салеми в [61] и Аллушем и Шаллитом в [5]. Результаты опубликованы в [52–56].

2. Доказана логарифмическая оценка на длину фиксированного контекста слова w в зависимости от $|w|$ в случае тернарного бесквадратного языка SF, которая является асимптотически неулучшаемой. Это является существенным усилением результата, полученного Сони и Шелтоном в [70]. Как следствие получена нижняя оценка на частоту ветвления префиксного (суффиксного) дерева тернарного бесквадратного языка — $2/9$, на основании которой, учитывая индекс роста языка SF, можно сделать вывод, что дерево языка SF устроено достаточно регулярно. Результат опубликован в [57].

3. Доказано, что любой буквенный шаблон длины 4 и менее неизбежен в тернарном бесквадратном языке. Для длин 5 и 6 дана классификация по избегаемости, а также предложена идея решения этой задачи для семибуквенных шаблонов, где требуется рассмотреть единственный шаблон для завершения классификации буквенных шаблонов в смысле избегаемости в целом для языка SF. Эта задача является логическим продолжением и обобщением результатов, полученных Туэ в [80]. Результат опубликован в [51].

Научная новизна

Все результаты диссертации являются новыми.

Теоретическая и практическая ценность

Диссертация носит теоретический характер. Полученные результаты могут быть использованы в дальнейших исследованиях по комбинаторике слов и при чтении специальных курсов для студентов математических специальностей.

Апробация результатов работы

Результаты диссертации были представлены на 8-й и 10-й международных конференциях по комбинаторике слов WORDS2011 и WORDS2015 (Прага, 2011; Киль, 2015), 1-м и 2-м российско-финских симпозиумах по дискретной математике (Санкт-Петербург, 2011; Турку, 2012), 37-й международной конференции по математическим основам компьютерных наук MFCS 2012 (Братислава, 2012), а также на семинарах «Алгебраические системы» и «Дискретная математика» (УрГУ/УрФУ, 2010-2015).

Публикации

Основные результаты диссертации опубликованы в работах [51–57], среди которых четыре работы [51, 54, 55, 57] опубликованы в рецензируемых изданиях из списка, рекомендованного ВАК. В совместных работах [52, 54, 55, 57] А.М. Шуру принадлежат постановка задачи, общая методика исследований и оптимизация некоторых предложенных автором конструкций. Доказательства всех основных утверждений принадлежат автору. Работы [53, 56] представляют собой тезисы статей [52, 55].

Глава 1

Предмаксимальные слова

В данной главе пойдёт речь о построении серий предмаксимальных слева слов уровня n для любого натурального n и предмаксимальных слов уровня (n, k) для любой пары $(n, k) \in \mathbb{N}^2$ в бинарном бескубном языке и тернарном бесквадратном языке. Иными словами, будет конструктивным образом доказано существование конечных поддеревьев и конечных ациклических подграфов произвольной высоты в соответствующих графах этих языков. Основными результатами Главы 1 являются следующие утверждения.

Теорема 10 *В языке* CF *существуют*
а) предмаксимальные слева слова любого уровня $n \in \mathbb{N}_0$*;*
б) предмаксимальные слова любого уровня $(n, k) \in \mathbb{N}_0^2$.

Теорема 11 *В языке* SF *существуют*
а) предмаксимальные слева слова любого уровня $n \in \mathbb{N}_0$*;*
б) предмаксимальные слова любого уровня $(n, k) \in \mathbb{N}_0^2$.

В первом разделе приводится общая схема построения для обоих языков. Во втором и третьем разделах рассматриваются конкретные техники применения общей схемы для бинарного бескубного и тернарного бесквадратного языков, соответственно.

1.1 Общая схема построения предмаксимальных слов

Мы доказываем утверждения а) теорем 10 и 11 построением последовательностей предмаксимальных слева слов, содержащих слова любого конечного уровня. Последовательности строятся в два этапа:

1) построение вспомогательной последовательности $\{w_n\}_0^\infty$ такой, что для всех n и для всех $k \leqslant n$ слово w_n имеет единственный левый контекст длины k;

2) достройка слова w_n до предмаксимального слева слова $\overline{w}_n$.

Для доказательства утверждений б) этих теорем мы соединяем слово $\overline{w}_n$ с реверсом слова $\overline{w}_k$ с помощью некоторого «буферного» слова.

Пусть слово w из соответствующего языка (CF или SF) имеет единственный левый контекст длины n, скажем, u, и два левых контекста[1] длины $n{+}1$. Тогда u – наидлиннейший фиксированный левый контекст w. Поясним содержание первого этапа. Мы строим последовательность $\{w_n\}_0^\infty$ индуктивно. Фиксированный левый контекст u_n слова w_n имеет длину $\geqslant n$. Потребуем, чтобы каждое слово w_n, $n > 0$, обладало следующими свойствами:

1. w_{n-1} – префикс w_n;

2. любой суффикс некоторого фиксированного ω^*-слова U является левым контекстом w_n;

3. наидлиннейший фиксированный левый контекст u_n слова w_n равен $\mathsf{U}[k..1]$ для некоторого $k \geqslant n$;

4. если $|u_n| > n$, то $w_{n+1} = w_n$ (*тривиальные* итерации).

Если $(n{+}1)$-ая итерация нетривиальна, то фиксированный контекст u_{n+1} длиннее, чем $u_n = \mathsf{U}[n..1]$. Следовательно, если $u_n[1] = x$, то оба слова yu_n и zu_n являются левыми контекстами w_n (в случае бинарного алфавита одна из букв y и z совпадает с x). Значит, нам нужно продолжить w_n вправо так, чтобы «запретить» появление некоторой буквы перед u_n в левом контексте. Пусть $yu_n = \mathsf{U}[n{+}1..1]$. Тогда мы запрещаем z на $(n{+}1)$-ой итерации. Основная идея состоит в том, чтобы «размножить» предыдущее слово вместе с его запрещаемым контекстом на соответствующее количество копий (две для бесквадратного языка, три для бескубного), то есть, говоря иначе, возвести zu_nw_n в запрещаемую степень, и в качестве w_{n+1} взять суффикс полученного слова длины $|w_{n+1}| - (n{+}1)$, удалив, тем самым, первое вхождение zu_n. Более точно, положить

$$w_{n+1} = \underbrace{\qquad w_n}\underbrace{zu_nw_n}\underbrace{zu_nw_n}, \tag{1.1a}$$

$$w_{n+1} = \underbrace{\qquad w_n}\underbrace{zu_nw_n}, \tag{1.1b}$$

[1] Очевидно, что трёх левых контекстов длины $n{+}1$ у слова w из языка SF быть не может, поскольку в любом слове из этого языка любые две соседние буквы различны

где схема 1.1a предназначена для случая бинарного бескубного языка, а 1.1b — для тернарного бесквадратного языка. Видно, что zu_n не может быть левым контекстом w_{n+1}, а значит, $|u_{n+1}| > |u_n|$.

Замечание 1 *Попытка построить последовательность $\{w_n\}_0^\infty$ напрямую по схеме (1.1) приводит к неудаче, так как возникают локальные кубы (1.1a) и квадраты (1.1b) на стыке некоторых w_n и zu_n.*

Попытаемся решить эту проблему следующим образом. Будем к каждой копии w_n присоединять справа некоторое «буферное» слово s_n. Таким образом, схема 1.1 примет вид:

$$w_{n+1} = \underbrace{\quad w_n s_n}\underbrace{z u_n w_n s_n}\underbrace{z u_n w_n s_n}, \tag{1.2a}$$

$$w_{n+1} = \underbrace{\quad w_n s_n}\underbrace{z u_n w_n s_n}. \tag{1.2b}$$

Далее для обоих языков будет представлено конструктивное доказательство существования таких последовательностей $\{s_n\}_0^\infty$, что слова w_n для любого натурального n соответственно бескубны (1.2a) или бесквадратны (1.2b). Второй этап построения последовательностей предмаксимальных слева слов опирается на ту же идею, что и первый. Для построения предмаксимального слева слова уровня n возьмём слово w_{n+1} с наидлиннейшим фиксированным левым контекстом длины $n+1$ и запретим появление самой левой буквы, положив

$$\overline{w}_n = \underbrace{\quad w_{n+1}\overline{s}_n}\underbrace{y u_n w_{n+1}\overline{s}_n}\underbrace{y u_n w_{n+1}\overline{s}_n}, \tag{1.3a}$$

$$\overline{w}_n = \underbrace{\quad w_{n+1}\overline{s}_n}\underbrace{y u_n w_{n+1}\overline{s}_n}. \tag{1.3b}$$

Аналогично первому этапу, специальные слова $\overline{s}_n$ используются для предотвращения появления локальных запрещённых степеней. Допустим, что мы нашли такую последовательность слов $\{\overline{s}_n\}_0^\infty$, что для любого натурального n $\overline{w}_n$ соответственно бескубно (1.3a) или бесквадратно (1.3b). Значит, в итоге мы получили, что любое слово $\overline{w}_n$ имеет однозначное продолжение влево длиной n букв, а любая буква в $(n+1)$-й позиции слева приведёт к образованию куба или квадрата соответственно. Для случая бинарного языка мы запретили обе возможные буквы, для тернарного мы запретили обе буквы, отличные от n-й буквы фиксированного контекста. Это и означает, что $\overline{w}_n$ — предмаксимальное слева слово уровня n.

Замечание 2 *Для любого n слово $\overleftarrow{\overline{w}_n} = \overline{w}_n(|\overline{w}_n|)\cdots\overline{w}_n(1)$ является предмаксимальным справа словом уровня n.*

Для получения двустороннего предмаксимального слова уровня (n, k) используем естественную идею: возьмём предмаксимальное слева слово уровня n, реверс предмаксимального слева слова уровня k и соединим их опять же при помощи специального «буфера».

Задача подбора слов U, построения последовательностей $\{s_n\}_0^\infty$, $\{\overline{s}_n\}_0^\infty$ и подбора буферных слов для двустороннего случая таких, чтобы слова w_n, $\overline{w}_n$ и двусторонние слова не содержали запрещённых степеней, решается для бинарного бескубного и тернарного бесквадратного языка применением разных техник, и далее эти случаи будут рассмотрены по отдельности.

1.2 Бинарный бескубный язык

В качестве слова U в случае построения серии предмаксимальных слов произвольного конечного уровня в бинарном бескубном языке возьмём реверс слова Туэ–Морса ${}_\infty^a\mathsf{T}$ без первой буквы. Положим $w_0 = aabaaba$ и заметим, что его наидлиннейший фиксированный левый контекст равен abb (первая буква ${}_\infty^a\mathsf{T}$ на самом деле «спрятана» в первой букве w_0). Легко убедиться, что ${}_\infty^a\mathsf{T}$ является левым контекстом $abaaba$ в CF, так как само слово ${}_\infty^a\mathsf{T}$ сильно бескубно. Буквы слова ${}_\infty^a\mathsf{T}$ будем нумеровать начиная с 0.

Для слов Туэ–Морса справедлива следующая лемма:

Лемма 1 *Буквы* ${}_\infty^a\mathsf{T}[n]$ *и* ${}_\infty^a\mathsf{T}[n-1]$ *совпадают тогда и только тогда, когда* $n = m \cdot 2^k$, *где* m *и* k *- нечетные натуральные числа.*

Доказательство. Из определения морфизма и слов Туэ–Морса следует, что буквы двух соседних n-блоков совпадают тогда и только тогда, когда n нечётно, при этом длина n-блока равна 2^n. □

Замечание 3 *Если единственный левый контекст длины* n *слова* w_n *начинается с* xx, *то* $|u_n| > n$, *т.к. буква перед* xx *также фиксирована. Следовательно, по лемме 1* $(n+1)$*-ая итерация тривиальна при* $n = m \cdot 2^k, n \geqslant 3$ *для нечетных* k, m. *Далее будет доказано, что во всех остальных случаях можно произвести нетривиальную итерацию.*

Пусть xu_na — суффикс слова ${}_\infty^a\mathsf{T}$ (то есть, фиксированный левый контекст слова w_{n+1}), тогда обозначим $p_n = xu_n, p'_n = \bar{x}u_n$.

Структура слова w_{n+1} обусловливает следующие ограничения на слова s_n и s_{n+1} (или s_{n+2}, если $s_{n+1} = \lambda$):

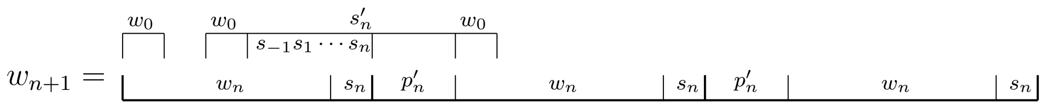
(s1) Т.к. $u_{n+1}w_{n+1}s_{n+1}$ является подсловом w_{n+2}, а u_{n+1} кончается на u_n и $u_nw_{n+1}x = (x_nw_ns_nx)^3$ по (1.2a), слово s_{n+1} должно начинаться с буквы x, которая является первой буквой p_n;

(s2) Т.к. $s_n x u_n$ - подслово w_{n+1}, получаем, что если u_n начинается с x ($\bar{x}x\bar{x}x$), то s_n кончается на $\bar{x}$ (x, соответственно).

Т.к. $u_n \in \mathsf{TM}$ — сильно бескубное слово, любой другой префикс u_n не определяет однозначно последнюю букву s_n. Будем называть итерацию *ограничивающей*, если на ней срабатывает ограничение (s2).

1.2.1 Построение буферных слов

Наша первая задача — найти последовательность s_n, удовлетворяющую ограничениям (s1) и (s2) такую, что любое слово вида $s_i s_{i+1} \cdots s_{i+j}$ бескубно. Из технических соображений удобно положить $w_3 = w_0 s_{-1} s_1$ ($s_0 = s_2 = \lambda$ по Замечанию 3). Следовательно, слово $s'_n = s_{-1}s_0 \cdots s_n$ входит в w_{n+1} между w_0 и следующим вхождением p'_n.



Т.к. n может быть сколь угодно большим, требуется построить бескубное ω-слово $s'_\infty = s_{-1}s_0 \cdots s_n \cdots$. При построении s'_∞ мы придерживаемся следующих правил:

1. для всех нетривиальных итераций $s_n \in \{\mathsf{T}_2^x, \mathsf{T}_2^x\mathsf{T}_2^x, \mathsf{T}_4^x, \mathsf{T}_2^x\mathsf{T}_2^{\bar{x}}\mathsf{T}_1^x, \mathsf{T}_1^x, \mathsf{T}_1^x\mathsf{T}_2^{\bar{x}}\}$; отсюда $s_n \in \mathsf{TM}$;

2. везде, где это возможно, полагаем s_n равным 2-блоку или произведению 2-блоков;

3. иначе, если s_n кончается на T_1^x, полагаем $s_{n+1} = \mathsf{T}_1^{\bar{x}}$ или $s_{n+1} = \mathsf{T}_1^{\bar{x}}\mathsf{T}_2^x$ (то же самое для s_{n+2}, когда $s_{n+1} = \lambda$);

4. если $s_n \neq \lambda$ и n-ая итерация не является ограничивающей, вводим искусственный запрет на последнюю букву s_n. А именно, если s_{n-1} кончается на x, то s_n должно заканчиваться на $\bar{x}$ (то же самое, если s_{n-2} кончается на x, а $s_{n-1} = \lambda$).

С учётом вышесказанного справедлива следующая

Лемма 2 (1) *Если s_n кончается на x, то либо s_{n+1} кончается на $\bar{x}$, либо $s_{n+1} = \lambda$ и s_{n+2} кончается на $\bar{x}$.*
(2) *Первая буква непустого слова s_n совпадает с последней для всех n, за исключением случая когда $p_n = x\bar{x}x\bar{x}\cdots$ или $p_n = xx\bar{x}x\cdots$.*

Таблица 1.1: Слова s_n для 32 последовательных итераций, начинающихся с некоторого номера k, кратного 32. Правая (левая) часть таблицы применяется, когда буквы $\mathsf{T}^b_\infty[k/32]$ и $\mathsf{T}^b_\infty[k/32{+}1]$ совпадают (различны). Строки, соответствующие тривиальным итерациям, опущены. Второе значение s_{k+29} (соответствующее итерации $k+30$) в левой части используется тогда и только тогда, когда $\mathsf{T}^b_\infty[k/32{+}1] = \mathsf{T}^b_\infty[k/32{+}2]$.

№ итерации.	Запреты		
(n)	Начало	Конец	s_{n-1}
k	$\bar{x}$	$\bar{x}$	T_2^x
$k+1$	тривиальная		
$k+2$	x	x	$\mathsf{T}_2^{\bar{x}}\mathsf{T}_2^{\bar{x}}$
$k+4$	$\bar{x}$	$\bar{x}$	T_2^x
$k+5$	$\bar{x}$	$x, \mathsf{T}_2^{\bar{x}}$	$\mathsf{T}_2^x\mathsf{T}_2^{\bar{x}}\mathsf{T}_1^x$
$k+6$	x	$\bar{x}$	$\mathsf{T}_1^{\bar{x}}$
$k+8$	x	x	$\mathsf{T}_2^{\bar{x}}$
$k+10$	$\bar{x}$	$\bar{x}$	$\mathsf{T}_2^x\mathsf{T}_2^x$
$k+12$	x	x	$\mathsf{T}_2^{\bar{x}}$
$k+13$	x	$\bar{x}, \mathsf{T}_2^x$	$\mathsf{T}_2^{\bar{x}}\mathsf{T}_2^x\mathsf{T}_1^{\bar{x}}$
$k+14$	$\bar{x}$	x	T_1^x
$k+16$	$\bar{x}$	$\bar{x}$	T_2^x
$k+17$	$\bar{x}$	x	T_1^x
$k+18$	$xx\bar{x}$	$\bar{x}$	$\mathsf{T}_1^{\bar{x}}$
$k+20$	$\bar{x}\bar{x}x$	x	$\mathsf{T}_2^{\bar{x}}$
$k+21$	x	$\bar{x}$	$\mathsf{T}_1^{\bar{x}}$
$k+22$	$\bar{x}\bar{x}x$	x	T_1^x
$k+24$	$xx\bar{x}$	$\bar{x}$	T_2^x
$k+26$	x	x	$\mathsf{T}_2^{\bar{x}}$
$k+28$	$\bar{x}$	$\bar{x}$	T_4^x
$k+29$	$\bar{x}$	$x, \mathsf{T}_2^{\bar{x}}$	$\mathsf{T}_2^x\mathsf{T}_2^{\bar{x}}\mathsf{T}_1^x$
$k+30$	x	$\bar{x}$	$\mathsf{T}_1^{\bar{x}}/\mathsf{T}_1^{\bar{x}}\mathsf{T}_2^x$

№ итерации.	Запреты		
(n)	Начало	Конец	s_{n-1}
k	x	x	$\mathsf{T}_2^{\bar{x}}$
$k+1$	x	$\bar{x}$	$\mathsf{T}_1^{\bar{x}}$
$k+2$	$\bar{x}\bar{x}x$	x	T_1^x
$k+4$	$xx\bar{x}$	$\bar{x}$	T_2^x
$k+5$	$\bar{x}$	x	T_1^x
$k+6$	$xx\bar{x}$	$\bar{x}$	$\mathsf{T}_1^{\bar{x}}$
$k+8$	$\bar{x}\bar{x}x$	x	$\mathsf{T}_2^{\bar{x}}$
$k+10$	$\bar{x}$	$\bar{x}$	T_2^x
$k+12$	x	x	$\mathsf{T}_4^{\bar{x}}$
$k+13$	x	$\bar{x}, \mathsf{T}_2^x$	$\mathsf{T}_2^{\bar{x}}\mathsf{T}_2^x\mathsf{T}_1^{\bar{x}}$
$k+14$	$\bar{x}$	x	T_1^x
$k+16$	$\bar{x}$	$\bar{x}$	T_2^x
$k+17$	$\bar{x}$	x	T_1^x
$k+18$	$xx\bar{x}$	$\bar{x}$	$\mathsf{T}_1^{\bar{x}}$
$k+20$	$\bar{x}\bar{x}x$	x	$\mathsf{T}_2^{\bar{x}}$
$k+21$	x	$\bar{x}$	$\mathsf{T}_1^{\bar{x}}$
$k+22$	$\bar{x}\bar{x}x$	x	T_1^x
$k+24$	$xx\bar{x}$	$\bar{x}$	T_2^x
$k+26$	x	x	$\mathsf{T}_2^{\bar{x}}$
$k+28$	$\bar{x}$	$\bar{x}$	T_4^x
$k+29$	$\bar{x}$	$x, \mathsf{T}_2^{\bar{x}}$	$\mathsf{T}_2^x\mathsf{T}_2^{\bar{x}}\mathsf{T}_1^x$
$k+30$	x	$\bar{x}$	$\mathsf{T}_1^{\bar{x}}$

Значения для слов s_n приведены в таблице 1.1. Непосредственно проверяется, что эти значения удовлетворяют правилам 1–4 и лемме 2. Согласно таблице 1.1, правило 3 применяется к s_n в том и только в том случае, когда p_n начинается с $x\bar{x}x\bar{x}$. Слово p_n сильно бескубно, значит, если оно имеет такой префикс, то p_{n-1} (или p_{n-2}, если $(n{-}1)$-ая итерация тривиальна) не имеет таких префиксов. В результате слово s_{n-1} (соответственно, s_{n-2}) кончается на 2-блок.

Теперь рассмотрим случай $p_n = x\bar{x}x\bar{x}\cdots$ более подробно. Без ограничения общности, пусть p_n начинается с b. Тогда $p_n = babaab\cdots$. Т.к. $p'_n = aabaab\cdots$, слово s_n не может оканчиваться на a или $baab$, значит, мы не можем закончить его на 2-блок и оно должно равняться одному из следующих слов: T_1^x, $\mathsf{T}_2^x\mathsf{T}_2^{\bar{x}}\mathsf{T}_1^x$.

Так как p_n — подслово слова ${}_\infty^a\mathsf{T}$, которое представляет собой бесконечное произведение 2-блоков, некоторый 2-блок T_2^a этого произведения кончается во второй позиции слова p_n. Сначала рассмотрим следующий случай вхождения p_n в ${}_\infty^a\mathsf{T}$:

$$ {}_\infty^a\mathsf{T} = \cdots \overbrace{abba}^{\mathsf{T}_2^a}\overbrace{ab\,ba}^{\mathsf{T}_2^a}\underbrace{\overbrace{baab}^{\mathsf{T}_2^b}\overbrace{baab}^{\mathsf{T}_2^b}\cdots}_{p_n} \tag{1.4}$$

Т.к. $p'_{n-1} = bbaab\cdots$, слово s_{n-1} кончается на $abba$. Следовательно, мы не можем положить $s_n = ab$ (иначе у s_n появится суффикс $baab$). Далее, p_{n-1} начинается с $abaab$, откуда $s_n[1] = a$ по (s1). Значит, согласно правилу 1, остается единственный вариант — положить s_n равным $\mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_1^a = abbabaabab$. Легко понять, что $s_{n+1} = ba$ удовлетворяет обоим условиям (s1) и (s2).

Если последний 2-блок под фигурной скобкой в (1.4) равен T_2^a, а не T_2^b, мы имеем следующую ситуацию с точностью до переименования букв:

$$ {}_\infty^a\mathsf{T} = \cdots \overbrace{baab}^{\mathsf{T}_2^b}\overbrace{ab\,ba}^{\mathsf{T}_2^a}\underbrace{\overbrace{baab}^{\mathsf{T}_2^b}\cdots}_{p_n} $$

Как и в предыдущем случае, положим $s_n = \mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_1^a$ и $s_{n+1} = \mathsf{T}_1^b$. Проблема возникает на $(n{+}5)$-ой итерации, так как

$$ p'_{n+4} = b\,bab\,bab\,aab\cdots, $$

т. е., s_{n+4} не может заканчиваться ни на ba, ни на ab. Здесь мы отступим от общего правила и используем следующий прием. На последующих трех итерациях (с $(n{+}5)$-ой по $(n{+}7)$-ю, последняя из которых тривиальна) мы должны добавить baa к фиксированному контексту. Мы сделаем это, последовательно запрещая трехбуквенные контексты вместо одной буквы. Слово $p_{n+3} = babbaba\cdots$ имеет три левых контекста длины 3: aab, baa и bba. Запретим bba на $(n{+}5)$-ой итерации и aab на $(n{+}6)$-ой. Для этого положим $P'_{n+4} = bba\,babbabaab\cdots$, $P'_{n+5} = aab\,babbabaab\cdots$. Это позволяет положить $s_{n+4} = ba, s_{n+5} = ab$. Такие итерации назовем *исключительными*.

Замечание 4 *Использование исключительных итераций приводит к одному локальному нарушению общего правила для u_n. А именно, $|u_{n+5}| = n{+}4$ (это слово совпадает с u_{n+4}). Ситуация исправляется на следующей итерации, где мы получаем $|u_{n+6}| = n{+}7$, а $(n{+}7)$-ая итерация тривиальна.*

Замечание 5 *Слово $\mathsf{T}_2^a\mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_2^a\mathsf{T}_2^a = \theta^2(aabaa)$ не является подсловом ${}_\infty^a\mathsf{T}$. Следовательно, подслово $\mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_2^a$ входит в ${}_\infty^a\mathsf{T}$ внутри подслов $\mathsf{T}_2^b\mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_2^a$*

или $\mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_2^a\mathsf{T}_2^b$. *Для каждого такого подслова мы используем две пары исключительных итераций.*

Напомним, что $s'_n = s_{-1}s_0s_1\dots s_n$. Рассмотрим равноблочный морфизм $\psi : \Sigma_2^+ \to \Sigma_2^+$ с длиной блока, равной 108, определяемый равенствами

$$\psi(a) = \mathsf{T}_4^a\mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_2^a\mathsf{T}_4^b\mathsf{T}_2^b\mathsf{T}_2^a\mathsf{T}_4^b\mathsf{T}_2^b\mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_4^a\mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_2^a, \tag{1.5a}$$

$$\psi(b) = \mathsf{T}_4^b\mathsf{T}_2^b\mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_4^a\mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_4^a\mathsf{T}_2^a\mathsf{T}_2^b\mathsf{T}_2^a\mathsf{T}_4^b\mathsf{T}_2^b\mathsf{T}_2^a\mathsf{T}_2^b. \tag{1.5b}$$

Заметим, что $\psi(b) = \overline{\psi(a)}$. Слово $\psi(aabbababbabbaabaababaabb)$ бескубно, как было проверено с помощью компьютера. Значит, по теореме 6, ψ — бескубный морфизм и слово $\psi(\mathsf{T}_\infty^b)$ бескубно. Согласно таблице 1.1, если $\mathsf{T}_\infty^b[j] \neq \mathsf{T}_\infty^b[j{+}1] \neq \mathsf{T}_\infty^b[j{+}2]$, то выполняется равенство $\psi(\mathsf{T}_\infty^b[j]) = s_{32j-1}s_{32j}\cdots s_{32j+30}$, а если $\mathsf{T}_\infty^b[j] \neq \mathsf{T}_\infty^b[j{+}1] = \mathsf{T}_\infty^b[j{+}2]$, то $\psi(\mathsf{T}_\infty^b[j\dots j{+}1]) = s_{32j-1}s_{32j}\cdots s_{32j+62}$. Следовательно, $\psi(\mathsf{T}_\infty^b) = s'_\infty$, и любое произведение последовательных “буферных” слов бескубно, что и требовалось. Заметим, что ψ-образ одной буквы соответствует (кроме одного исключения, связанного с определенным значением s_{k+29}) 32 последовательным итерациям, на протяжении которых к фиксированному контексту добавляется 5-блок. Таким образом, мы определили слова s_n, а значит, и слова w_n для всех натуральных n.

1.2.2 Доказательство корректности конструкции

Корректность полученной конструкции опирается на следующую основную лемму:

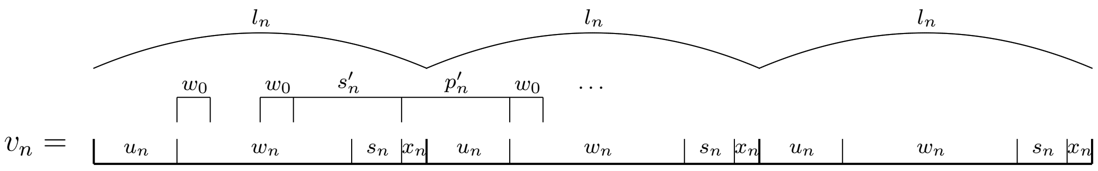
Лемма 3 *Слово u_nw_n бескубно для любого $n \in \mathbb{N}_0$.*

Доказательство. Докажем по индукции, что все слова $v_n = (u_nw_ns_nx_n)^3$, где x_n — буква, запрещаемая на $(n{+}1)$-й итерации, не имеют собственных подслов, являющихся кубами. Из этого будет немедленно следовать утверждение леммы. Естественно, что достаточно рассмотреть только случай, когда $(n{+}1)$-я итерация нетривиальна. База индукции $n \leq 4$ может быть легко проверена вручную или с помощью компьютера. Докажем шаг индукции. Рисунок 1.1 иллюстрирует структуру слова v_n.

Проведём доказательство от противного. Предположим, что некоторое слово v_n, где $n \geq 5$, содержит некоторый куб t^3, который, в свою очередь, не имеет собственных подслов–кубов.
Утверждение 1. Слово u_nw_n не является подсловом t^3.

Доказательство Подслово t^3 в v_n имеет периоды $q = |t|$ и $l_n = |v_n|/3$, но не удовлетворяет свойству взаимодействия периодов (иначе t являлось бы

Рис. 1.1: Структура слова $v_n = (u_n w_n s_n x_n)^3$



степенью более короткого слова, а значит, t^3 содержало бы другие кубы внутри). Следовательно, $|t^3| = 3q \leqslant q + l_n - 2$ по теореме Файна–Вильфа, откуда $q \leqslant l_n/2 - 1$. С другой стороны, по определению w_n самый длинный собственный суффикс слова $u_n w_n$ совпадает с самым длинным собственным префиксом слова v_{n-1} (обозначим его как v'). Если t^3 содержит v', то v' имеет периоды q и $l_{n-1} = |v_{n-1}|/3$. Применяя теорему Файна–Вильфа снова, получаем $l_{n-1} \leqslant q/2 - 1$. Исключая q из двух полученных неравенств, имеем $l_n \geqslant 4l_{n-1} + 3$. Но $l_n = |v_{n-1}| + |s_n| + 1 \leqslant 3l_{n-1} + 17$. Следовательно, $l_{n-1} \leq 14$. Для $n \geq 5$ такое неравенство не выполняется. Таким образом, заключаем, что t^3 не содержит слова $u_n w_n$.

Утверждение 2. Слово s'_n встречается в v_n только три раза.

Доказательство Слово s'_n входит в блок $u_n w_n s_n x_n$ непосредственно перед его последней буквой (см. Рис. 1.1). Слово v_n содержит три таких «регулярных» вхождения. Покажем, что других вхождений нет. Напомним, что s'_n является произведением 2-блоков (за исключением, возможно, последнего «непарного» 1-блока), и при $n \geq 5$ s'_n начинается с 4-блока. Значит, s'_n не содержит подслова w_0 и, более того, оно не может начинаться внутри w_0. Далее, вручную или с помощью компьютера можно проверить, что s'_∞ не содержит подслов Туэ–Морса длины >48. Теперь, глядя на структуры слов s'_n и v_n можно сделать вывод, что любое «нерегулярное» вхождение s'_n в v_n должно быть префиксом некоторого слова $s'_k p'_k w_0$, где $k < n$. Слово s'_k — собственный префикс s'_n. Слово p'_k получено из подслова Туэ–Морса инверсией первой буквы, значит, оно не может начинаться с 2-блока. Следовательно, $k = n - 1$, и s_n должен быть 1-блоком, равным префиксу p'_k (во всех остальных случаях слово $s_{k+1} \cdots s_n$ начинается с 2-блока). Согласно Таблице 1.1 во всех случаях, когда s_n представляет из себя 1-блок, p'_{n-1} начинается с квадрата буквы, значит, и такой вариант невозможен.

Утверждение 3. Слово $u_n w_n s_n x_n$ бескубно.

Доказательство Слово v_{n-1} бескубно по предположению индукции. Слово $u_n w_n$ получается из v_{n-1} удалением последней буквы и присоединением первой буквы слова u_n к началу, при этом $u_n(1) \neq x_{n-1}$, а значит, $u_n w_n$ бескубно. Используя снова тот факт, что s'_n — «почти» произведение 2-блоков, заключаем, что $s'_n x_n$ тоже бескубно. Значит, если в $u_n w_n s_n x_n$ есть куб в качестве подслова, то он содержит внутри себя суффикс s'_{n-1} слова w_n. Этому суффиксу предшествует $w_0 = aabaaba$ (см. Рис. 1.1); которое обрывает все периоды s'_{n-1} и не порождает куба. Значит, предполагаемый куб должен содержать более одного вхождения подслова s'_{n-1}. Применяя Утверждение 1 к словам s'_{n-1} и v_{n-1}, видим, что куб имеет период $l_{n-1} = (|u_n w_n|+1)/3$. Но это невозможно по условию (s1). Утверждение доказано.

Соединяя Утверждения 3 и 1, получаем, что t^3 содержится строго внутри $u_n w_n s_n x_n u_n w_n$, причём кончается строго позже x_n. Далее, если s'_n — подслово t^3, то среднее вхождение t в t^3 находится внутри s'_n (в противном случае t^3 содержит ещё одно вхождение s'_n, что противоречит Утверждению 2). В этом случае все позиции подслов aa и bb в t имеют одинаковую чётность, т.к. t — подслово произведения 1-блоков. Но самое правое вхождение t в t^3 содержит суффикс слова s'_n и префикс слова $x_n u_n = p'_n$ следом за ним. Буква x_n нарушает эту чётность позиций (u_n также является произведением 1-блоков). Более того, s'_n за исключением, возможно, двух последних букв допускает разбиение на 2-блоки; значит, все подслова aa и bb в правом вхождении t находятся по одну сторону от x_n. Заметим, что никакой 2-блок не является префиксом $x_n u_n$.

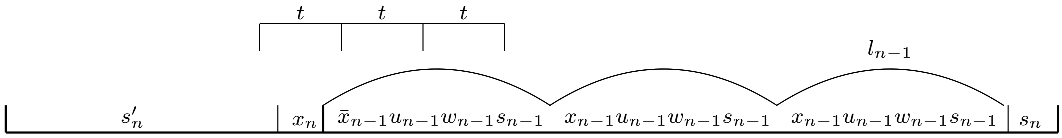
Случай 1. Все подслова aa и bb в t^3 находятся слева от x_n. Тогда p'_n начинается на некоторый 1-блок. Согласно Таблице 1.1, в этом случае этот 1-блок и последние 2 буквы слова s'_n образуют 2-блок, при этом само s'_n кончается на 2-блок. Значит, s'_n содержит подслово $\mathsf{T}_2^x 2\mathsf{T}_1^{\bar{x}}\mathsf{T}_2^x$, что невозможно по определению морфизма ψ.

Случай 2. Все подслова aa и bb в t^3 находятся справа от x_n. В таком случае подслово, находящееся на границе s'_n и $x_n u_n$ повторяется строго внутри s'_n и нарушает разбиение s'_n на 2-блоки, снова получили противоречие.

Таким образом, заключаем, что s'_n не является подсловом t^3. Отсюда, t^3 начинается внутри подслова $s'_n x_n$. А где оно кончается? Легко видеть, что слово

$$u_n w_n = \bar{x}_{n-1} u_{n-1} w_{n-1} s_{n-1} x_{n-1} u_{n-1} w_{n-1} s_{n-1} x_{n-1} u_{n-1} w_{n-1} s_{n-1}$$

имеет те же три вхождения s'_{n-1}, что и v_{n-1}. Значит, если t^3 содержит s'_{n-1}, то второе вхождение t находится внутри s'_{n-1}. Но это невозможно, поскольку s'_{n-1} является достаточно коротким суффиксом $w_{n-1}s_{n-1}$ и все слово $u_n w_n$ бескубно. Следовательно, t^3 должно кончаться внутри префикса $\bar{x}_{n-1} u_{n-1} w_{n-1} s_{n-1}$ слова $u_n w_n$, как изображено на следующем рисунке.



Используя те же рассуждения о чётности позиций подслов aa и bb, что и выше, можно заключить, что $s'_n x_n u_n = s'_n p'_n$ бескубно и, более того, t^3 должно содержать префикс $aabaa$ слова w_{n-1}. Требуется рассмотреть два случая: либо $aabaa$ — подслово t, либо $aabaa$ встречается в t^3 только дважды, на границах смежных подслов t. Второй случай невозможен, поскольку два ближайших вхождения $aabaa$ в w_{n-1} разделены подсловом $babaababbaabbabaabaabb$, которое не содержит p'_n в качестве суффикса. В первом случае мы получаем, что некоторому (но не самому левому) вхождению $aabaa$ в t^3 предшествует конкатенация некоторого суффикса слова s'_n и слова p'_n. Если это вхождение $aabaa$ является префиксом некоторой копии w_0, то ему предшествует p'_k, $k < n$. Но p'_k не является суффиксом p'_n, получили противоречие. Оставшаяся возможная позиция вхождения $aabaa$ находится на границе некоторых слов s'_k и p'_k. Но в таком случае s'_k содержит подслово, находящееся на границе s'_n and p'_n, и условие чётности говорит, что s'_k не может быть представлено в виде произведения 2-блоков. Это заключительное противоречие доказывает, что t^3 не может быть подсловом v_n. Лемма доказана. □

По построению, слово u_n является фиксированным продолжением влево слова w_n. Рассмотрим теперь второй этап, а именно, дополнение такого «почти однозначно» продолжаемого влево слова w_n до предмаксимального слова.

Исходя из 1.3a, нам снова нужно найти последовательность «буферных слов» $\{\overline{s}_n\}_0^\infty$ такую, что для всех n слово $\overline{w}_n$ будет бескубным.

В отличие от первого этапа, конструкция (1.3) применяется только один раз, и нам не нужно строить бескубное ω-слово. Следовательно, слова $\overline{s}_n$ для разных n независимы друг от друга и вид $\overline{s}_n$ зависит только от последней итерации согласно Таблице 1.1. Эта зависимость представлена в Таблице 1.2. Мы подбираем $\overline{w}_n$ так, чтобы оно было левым контекстом p_n в ${}_\infty^a\mathsf{T}$ (напомним, что $p_n = {}_\infty^a\mathsf{T}(n+1 \dots 1)$).

Вышеизложенная идея даёт результат всегда, когда $|u_n| = n$. Благодаря следующему очевидному замечанию достаточно построить максимальные влево слова уровня n для всех n таких, что $|u_n| = n$. Это значит, что $(n+1)$-ая итерация, на которой строится слово w_{n+1}, не тривиальна и не исключительна. Итак, мы не рассматриваем построение слов $\overline{w}_n$ для других значений n.

Замечание 6 *Для доказательства теоремы 10,a достаточно показать существование предмаксимальных слов уровня n для бесконечного числа раз-*

Таблица 1.2: «Заключительные» суффиксы $\overline{s}_n$ для соответствующих итераций Таблицы 1.1. Первый столбец содержит номер итерации, k — любое число, кратное 32.

№ итерации. (n)	Запреты (Начало)	$\overline{s}_{n-1}$
k		
$k+1$	$\overline{x}$	$x\overline{x}$
$k+3$	x	$\overline{x}$
$k+4$	x	λ
$k+5$	$\overline{x}$	$x\overline{x}\overline{x}x$
$k+7$	$\overline{x}$	$x\overline{x}$
$k+9$	x	$\overline{x}x$
$k+11$	$\overline{x}$	x
$k+12$	$\overline{x}$	λ
$k+13$	x	λ
$k+15$	x	$\overline{x}$
$k+16$	x	λ
$k+18$	$xx\overline{x}$	$x\overline{x}$
$k+19$		
$k+20$	$\overline{x}$	λ
$k+23$	$\overline{x}\overline{x}x$	$\overline{x}x$
$k+25$	$\overline{x}$	$x\overline{x}$
$k+27$	x	$\overline{x}$
$k+28$	x	λ
$k+29$	$\overline{x}$	λ
$k+31$	$\overline{x}$	x

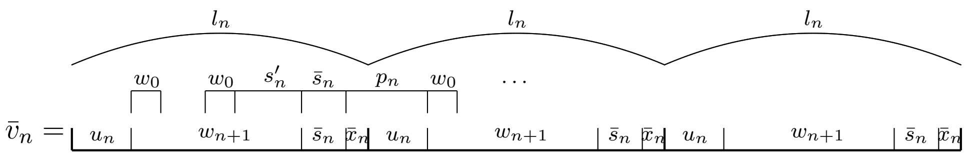
№ итерации. (n)	Запреты (Начало)	$\overline{s}_{n-1}$
k	$\overline{x}$	λ
$k+1$		
$k+3$	$\overline{x}\overline{x}x$	$\overline{x}$
$k+4$	x	λ
$k+5$		
$k+7$	$xx\overline{x}$	$x\overline{x}$
$k+9$	x	$\overline{x}x$
$k+11$	$\overline{x}$	x
$k+12$	$\overline{x}$	λ
$k+13$	x	λ
$k+15$	x	$\overline{x}$
$k+16$	x	λ
$k+18$		
$k+19$	$xx\overline{x}$	x
$k+20$	$\overline{x}$	λ
$k+23$	$\overline{x}\overline{x}x$	$\overline{x}x$
$k+25$	$\overline{x}$	$x\overline{x}$
$k+27$	x	$\overline{x}$
$k+28$	x	λ
$k+29$	$\overline{x}$	λ
$k+31$	$\overline{x}$	$x\overline{x}$

личных значений n. В самом деле, если w является предмаксимальным слева словом уровня n и $a_1 \cdots a_n w$ — максимальное слева слово, то слово $a_n w$ является предмаксимальным слева словом уровня $n{-}1$.

Применяя знания о том, что $w_{n+1} \in \mathsf{CF}$, $\overline{s}_n p_n \in \mathsf{TM}$, и суффикс s'_n слова w_{n+1} не имеет длинных подслов Туэ–Морса (это свойство присуще любому ψ-образу), можно доказать следующую лемму. Её доказательство подобно доказательству Леммы 3.

Лемма 4 *Слово $u_n \overline{w}_n$ бескубно для любого $n \in \mathbb{N}_0$.*

Рис. 1.2: Структура слова $\bar{v}_n = (u_n w_n \bar{s}_n \bar{x}_n)^3$



Доказательство. Докажем по индукции, что слова $\bar{v}_n = (u_n w_n \bar{s}_n \bar{x}_n)^3$, где $\bar{x}_n = p_n(1)$, не имеют собственных подслов, являющихся кубами. Из этого будет немедленно следовать утверждение леммы. База индукции $n \leq 4$ может быть проверена вручную или с помощью компьютера. Докажем шаг индукции.

Доказательство основано на положениях слова w_0 в $\bar{w}_n$. Все вхождения w_0 в w_{n+1} можно разделить на две группы: *регулярные* вхождения, которые появляются в результате итеративного применения конструкции (1.2), и *нерегулярные* вхождения, возникающие на границе некоторых слов s'_k и p'_k. Слово $\bar{w}_n$ также может содержать одно нерегулярное вхождение w_0 внутри каждого подслова $s'_n \bar{s}_n p_n$. Заметим, что за каждым регулярным вхождением w_0 находится подслово ba, а за нерегулярным — bb, так как слово Туэ–Морса не имеет подслов $ababa$.

Проведём доказательство от противного. Пусть $u_n \bar{w}_n$ содержит некоторый куб t^3.

Утверждение 1. Слово $p_n w_0 ba$ не является подсловом t^3.

Доказательство Каждому регулярному вхождению w_0 в w_{n+1} предшествует некоторое p'_k, где $k \leqslant n$. Ни одно из этих слов p'_k не является суффиксом p_n. Следовательно, в $\bar{v}_n$ слово $p_n w_0 ba$ встречается только два раза, а расстояние между этими вхождениями равно $l_n > |t|$, значит, t^3 не может включать в себя их оба. Пусть t^3 содержит одно вхождение $p_n w_0 ba$. Тогда среднее подслово t находится внутри $p_n w_0 ba$. Так как $|s'_n \bar{s}_n| > |p_n w_0 ba|$, слово t^3 в $u_n \bar{w}_n$ начинается внутри s'_n. Как следствие, t^3 содержит только одно регулярное вхождение w_0, а значит, только одно вхождение слова $w_0 ba$. Тогда t встречается внутри $w_0 ba$ и $w_0 ba = aabaababa$ имеет период $|t|$, что невозможно.

Утверждение 2. Слово t^3 содержит не более одного вхождения слова s'_n.

Доказательство В самом деле, в w_{n+1} слово s'_n встречается ровно три раза (см. Утверждение 2 в доказательстве Леммы 3), и расстояние между этими вхождениями равно периоду слова w_{n+1}. Этот период нарушается первой буквой слова $\bar{s}_n p_n$, значит, у t^3 такого периода быть не может. С другой стороны, t^3 не может содержать подслов s'_n из разных вхождений w_{n+1} в $\bar{v}_n$ согласно

Утверждению 1. Более того, так как слово $u_n w_{n+1}$ бескубно по Лемме 3, то единственное вхождение s'_n, которое может встретиться в t^3, — это суффикс w_{n+1}.

Допустим, что t^3 не содержит s'_n. Тогда среднее вхождение t в t^3 находится внутри s'_n, а значит, оно может быть разбито на 2-блоки за исключением, возможно, некоторых коротких начала и конца. Тогда t не имеет подслов $aabaa$ и $ababa$. Следовательно, t^3 может содержать не более двух последних букв слова w_0, предшествуюшего s'_n. Принимая во внимание Утверждение 1, получаем, что t^3 — подслово в слове $bas'_n\bar{s}_n p_n w_0 b$, в котором $aabaa$ встречается не более двух раз. Легко проверить, что $aabaa$ из w_0 не может находиться на границе двух копий t (иначе длина t была бы слишком мала). Следовательно, t^3 содержится в слове $r = bas'_n\bar{s}_n p_n aaba$.

Так как $p_n a$ является суффиксом ${}_{\infty}^{a}\mathsf{T}$, слово $p_n aaba$ сильно бескубно. Слово bas'_n бескубно как суффикс w_{n+1}. Значит, t^3 должно находиться где-то на границе этих слов (включая короткое или пустое слово $\bar{s}_n$). Все эти слова можно разбить на 1-блоки (за исключением последней буквы $p_n aaba$), при этом слово $\bar{s}_n$, если оно не пустое, является 1-блоком. Если само слово r не может быть разбито на 1-блоки, то в середине его есть слово вида $xxuxx$, где u — слово нечётной длины, не содержащее квадратов букв. По Таблице 1.1 проверяется, что если такое подслово существует, то оно не может быть продлено до куба с периодом $|xxu|$. С другой стороны, такое подслово встречается в r только один раз, а значит, не может содержаться в более длинном кубе.

Теперь рассмотрим случай, когда r может быть представлено в виде произведения 1-блоков. Заметим, что слова bas'_n и $p_n aaba$ разбиваются на 2-блоки с точностью до коротких суффиксов и префиксов. Если само r не допускает такого разбиения, значит, оно содержит внутри себя подслово вида $x\bar{x}x\bar{x}ux\bar{x}x\bar{x}$, где длина u чётна, но не делится на 4, и u не содержит квадратов с периодом 2. Как и в предыдущем случае, такое подслово единственно и не порождает ни коротких, ни длинных кубов.

Наконец, предположим, что r разбивается на 2-блоки. Так как p_n не начинается с 2-блока (иначе $|u_n| > n$), то s'_n должно кончаться на $\mathsf{T}_2^{\bar{x}}\mathsf{T}_1^{x}\mathsf{T}_1^{\bar{x}}$. Но тогда r имеет подслово $\theta^2(baababba)$ внутри себя (граница между bas'_n и $p_n aaba$ находится через две позиции справа от центра этого слова). Следовательно, r не содержит короких кубов, в то время как длинные кубы возникнуть не могут (максимальная длина подслова из языка Туэ–Морса в s'_n — 48).

Этим завершается доказательство леммы. $\square$

Так как слово $p_n\overline{w}_n$ является кубом по построению (1.3a) и в то же время $p_n = u_{n+1}$ — фиксированный левый контекст слова w_{n+1}, можно заключить, что u_n — наидлиннейший левый контекст $\overline{w}_n$. А значит, Теорема 10,a доказана.

Замечание 7 *Наша конструкция даёт верхнюю оценку длины кратчайшего предмаксимального слева слова любого заданного уровня n. Результаты из [18] предполагают, что эта длина зависит от n экспоненциально. Пусть $l(n) = |w_n|$. Для нетривиальных итераций имеем $l(n) = 3l(n-1) + O(n)$. Хорошо известно, что две последовательных буквы слова Туэ–Морса равны с вероятностью $1/3$. Значит, чтобы получить w_n, нужно сделать примерно $2n/3$ нетривиальных итераций. Отсюда, функция $l(n)$ экспоненциальна по основанию $3^{2/3} \approx 2.08$. Такое же свойство справедливо и для $|\overline{w}_n| = 3l(n+1) + O(n)$.*

1.2.3 Построение двусторонних предмаксимальных слов

Доказательство Теоремы 10,б. Аналогично Замечанию 6, достаточно построить предмаксимальные слова уровня (n_i, n_i) для некоторой бесконечной подпоследовательности $n_1 < n_2 < \ldots < n_i < \ldots$ натуральных чисел. Возьмём подпоследовательность $n_i = 32i + 3$ (Таблица 1.2 показывает, что $\overline{s}_{n_i} = \lambda$, что делает построение более простым). Напомним, что мы соединяем предмаксимальное слева и предмаксимальное справа слова через некоторое промежуточное слово. Но в качестве предмаксимальных слов использовать $\overline{w}_n$ для этой цели нельзя, поскольку все слова $u_n\overline{w}_n$ оказываются максимальными справа.

Изменим последний шаг в построении предмаксимальных слева слов следующим образом. Из доказательства Леммы 3 следует, что слово $u_n w_n s_n \cdots s_{n+l}$ бескубно для любого l. Итак, положим

$$\widetilde{w}_{n_i} = \underbrace{w_{n_i+1}s_{n_i+1}s_{n_i+2}}\underbrace{p_{n_i}w_{n_i+1}s_{n_i+1}s_{n_i+2}}\underbrace{p_{n_i}w_{n_i+1}s_{n_i+1}s_{n_i+2}}.$$

По Таблице 1.1 $s_{n_i+3} = \lambda$ и $s_{n_i+4}(1) \neq s_{n_i+1}(1) = x$. Доказательство того, что $x_{n_i}\widetilde{w}_{n_i} \in \mathsf{CF}$, воспроизводит доказательство Леммы 4. Напомним, что $s_{n_i+1}(1) = p_{n_i}(1)$ согласно (s1), откуда следует, что эта буква нарушает периоды w_{n_i+1} (см. (1.2a)). С другой стороны, буква $\bar{x}$ нарушает глобальный период слова $\widetilde{w}_{n_i}$. Следовательно, из того, что $x_{n_i+1}w_{n_i+1}s_{n_i+1}\cdots s_{n_i+l} \in \mathsf{CF}$, получаем $u_{n_i}\widetilde{w}_{n_i}s_{n_i+3}\cdots s_{n_i+l} \in \mathsf{CF}$ для любого l. Значит, $\widetilde{w}_{n_i}$ — бесконечно продолжаемое вправо, предмаксимальное слева слово уровня n_i.

Выберем чётное m такое, что $|u_{n_i}\widetilde{w}_{n_i}| < 2^{m-2}$, и рассмотрим слово $\widetilde{w}_{n_i,n_i} = \widetilde{w}_n\mathsf{T}_m^{\bar{x}}\overleftarrow{\widetilde{w}_n}$:

Осталось доказать, что слово $x_{n_i}\widetilde{w}_{n_i,n_i}\overleftarrow{u_{n_i}}$ бескубно. Из выбора m и сильно бескубности $\mathsf{T}_m^{\bar{x}}$ следует, что никакой куб не содержит подслова $\mathsf{T}_m^{\bar{x}}$. Значит, благодаря симметрии достаточно проверить, что слово $v = u_{n_i}\widetilde{w}_{n_i}\mathsf{T}_m^{\bar{x}}$ бескубно. От противного, предположим, что v содержит куб t^3. Напомним, что слово

$u_{n_i}\widetilde{w}_{n_i}$ бескубно. Так как первая буква $\mathsf{T}_m^{\bar{x}}$ нарушает период $u_{n_i}\widetilde{w}_n$, имеем $|t| < \operatorname{per}(\widetilde{w}_{n_i})$. Рассмотрим самое правое подслово $aabaa$ в v, оно находится внутри подслова w_0 непосредственно перед суффиксом s'_{n_i+2} слова $\widetilde{w}_n$. Если это подслово содержится в t^3, то через $|t|$ букв левее снова встречаем $aabaa$, за которым следует s'_{n_i+2}. Тогда $|t| = \operatorname{per}(\widetilde{w}_{n_i})$, противоречие. Следовательно, t^3 не имеет подслов $aabaa$, т. е., само является подсловом $abaaba\, s'_{n_i+2}\mathsf{T}_m^{\bar{x}}$. Как мы уже знаем, слово s'_{n_i+2} не содержит подслов Туэ–Морса длины > 48. Значит, период t^3 ограничен константой. Отсутствие кубов с короткими периодами проверяется компьютерным перебором.

Таким образом, слово $\widetilde{w}_{n_i,n_i}$ предмаксимально уровня (n_i, n_i). Теорема доказана. □

1.3 Тернарный бесквадратный язык

1.3.1 Коды Пансьё и маршрутные коды

В этом подразделе мы рассмотрим кодирование тернарных бесквадратных слов с помощью бинарных кодов Пансьё над алфавитом $\{0,1\}$ и их преобразование в так называемые маршрутные коды над алфавитом $\{1,2,3\}$. Эти способы кодирования активно используются в доказательствах большинства основных результатов диссертации, касающихся языка SF.

Любое тернарное бесквадратное слово u может быть закодировано бинарным *кодом Пансьё* $\mathsf{cwd}(u)$ длины $|u| - 2$:

$$\mathsf{cwd}(u)[i] = \begin{cases} 0 & \text{, если } u[i] = u[i{+}2], \\ 1 & \text{, иначе.} \end{cases}$$

Пример 3

$$\begin{array}{rcccccccccl} u & = & a & b & c & b & a & c & b & c & \dots, \\ \mathsf{cwd}(u) & = & 1 & 0 & 1 & 1 & 1 & 0 & & & \dots. \end{array}$$

Этот тип кодирования был предложен в [50] для бо́льших алфавитов и изучен в [75] для случая трёхбуквенного алфавита. Приведём несколько фактов из [75], которые понадобятся в дальнейшем.

Во-первых, кодировка Пансьё может быть естественным образом перенесена на циклические слова. Код Пансьё для бесквадратного циклического слова (u) – это бинарное циклическое слово $(\mathsf{cwd}(u))$ длины $|u| \geqslant 3$, определяемое условиями

$$(\mathsf{cwd}(u))[i] = \begin{cases} 0 & \text{, если } u[i] = u[(i+2) mod |n|], \\ 1 & \text{, иначе.} \end{cases}$$

Пример 4

$$\begin{array}{ccc} b & c & b \\ a & & a \\ c & b & c \end{array} \longrightarrow \begin{array}{ccc} 0 & 1 & 1 \\ 1 & & 1 \\ 1 & 1 & 0 \end{array}$$

Заметим, что код Пансьё может быть построен для любого тернарного слова длины $\geqslant 3$, не содержащего квадратов букв. Коды Пансьё бесквадратных слов (обычных или циклических) также называются бесквадратными.

Рассмотрим бесквадратные циклические коды Пансьё. Они не содержат подслов 00 и 1111, кодирующих квадраты с периодами 2 и 3, соответственно. Нули в коде Пансьё соответствуют «прыжкам» одной буквы через другую в кодируемом слове. Существует шесть таких прыжков, представленных подсловами aba, bcb, cac, aca, bab и cbc. Мы называем первые три прыжка *правыми*, а остальные – *левыми*. Правый прыжок в бесквадратном циклическом слове всегда следует за левым, а левый – за правым. В частности, отсюда следует, что число нулей в бесквадратном циклическом коде Пансьё всегда чётно. Следующий прыжок получается из предыдущего

– заменой средней буквы ($aba \leftrightarrow aca$), если нули разделены одной единицей,

– заменой крайних букв ($aba \leftrightarrow cbc$), если нули разделены двумя единицами,

– переменой ролей букв ($aba \leftrightarrow bab$), если нули разделены тремя единицами.

Для описания бесквадратных циклических кодов Пансьё используется взвешенный полный двудольный граф $K_{3,3}$. Левые [правые] прыжки соответствуют нижней [соответственно, верхней] доле графа. Число единиц, необходимое для кодирования пары последовательных прыжков, равно весу ребра, соединяющего соответствующие вершины. Каждый бесквадратный циклический код

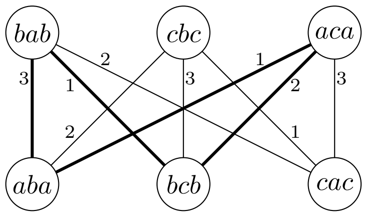


Рис. 1.3: Граф прыжков в тернарных бесквадратных циклических словах. Жирными линиями выделен замкнутый маршрут (1213).

Пансьё длины не менее 4 соответствует замкнутому маршруту во взвешенном графе, изображённом на рис. 1.3.

Маршрут в построенном графе однозначно определяется начальной вершиной и последовательностью весов рёбер. Ввиду симметрии эта последовательность определяет, будет ли маршрут замкнутым, независимо от начальной вершины. Поскольку нас интересуют коды Пансьё, а не кодируемые ими слова, мы можем рассматривать маршруты просто как последовательности весов, т.е. как слова над алфавитом $\{1, 2, 3\}$. Такие слова будем называть *маршрутными кодами*. Любой замкнутый маршрут представляет собой комбинацию простых циклов, следующих друг за другом или вложенных друг в друга (замкнутый маршрут длины 2 тоже рассматривается как цикл). Каждый простой цикл задаётся циклическим словом над $\{1, 2, 3\}$. Следующая лемма наиболее важна в рассуждениях, связанных с бесквадратными маршрутными кодами.

Лемма 5 ([75]) *Замкнутый маршрут, не имеющий*
(а) подслов 11, 222, 223, 322, 333,
(б) подслов вида XYX таких, что $|Y| = 2$, $|X|$ чётна и (XY) является меткой замкнутого маршрута,
определяет бесквадратный циклический код Пансьё.

Будем получать маршрутный код $\mathsf{cwk}(u)$ для произвольного (не циклического) тернарного бесквадратного слова u следующим образом. Берём код Пансьё $\mathsf{cwd}(u)$ этого слова, заменяем каждый блок из единиц его длиной и опускаем все нули. Для однозначности необходимо хранить информацию о первой и последней букве $\mathsf{cwd}(u)$ (например, если $\mathsf{cwd}(u)[1] = 1$, то первый блок из единиц в $\mathsf{cwd}(u)$ может быть частью более длинного блока). Мы подчёркиваем первую [последнюю] букву $\mathsf{cwk}(u)$, если $\mathsf{cwd}(u)$ начинается [соответственно, кончается] на 1.

Пример 5

$$u = abcacbacaba, \quad \mathsf{cwd}(u) = 110111010, \quad \mathsf{cwk}(u) = \underline{2}31.$$

Замечание 8 *Упомянутый во введении результат Туэ [80] о существовании тернарных бесквадратных слов, избегающих два подслова длины 3, имеет ясную и простую интерпретацию в описанных выше терминах. А именно, в описанном взвешенном графе $K_{3,3}$ для любых двух фиксированных вершин из разных долей существует маршрут, соответствующий бесквадратному слову, ни разу не посещающий эти вершины. Также теперь легко видеть, почему три и более различных подслова вида xyx одновременно неизбежны в бесквадратных словах.*

1.3.2 Обзор основной конструкции для предмаксимальных слов

Перейдём к рассмотрению особенностей основной конструкции для построения предмаксимальных слов в случае языка SF.

Положим $w_0 = cabcacba$ и заметим, что у этого слова есть фиксированный левый контекст ba длины 2. Комбинация уравнения 1.3b и Предложения 1 даёт нам следующую идею: найти такое слово s_n, что циклическое слово $(zu_nw_ns_n)$ бесквадратно, и положить $w_{n+1} = w_ns_nzu_nw_ns_n$. Как и в случае бескубных слов, бозначим $p'_n = zu_n$, $p_n = yu_n$. В качестве слова U для построения серии предмаксимальных слов произвольного конечного уровня в тернарном бесквадратном языке возьмём реверс слова Аршона $\overleftarrow{\mathsf{A}}$. Напомним что p_n является суффиксом U для любого n.

Замечание 9 $(n{+}1)$*-ая итерация тривиальна, если u_n начинается с aba, $abcab$ или $abcbabc$ с точностью до переименования букв. Во всех других случаях мы сможем произвести нетривиальную итерацию благодаря тому, что слово Аршона $(7/4)^+$-свободно, а значит, слово u'_{n+1} бесквадратно.*

Слово s_n должно выполнять три функции:

- «замыкать» соответствующий маршрутный код, если он не замкнут;
- предотвращать появление квадратов после склеивания начала слова с его концом;
- обеспечивать выполнимость следующей нетривиальной итерации (см. следующее замечание).

Замечание 10 *Так как слово $u_nw_{n+1} = u_nw_ns_nzu_nw_ns_n$ является квадратом без одной буквы, его можно продолжить вправо бесквадратным образом только буквой, отличной и от последней буквы s_n, и от z, то есть оно имеет единственный правый контекст длины 1. Мы приписываем этот* *контекст* перед *добавлением буферного слова на следующей нетривиальной итерации.*

Мы будем описывать s_n в терминах маршрутных кодов (иногда прибегая также к кодам Пансьё). Заметим, что $|(\mathsf{cwd}(u))| = |\mathsf{cwd}(u)| + 2$, таким образом, нам нужно добавить не менее двух символов справа к коду Пансьё слова zu_nw_n, чтобы w_n было префиксом w_{n+1}. Пусть $v = \mathsf{cwd}(zu_nw_ns_n)$. Тогда код Пансьё $\mathsf{cwd}(w_{n+1})$ получается следующим образом: берём слово v^2, удаляем из него первые $n{+}1$ символов (они кодируют p'_n) и удаляем два последних символа, которые соединяют конец циклического слова с его началом, то есть кодируют две первых буквы слова. Согласно замечанию 10, мы всегда начинаем построение кода Пансьё следующего буферного слова с его первого символа, этот символ кодирует «неизбежную» следующую букву с предыдущей нетривиальной итерации. Так как эта буква не равна z, добавляемый символ отличается от второго с конца символа v (этот символ кодировал z в циклическом коде Пансьё).

Начиная с седьмой итерации (она соответствует третьей нетривиальной), все нетривиальные итерации производятся по общей схеме, описанной в разделе 1.3.5. Будем называть такие итерации *общими*. Первая, вторая, пятая и шестая итерации тривиальны. Первые две нетривиальные итерации производятся так:

$$w_2 = w_0 = cabcacba, \quad p'_2w_2 = a\,ba\,cabcacba, \quad \mathsf{cwd}(p'_2w_2) = 010111011.$$

Добавляя 0111 к концу этого кода Пансьё, мы получаем замкнутый маршрутный код 132$\underline{3}$, соответствующий бесквадратному циклическому слову по лемме 5. Удваиваем код Пансьё, удаляем первые три символа и два последних, добавляем новый последний символ и декодируем результат, чтобы получить слово w_3 (для удобства код Пансьё p'_3w_3, нужный для следующей итерации, приведён здесь же):

	010111011011 10101110110111
$\mathsf{cwd}(w_3)$	111011011 1010111011010
w_3	$cabcacbabca\ bacabcacbabcb$
$\mathsf{cwd}(p'_3w_3)$	0110111011011 1010111011010

«Неизбежная» буква после w_3 – это b (a запрещается на этой итерации). Теперь рассмотрим четвёртую итерацию. Код Пансьё $\mathsf{cwd}(p'_3w_3)$ соответствует замкнутому маршрутному коду 23231321, но мы добавили 0 к коду Пансьё, чтобы закодировать «неизбежную» букву. В итоге после кодирования последней буквы код Пансьё содержит 0 в начале и в конце, в таком виде замкнуть его в циклический код нельзя. Поэтому добавляем в конец 1110111 и получаем замкнутый маршрутный код $\mathsf{232313213\underline{3}}$. По лемме 5, он кодирует бесквадратный циклический код Пансьё.

	01101110110111010111011010111011	10110111011011101011101101011110111
$\mathsf{cwd}(w_4)$	1110110111010111011010111011	10110111011011101011101101011101 0
w_4	$cabcacbabcabacabcacbabcbacbcabc$	$bacabcacbabcabacabcacbabcbacbcac$

После декодирования получаем $w_4 = w_3 bacbca\, bcbaw_3 bacbca$. Легко проверить, что слово w_3 имеет фиксированный левый контекст длины 3 cba, являющийся суффиксом $\overleftarrow{\mathsf{A}}$, а слово w_4 имеет фиксированный левый контекст $bcacba$ с тем же свойством (пятая и шестая итерации тривиальны).

1.3.3 Маршрутный код слова Аршона

Для описания дальнейших итераций нам понадобится маршрутный код реверса слова Аршона. Справедливо следующее

Замечание 11 *Подслова Аршона могут встречаться в* A *на позициях различной чётности и, следовательно, могут иметь 2 разных* α*-образа. Обозначим как* $\alpha_o(w)$ *[*$\alpha_e(w)$*] образ подслова* w*, если* w *начинается в* A *в нечётной [соответственно, чётной] позиции. Очевидно, что подслово Аршона* $\alpha_o(w)$ *[*$\alpha_e(w)$*] начинается в нечётной [соответственно, чётной] позиции. Следовательно,* w *имеет не более двух различных* α^k*-образов, обозначаемых как* $\alpha_o^k(w)$ *и* $\alpha_e^k(w)$*, для любого* $k > 0$.

Оказывается, что маршрутный код слова Аршона и его реверса имеет очень простое описание.

Лемма 6 *Пусть* $f = \mathsf{132}$*,* $g = \mathsf{123}$ *и* B *- бесконечное вправо слово, порождённое морфизмом* $\beta : \{f,g\}^* \to \{f,g\}^*$*, заданным правилами:* $\beta(f) = fgf$*,* $\beta(g) = fgg$*. Тогда* $\mathsf{cwk}(\mathsf{A}) = \underline{\mathsf{2}}\mathsf{B}$ *и* $\mathsf{cwk}(\overleftarrow{\mathsf{A}}) = \overleftarrow{\mathsf{B}'}\mathsf{1}\underline{\mathsf{2}}$*, где* B' *получается из* B *переименованием букв* f *и* g*.*

Доказательство. Согласно замечанию 11, для любого $k > 0$ α^k-образ буквы x определяется чётностью позиции x. Будем называть слова $\alpha_o^k(x)$ и $\alpha_e^k(x)$ *α^k-блоками*. Обратим внимание на два очевидных свойства блоков Аршона и функции α.

(а) α-блоки можно рассматривать как перестановки алфавита Σ: нечётные α-блоки являются перестановками одной чётности, а чётные α-блоки – перестановками противоположной чётности.

(б) x – первая буква $\alpha_o(x)$ и последняя буква $\alpha_e(x)$.

Предположим, что $x \in \Sigma$ и $\alpha_o(x) = xyz$. Используя свойства (а),(б), легко проверить, что

$$\begin{array}{rclcl} \alpha_o^2(x) & = & xyz\,xzy\,zxy & = & \alpha_o(x)\,\alpha_e(y)\,\alpha_o(z), \\ \alpha_e^2(x) & = & yxz\,yzx\,zyx & = & \alpha_e(z)\,\alpha_o(y)\,\alpha_e(x). \end{array}$$

Заметим, что $\alpha_e^2(x) = \overleftarrow{\alpha_o^2(x)}$. Получим по индукции равенство $\alpha_e^k(x) = \overleftarrow{\alpha_o^k(x)}$ для любого k. Пусть $\alpha_e^{k-1}(x) = \overleftarrow{\alpha_o^{k-1}(x)}$. Ввиду симметричности, то же справедливо для α^{k-1}-образов y и z. Но тогда

$$\begin{aligned} \alpha_o^k(x) = \alpha_o^{k-1}(x)\,\alpha_e^{k-1}(y)\,\alpha_o^{k-1}(z) = & \\ = \overleftarrow{\alpha_e^{k-1}(x)}\,\overleftarrow{\alpha_o^{k-1}(y)}\,\overleftarrow{\alpha_e^{k-1}(z)} &= \overleftarrow{\alpha_e^{k-1}(x)\,\alpha_o^{k-1}(y)\,\alpha_e^{k-1}(z)} = \overleftarrow{\alpha_e^k(x)}, \end{aligned}$$

что и требовалось.

Теперь заметим, что все α^2-блоки эквивалентны в том смысле, что они являются образами друг друга при перестановках алфавита. Тогда каждый α^2-блок имеет код Пансьё 1101011 и маршрутный код $\underline{2}1\underline{2}$. При соединении двух α^2-блоков должны быть вставлены два символа между их кодами Пансьё. Эти два символа – 01 или 10, так как коды Пансьё бесквадратных слов не содержат ни 00, ни 1111. Следовательно, код Пансьё двух последовательных α^2-блоков будет выглядеть как $\underline{2}1321\underline{2}$ или $\underline{2}1231\underline{2}$. Отсюда $\mathsf{cwk}(\mathsf{A}) \in \underline{2}(f+g)^*$.

Теперь рассмотрим α^3-блоки. Имеем $\alpha_o^3(x) = \alpha_o^2(x)\alpha_e^2(y)\alpha_o^2(z)$. Глядя на начала и концы соответствующих α^2-блоков, можно увидеть, что первая вставка в код Пансьё – это 10, а вторая – 01. Следовательно, маршрутный код нечётного α^3-блока равен $\underline{2}1321231\underline{2}$. Так как этот маршрутный код является палиндромом, а нечётные и чётные α^3-образы представляют из себя реверсы друг друга, все α^3-блоки имеют такой маршрутный код. Из определения B следует, что если $i \equiv 0 \pmod 3$, то $\mathsf{B}[i] = \mathsf{B}[i/3]$, $\mathsf{B}[i+1] = f$, $\mathsf{B}[i+2] = g$. Таким образом, мы только что доказали два последних равенства, осталось доказать первое.

Требуется определить символы, вставляемые в код Пансьё на границе двух α^k-блоков, где $k > 2$. Рассмотрим α^{k+1}-блок. Как и в предыдущих рассуждениях, благодаря симметрии достаточно рассмотреть какой-то один нечётный блок:

$$\mathsf{A} = \begin{array}{|c|c|c|c|c|c|c|c|c|} \multicolumn{9}{c}{\alpha_o^{k+1}(x)} \\ \hline \multicolumn{3}{|c|}{\alpha_o^k(x)} & \multicolumn{3}{c|}{\alpha_e^k(y)} & \multicolumn{3}{c|}{\alpha_o^k(z)} \\ \hline \alpha_o^{k-1}(x) & \alpha_e^{k-1}(y) & \alpha_o^{k-1}(z) & \alpha_e^{k-1}(x) & \alpha_o^{k-1}(z) & \alpha_e^{k-1}(y) & \alpha_o^{k-1}(z) & \alpha_e^{k-1}(x) & \alpha_o^{k-1}(y) \\ \hline \end{array}$$

Первая граница находится в середине подслова $\alpha_o^{k-1}(z)\alpha_e^{k-1}(x)$. Это подслово появляется в $\alpha_o^k(z)$ (см. рисунок) и, значит, никогда не появляется в чётном α^k-блоке по свойству (a). Следовательно, вставка в код Пансьё на границе двух первых α^k-блоков в α^{k+1}-блоке такая же, как вставка на границе двух первых α^{k-1}-блоков в α^k-блоке. Рассуждения для вставки на второй границе α^k-блоков аналогичны (подслово $\alpha_e^{k-1}(y)\alpha_o^{k-1}(z)$, встречающееся в $\alpha_o^k(x)$ и не встречающееся в чётном α^k-блоке). Значит, мы подтвердили требуемое равенство $\mathsf{B}[i] = \mathsf{B}[i/3]$. Доказательство равенства $\mathsf{cwk}(\mathsf{A}) = \underline{2}\mathsf{B}$ завершено. Формула для $\mathsf{cwk}(\overleftarrow{\mathsf{A}})$ непосредственно следует из определений и формулы для $\mathsf{cwk}(\mathsf{A})$. □

1.3.4 Свойства слов p'_n

Пользуясь знанием о структуре маршрутного кода слова Аршона, докажем некоторые свойства слов p'_n.

Лемма 7 *Пусть v – префикс маршрутного кода $\mathsf{cwk}(p'_n)$, являющийся замкнутым маршрутом длины $\geqslant 6$, а v' – самый длинный префикс $\mathsf{cwk}(p'_n)$ с периодом $|v|$. Тогда $|v'| \le 2|v| - 4$.*

Доказательство. Напомним, что слово p'_n бесквадратно. Так как слова $p_n = \overleftarrow{\mathsf{A}}[n..1]$ и p'_n различаются между собой только первой буквой, то же справедливо и для их кодов Пансьё $\mathsf{cwd}(p'_n)$ и $\mathsf{cwd}(p_n)$. Рассмотрим все возможные на нетривиальных итерациях префиксы $\mathsf{cwd}(p_n)$. Его префиксами не могут быть 10, 0111 и 110101, потому что 00, 1111 и 010101 кодируют квадраты. Также, $\mathsf{cwd}(p_n)$ не может начинаться с 01010, 0110110 или 1110111 по лемме 6. Оставшиеся случаи рассмотрены ниже.

- 01011. Согласно лемме 6, это начало блока 123 или 132 в маршрутном коде. Так как $\mathsf{cwd}(p'_n) = 11011\ldots$, получаем $\mathsf{cwk}(p'_n) = \underline{2}2\ldots$ либо $\mathsf{cwk}(p'_n) = \underline{2}32\ldots$. Подслова 22, 33, 323 и 232 не встречаются внутри маршрутного кода слова p_n. Следовательно, $|v'| \le |v| + 2 \le 2|v| - 4$.

- 11011. По лемме 6, эти блоки единиц образуют часть фрагмента 231 или 321 в маршрутном коде. Меняя первую букву, получаем 121 или 131. Эти подслова не встречаются внутри $\mathsf{cwk}(p_n)$. Следовательно, $|v'| \le |v| + 2 \le 2|v| - 4$.

- 0110111. Здесь $\mathsf{cwk}(p'_n) = \underline{3}3\ldots$. Подслово 33 не встречается внутри $\mathsf{cwk}(p_n)$. Значит, $|v'| \le |v| + 1 < 2|v| - 4$.

- 1110110. Здесь $\mathsf{cwk}(p'_n) = 22\ldots$. Подслово 22 также не встречается в $\mathsf{cwk}(p_n)$. Значит, $|v'| \le |v| + 1 < 2|v| - 4$.

- 011010. Имеем $\mathsf{cwk}(p'_n) = \underline{3}1\ldots$

- 111010. Имеем $\mathsf{cwk}(p'_n) = \mathsf{21}\ldots$.

Два последних случая похожи, поэтому будем их рассматривать одновременно. Пронумеруем блоки $f = \mathsf{132}$ и $g = \mathsf{123}$ маршрутного кода $\mathsf{cwk}(p'_n)$ слева направо, начиная с 1. Тогда первый блок начинается во второй позиции $\mathsf{cwk}(p'_n)$. Следовательно, если $|v'| > |v| + 1$, то $|v|$ делится на 3, иначе внутри $\mathsf{cwk}(p_n)$ оказалось бы подслово 11, 121 или 131, что невозможно по лемме 6. Так как v – замкнутый маршрут, его длина чётна. Отсюда заключаем, что $|v|$ делится на 6. Непосредственно по рис. 1.3 можно проверить, что замкнутые маршруты длины 6, которые разбиваются на блоки f и g – это 123123, 132132 и сопряжённые с ними и других нет. Замкнутые маршруты длины 12 представляют из себя комбинации циклов 123123, 132132 и сопряжённые этих комбинаций. Непосредственно из описания морфизма β следует, что нигде в $\mathsf{cwk}(\mathsf{A})$ не встречается подслов fff, ggg, $fggffg$, $ffggff$ и $ggffgg$. Следовательно, самый длинный префикс маршрутного кода $\mathsf{cwk}(p'_n)$ с периодом $|v| = 6$ имеет длину 8 ($\mathsf{2}ff\mathsf{1}$ или $\mathsf{3}gg\mathsf{1}$), а самый длинный префикс $\mathsf{cwk}(p'_n)$ с периодом $|v| = 12$ имеет длину 20 ($\mathsf{3}gffggf\mathsf{1}$).

Пусть, наконец, $|v| = 6k$ для некоторого $k \geq 3$. Заметим, что блок g или f в маршрутном коде соответствует подслову длины 9 в исходном слове (6 единиц и 3 нуля). Слово v кодирует $\frac{6k}{3} \cdot 9 = 18k$ букв, $k \geqslant 3$. Рассуждая от противного, предположим, что самый длинный префикс $\mathsf{cwk}(p'_n)$ с периодом $|v|$ имеет длину как минимум $2|v|{-}3$. Суффикс v длины 3 кодирует не более 10 букв (в нём не менее одной единицы и не более двух троек, так как первое вхождение v в $\mathsf{cwk}(p'_n)$ заканчивается за 1 символ до начала следующего блока f или g). Если p'_n имеет префикс длины $36k{-}10$ с периодом $18k$, то p_n содержит подслово длины $36k{-}11$ с таким же периодом, а именно, подслово $p_n[2..36k{-}10]$. Но $(36k - 11)/18k > 7/4$, что противоречит $(7/4)^+$-свободности слова Аршона. Это противоречие завершает доказательство. $\square$

Замечание 12 *Если в условиях Леммы 7 заменить* $\mathsf{cwk}(p'_n)$ *на* $\mathsf{cwk}(p_n)$, *то все рассуждения в ее доказательстве, начиная с того, что длина кратна 6, останутся верными; более того, мы получим чуть лучшую оценку* $|v'| \leqslant 2|v| - 5$.

Замечание 13 *Дополняя результат леммы 7, покажем, что можно избежать «коротких» квадратов, продолжая* p'_n *влево.*
(1) По лемме 6 $\mathsf{cwk}(p'_n)$ *не может иметь префиксы* 11, $\underline{1}$1, 223, 222, $\underline{2}$22, 322, $\underline{3}$22, 333 *и* $\underline{3}$33. *Если* $\mathsf{cwk}(p'_n) = \underline{2}\mathsf{23}\ldots$, *заменим* $\underline{2}$ *на* 3 *(т.е.,* $\mathsf{cwd}(s_n)$ *закончится на* $\underline{1}$*). Таким образом, мы можем избежать подслов в маршрутном коде, перечисленных в лемме 5,а).*
(2) В графе $K_{3,3}$, *изображённом на рис. 1.3, есть всего три различных цикла длины 4, они представлены маршрутными кодами* 1213, 1232, 1323 *и*

сопряжёнными с ними. Шесть из этих двенадцати маршрутных кодов могут быть префиксами $\mathsf{cwk}(p'_n)$*:* 1213, 1312, 2123, $\underline{3}132$, $\underline{2}321$*, а также* $\underline{2}231$*, который будет заменён на* $\underline{3}231$ *согласно предыдущему пункту. Первые четыре маршрутных кода не могут быть продолжены до периодического слова по лемме 6.* $\underline{2}321$ *и* 3231 *могут быть продолжены до «нехороших» в смысле леммы 5,б маршрутных кодов* $\underline{2}32123$ *и* 323132*, соответственно. В обоих случаях за этими периодическими подсловами в маршрутном коде слова* p'_n *следует* 1*. Когда* $\mathsf{cwk}(p'_n) = \underline{2}32123\ldots$*, мы меняем* $\underline{2}$ *на* 3*. Когда* $\mathsf{cwk}(p'_n) = 323132\ldots$*, мы заканчиваем* $\mathsf{cwk}(s_n)$ *на* 32 *и обрываем этот период с другой стороны (Лемма 5 не является критерием, маршрутный код* 323132 *кодирует «квадрат без одной буквы» – слово вида* wxw*).*

1.3.5 Построение буферных слов и доказательство бесквадратности

Построение буферных слов s_n опирается на следующую лемму, которая говорит о том, что к слову u'_n можно присоединить слева подслово Аршона любой наперёд заданной длины через некоторое слово-вставку.

Лемма 8 *Для любого слова* p'_n *существует слово* t'_n *такое, что для любого* m *со свойством* $\mathsf{cwd}(\mathsf{A})[m{+}1] = 0$ *слово* $\mathsf{A}[1..m]t'_n$ *является левым контекстом* p'_n*.*

Доказательство. Нам необходимо найти код Пансьё C такой, что код $\mathsf{cwd}(A)[1..m]C\mathsf{cwd}(p'_n)$ бесквадратен для любого m, удовлетворяющего условию $\mathsf{cwd}(\mathsf{A})[m{+}1] = 0$. То есть мы полагаем, что C всегда начинается с 0. На самом деле, мы можем перейти от кодов Пансьё к маршрутным кодам и искать маршрутный код C' такой, что $\mathsf{cwk}(A)[1..m]C'\mathsf{cwk}(p'_n)$ кодирует бесквадратное слово.

Мы знаем, что $\mathsf{cwk}(p'_n)$ совпадает с $\mathsf{cwk}(p_n)$ во всех позициях, кроме первой, а $p_n = \overleftarrow{\mathsf{A}}[n+1..1]$. Маршрутный код $\mathsf{cwk}(p'_n)$ имеет одну из следующих форм:

1. $21\underline{2}\ldots$
2. $\underline{3}1\underline{2}\ldots$
3. $121\ldots$
4. $131\ldots$
5. $221\ldots$
6. $\underline{3}31\ldots$
7. $\underline{2}23\ldots$
8. $\underline{2}32123\ldots$
9. $\underline{2}32132\ldots.$

Следование форм $\mathsf{cwk}(p'_n)$ друг за другом на последовательных нетривиальных итерациях изображено на рис. 1.4. Для каждой вершины маршрутным кодом C' помечено исходящее ребро. Этот код не зависит от вершины, в которую ведёт данное ребро. Данные маршрутные коды подбираются таким образом, чтобы быть как можно более «непохожими» на маршрутный код слова Аршона, то есть создавать некий «барьер» для периода, который может появиться при соединении подслова Аршона с p'_n: их префиксы длины 3 и суффиксы длины 4 (3 в большинстве случаев) не встречаются в cwk(A). Покажем, что маршрутный код C', соответствующий форме 1 маршрутного кода слова p'_n, удовлетворяет условиям леммы. Доказательства для всех остальных случаев аналогичны.

Вершина 1 соответствует префиксу $\underline{2}1\underline{2}$ и вставке $C' =$ 3233233. Легко заметить, что маршрутный код $\mathsf{cwk}(A)[1..m]C'\mathsf{cwk}(p'_n)$ по построению не содержит подслов, перечисленных в лемме 5,а). Более того, $\mathsf{cwk}(A)[1..m]C'$ кодирует бесквадратное слово по лемме 5,б). В самом деле, по замечанию 12 любой суффикс XYX слова $\mathsf{cwk}(A)[1..m]$ такой, что XY - замкнутый маршрут, имеет длину не более $2|XY| - 5$, и этот суффикс может быть продолжен не более чем двумя символами маршрутного кода C', потому что 323 не встречается внутри $\mathsf{cwk}(A)[1..m]$ (в случае, если $\mathsf{cwk}(A)[1..m]$ заканчивается на 1; в остальных случаях продолжение $\mathsf{cwk}(A)[1..m]$ в C' ещё короче). Далее, $C'\mathsf{cwk}(p'_n)$ кодирует бесквадратное слово, так как суффикс 33 слова C' не встречается нигде внутри $\mathsf{cwk}(p'_n)$.

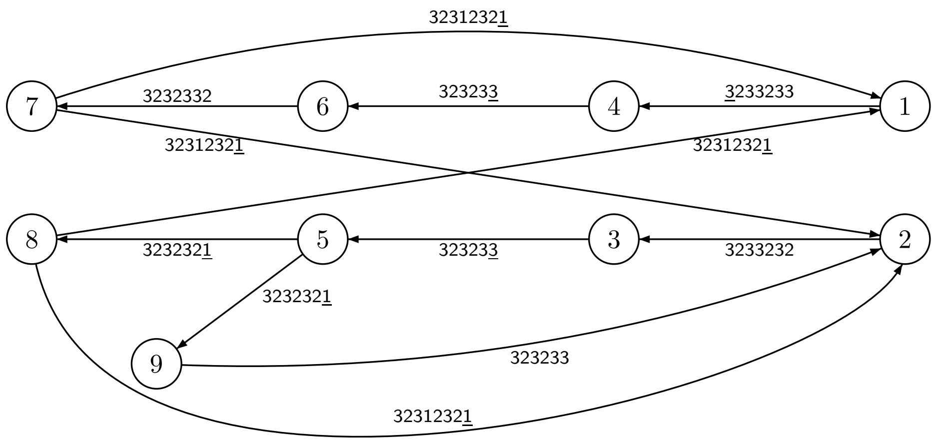


Рис. 1.4: Нетривиальные итерации. Вершины соответствуют формам $\mathsf{cwk}(p'_n)$, рёбра помечены значениями C'.

Наконец, допустим, что $\mathsf{cwk}(A)[1..m]C'\mathsf{cwk}(p'_n)$ содержит периодическое подслово XYX, строго содержащее C'. Тогда подслово 3323 слова C' находится в Y, так как ни 323, ни 33 не могут повториться внутри маршрутного кода слова Аршона. Таким образом, $|XYX| \leq 2|XY| - 4$, а значит, рассматриваемый маршрутный код кодирует бесквадратное слово по лемме 5,б. После декодирования получим бесквадратное слово вида $\mathsf{A}[1..m]t'_n p'_n$, что и требовалось. □

Пусть $(n+1)$-ая итерация нетривиальна и имеет номер m среди общих нетривиальных итераций. В соответствии с доказанной леммой, будем искать буферное слово s_n в виде $\mathsf{A}[i..j]t'_n$. Используя представление маршрутного кода $\mathsf{cwk}(\mathsf{A})$ через морфизм β из леммы 6, возьмём маршрутный код $\mathsf{B}[m..2m{-}1]C'\mathsf{cwk}(p'_n)$, который кодирует бесквадратное слово по лемме 8. Рассмотрим маршрутный код $\mathsf{B}[m..2m{-}1]C'\mathsf{cwk}(p'_n w_n)$. Если этот маршрут замкнут, докажем, что он кодирует бесквадратное циклическое слово. В противном случае сначала замкнём его, заменив $\mathsf{B}[m..2m{-}1] = \mathsf{cwk}(\mathsf{A})[3m{-}1..6m{-}2]$ на некоторое более длинное подслово маршрутного кода B, начинающегося с той же позиции, - $\mathsf{cwk}(\mathsf{A})[3m{-}1..k_m]$. Для замыкания маршрута надо добавить ребра, переводящие текущую вершину в ту, с которой маршрут начинался (т.е., в общем случае, в любую наперед заданную). Из рис. 1.3 видно, что любые две различные вершины графа $K_{3,3}$ могут быть соединены путями 1, 2, 3, 12 или 13. 1 — это следующая буква маршрутного кода Аршона после позиции $6m{-}2$. Каждый из оставшихся четырёх путей может быть заменён на подслово $\mathsf{cwk}(\mathsf{A})$, непосредственно следующее за $\mathsf{B}[m..2m{-}1]$:

2	может быть заменён на	123 или 13213 или 1321231
3	может быть заменён на	132 или 12312 или 1231321
12	может быть использован или заменён на	1321
13	может быть использован или заменён на	1231

(1.6)

Вставляя подходящее подслово после $\mathsf{B}[m..2m{-}1]$ в рассматриваемом маршрутном коде, получаем замкнутый маршрутный код $v_m = \mathsf{cwk}(\mathsf{A})[3m{-}1..k_m]$ $C'\mathsf{cwk}(p'_n w_{n-1})$, причём его префикс $\mathsf{cwk}(\mathsf{A})[3m{-}1..k_m]C'\mathsf{cwk}(p'_n)$ кодирует бесквадратное слово по лемме 8. Рассмотрим границу между $\mathsf{cwk}(w_n)$ и $\mathsf{B}[m..2m{-}1]$ в циклическом слове (v_m).

Слово w_n заканчивается на буферное слово, построенное на $(m{-}1)$-ой общей нетривиальной итерации из маршрутного кода, оканчивающегося на некоторый суффикс C''. Отметим, что $\mathsf{cwk}(w_n)$ оканчивается на «искажённый» вариант $\mathsf{cwk}(C'')$ согласно замечанию 10. А именно, последняя буква кода Пансьё, соответствующего C'', удалена, а предпоследняя — изменена. Тем самым, суффикс 32$\underline{1}$ [соответственно, 232, 23$\underline{3}$, 233 или 332] слова C'' поменялся на 33 [соответственно, 231, 231, 232 и 331] (C'' имеет один из этих суффиксов согласно Рис. 1.4). Слово $\mathsf{B}[m]$ начинается с 1, а значит, если $\mathsf{cwk}(w_n)$ закончилось на 1, мы сначала добавим 33 к искажённому C''. Обозначим как $\widetilde{C}_{m-1}$ слово, полученное таким образом из C''.

Лемма 9 *Следующий маршрутный код кодирует бесквадратное слово для любого m:*

$$D_m = \mathsf{cwk}(\mathsf{A})[2..k_1]\widetilde{C}_1\mathsf{cwk}(\mathsf{A})[5..k_2]\widetilde{C}_2\cdots\mathsf{cwk}(\mathsf{A})[3m{-}1..k_m]\widetilde{C}_m.$$

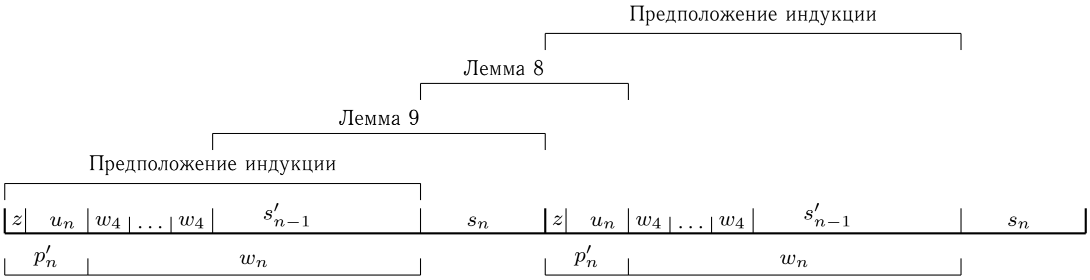
Доказательство. Лемма 5 справедлива и для маршрутных кодов обычных слов. По построению D_m не содержит подслов, перечисленных в лемме 5,а). Предположим, что D_m содержит периодическое подслово XYX, описанное в лемме 5,б). Для $|XY| = 4$ отсутствие таких подслов может быть проверено непосредственно по построению маршрутных кодов C' из доказательства леммы 8. Пусть $|XY| \geq 6$. Для коротких подслов рассуждаем так же, как в доказательстве леммы 8. Напомним, что такое подслово в маршрутном коде Аршона имеет длину не более $2|XY| - 5$ и может быть продолжено в $\widetilde{C}_i$ не более чем двумя буквами. Если такое подслово содержится в $\mathsf{cwk}(\mathsf{A})[3i{-}1..k_i]\widetilde{C}_i\mathsf{cwk}(\mathsf{A})[3i{+}2..k_{i+1}]$ и включает в себя $\widetilde{C}_i$, то не менее трёх букв $\widetilde{C}_i$ находится в Y. Для более длинных подслов XYX слово Y ещё длиннее, так как слово $\widetilde{C}_i$ может совпасть только с некоторым $\widetilde{C}_{i+j}$. Но из равенства $\widetilde{C}_i = \widetilde{C}_{i+j}$ следует, что $j \geq 3$ (см. рис. 1.4), а фрагменты маршрутного кода $\mathsf{cwk}(\mathsf{A})$, окружающие $\widetilde{C}_i$, строго короче, чем соответствующие фрагменты вокруг $\widetilde{C}_{i+j}$. □

Лемма 10 *Для любого $n \in \mathbb{N}$ слово $p_n w_{n+1}$ бесквадратно.*

Доказательство. Требуется доказать два утверждения:
(а) $p'_n w_{n+1}$ - минимальный квадрат;
(б) $p_n w_{n+1}$ не начинается с квадрата.
Рассматриваемые слова «помечены» вхождениями w_4 — всего имеется 2^{m-4} «регулярных» вхождения в каждое подслово w_k, где m – номер общей нетривиальной итерации, соответствующий k-ой итерации. Эти вхождения w_4 получаются при удвоении слова на каждой нетривиальной итерации. Больше нигде, кроме как на этих «регулярных позициях», w_4 не встречается. Докажем это. Имеем $\mathsf{cwk}(w_4) = 3231321332323132131$. Отсюда видно, что w_4 не может совпасть ни с каким подсловом Аршона. Далее, по рис. 1.4 непосредственно проверяется, что $\mathsf{cwk}(w_4)$ не содержит подслов, равных какому-нибудь из слов C'. Более того, этот маршрутный код не содержит ни одного «искажённого» при переходе к следующей нетривиальной итерации варианта C' с учётом добавленного к нему при необходимости цикла 33.

Каждому w_4 в $p_n w_{n+1}$ предшествует некоторое p_i, $i < n$ и только первому вхождению предшествует p_n. Последнее свойство, дополненное прямой проверкой для коротких периодов, даёт нам утверждение (б). Ввиду леммы 1, чтобы доказать утверждение (а), нам нужно убедиться, что циклическое сло-

Рис. 1.5: Структура слова $v^2 = (p'_n w_n s_n)^2$



во $(v) = (p'_n w_n s_n)$ бесквадратно. Будем делать это по индукции по n, база $n \le 6$ доказана в подразделе 1.3.2.

Докажем шаг индукции. Рисунок 1.5 иллюстрирует структуру слова v^2. Сверху обозначены подслова, для которых уже доказано в предыдущих леммах, что они бесквадратны. s'_{n-1} обозначает слово $s_6 \cdots s_{n-1}$.

Предположим обратное: пусть квадрат $XX = X^{(1)}X^{(2)}$ является подсловом (v). Мы уже доказали, что подслово $p'_n w_4$ встречается в (v) только один раз — в начале слова v. То же верно и для s_n: его маршрутный код состоит из некоторого подслова $\mathsf{B}[m..k_m]$ и вставки C', а мы ни разу не вставляем два одинаковых подслова из B. Следовательно, слова s_n и $p'_n w_4$ не могут быть подсловами X. Значит, есть всего три варианта расположения XX:

1. $X^{(2)}$ заканчивается в s_n;
2. $X^{(2)}$ начинается в s_n и заканчивается в p'_n;
3. $X^{(2)}$ начинается в p'_n.

Сначала докажем, что бесквадратно не только слово $s'_n p'_n$, но и $s'_n p'_n w_4$. Пусть, от противного, $s'_n p'_n w_4$ содержит квадрат. Понятно, что этот квадрат должен захватывать часть w_4. Но перед ним находится подслово Аршона u_n. Очевидно, что X содержит какую-то часть u_n, а она является подсловом Аршона. Значит, это подслово должно «наложиться» на подслово Аршона из некоторого s_k. Но за ним следует либо ещё одно подслово Аршона, которое не может совпасть с началом w_4, либо вставка из Леммы 8, которая также не может совпасть с началом w_4, что легко проверяется в терминах их маршрутных кодов.

Рассмотрим первый случай. $X^{(2)}$ начинается позже, чем p'_n (случай, в котором $X^{(2)}$ начинается в p'_n, обсуждается ниже). $X^{(2)}$ не может содержать s'_{n-1}, так как ему негде повториться раньше вместе с префиксом s_n. Значит, $X^{(2)}$

начинается внутри s'_{n-1}, то есть X не содержит w_4, следовательно, $X^{(1)}$ начинается в w_4, находящемся непосредственно перед s'_{n-1}. Значит, весь квадрат XX строго содержится в подслове $w_4 s'_{n-1} s_n$. Применяя рассуждения, аналогичные доказательству леммы 9 (суффикс w_4, входящий в X, может играть ту же роль, что и вставки, соответствующие маршрутным кодам C''), приходим к противоречию.

Во втором случае «подавляющий Аршона» суффикс s_n (т.е. слово, в которое декодируется добавленный на этой итерации маршрутный код C' из доказательства леммы 8) может повториться левее не позже, чем в суффиксе s_{n-3}. Но тогда период квадрата больше, чем $|s_n p'_n|$, — снова получили противоречие.

В третьем случае $X^{(2)}$ содержит w_4, следующее за p'_n. Самое близкое к нему вхождение w_4 слева находится перед s'_n. Значит, X содержит s'_n. Но как отмечалось выше, даже s_n не может быть подсловом X. Получаем противоречие, которое и завершает доказательство. □

1.3.6 Построение предмаксимальных слов

По построению слово u_n — фиксированный левый контекст w_n. Рассмотрим теперь второй этап, то есть, достройку такого «почти однозначно» продолжаемого слова w_n до предмаксимального слева слова. Основная идея такая же, как на первом этапе. Чтобы получить предмаксимальное слово уровня n, мы строим слово w_{n+1} за $n+1$ итерацию, а затем запрещаем продолжение слова $u_n w_{n+1}$ первой буквой слова u_{n+1}. Обозначим полученное предмаксимальное слово уровня n через $\bar{w}_n$. Тогда

$$\bar{w}_n = \underbrace{w_{n+1}\bar{s}_{n+1}}\underbrace{u_{n+1}w_{n+1}\bar{s}_{n+1}}, \tag{1.7}$$

где $\bar{s}_{n+1}$ – буферное слово, вставляемое подобно s_{n+1}, чтобы замкнуть маршрутный код и предотвратить появление квадратов на границе. В отличие от первого этапа конструкция (1.7) используется только один раз.

Замечание 14 *Чтобы доказать утверждение а) теоремы 11, достаточно показать существование предмаксимальных слева слов уровня n для бесконечного числа различных значений n. В самом деле, если w – предмаксимальное слева слово уровня n и $a_1 \cdots a_n w$ – максимальное слева слово, то слово $a_n w$ является предмаксимальным слева уровня $n-1$.*

Согласно этому замечанию будем строить предмаксимальные слова уровня n только для таких n, что

(*) $(n{+}1)$-ая итерация первого этапа является общей нетривиальной и $\mathsf{cwk}(u'_{n+1}) = 21\underline{2}\ldots$ (соответствует вершине 1 на рис. 1.4).

Тогда $u_{n+1} = \underline{3}1\underline{2}\ldots$. Пусть эта итерация имеет номер m среди нетривиальных итераций общего вида. Тогда будем получать буферное слово $\bar{s}_n$ из маршрутного кода $\mathsf{B}[m{+}1..2m{+}1]32332$ (если необходимо замкнуть маршрут, продолжим используемое подслово маршрутного кода слова Аршона согласно (1.6)). Маршрутный код 32332 обозначим как $\overline{C}$. Корректность конструкции (то есть то, что такая вставка приводит к бесквадратному циклическому слову и это слово является предмаксимальным) доказывается полностью аналогично лемме 10.

Лемма 11 *Циклическое слово $(\bar{v}) = (p_n\bar{w}_n)$ бесквадратно для любого n, удовлетворяющего условию (*).*

Утверждение. Слово $s'_{n+1}\bar{s}_n p_n$ бесквадратно для любого n из условия (*).

Доказательство Пусть, как и прежде, m — номер итерации среди нетривиальных итераций общего вида, на которой было построено слово w_{n+1}. Требуется доказать, что маршрутный код

$$\overline{D}_m = \mathsf{cwk}(\mathsf{A})[2..k_1]\widetilde{C}_1 \cdots \mathsf{cwk}(\mathsf{A})[3m{-}1..k_m]C'\mathsf{cwk}(\mathsf{A})[3m{-}2..6m{+}4]\overline{C}S\mathsf{cwk}(p_n)$$

кодирует бесквадратное слово. Подслово $C' = 3233233$ — это маршрутный код суффикса слова s_{n+1}, согласно Рис. 1.4, а S — продолжение маршрутного кода до замкнутого по (1.6). Напомним, что $p_n = \underline{3}1\underline{2}\ldots = \overleftarrow{\mathsf{A}[1..n]}$, причём $|\mathsf{cwk}(p_n)| < |\mathsf{cwk}(\mathsf{A})[3m{-}2..6m{+}4]|$. По Лемме 9 код

$$D_m = \mathsf{cwk}(\mathsf{A})[2..k_1]\widetilde{C}_1\mathsf{cwk}(\mathsf{A})[5..k_2]\widetilde{C}_2 \cdots \mathsf{cwk}(\mathsf{A})[3m{-}1..k_m]\widetilde{C}_m$$

соответствует бесквадратному слову, а

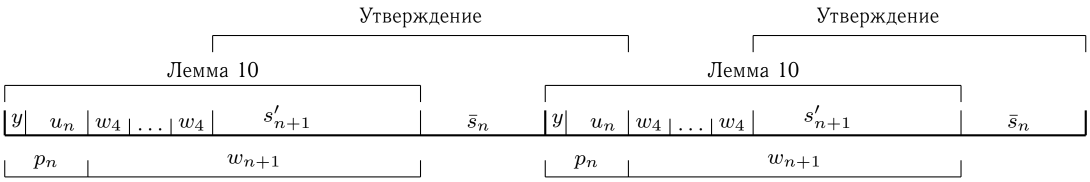
$$\overline{D}_m = D_{m-1}\mathsf{cwk}(\mathsf{A})[3m{-}1..k_m]C'\mathsf{cwk}(\mathsf{A})[3m{-}2..6m{+}4]\overline{C}S\mathsf{cwk}(p_n).$$

$\overline{D}_m$ имеет такую же структуру, что и D_m: это куски маршрутного кода слова Аршона (каждый следующий длиннее предыдущего), разделённые вставками C' из Леммы 8 и $\overline{C}$. То есть мы можем полностью воспроизвести доказательство Леммы 9.

Доказательство Леммы 11. Рисунок 1.6 иллюстрирует структуру слова $\bar{v}^2$. Сверху обозначены слова, бесквадратность которых доказана выше.

От противного, пусть для некоторого n из условия (*) $(\bar{v})$ не бесквадратно, а значит, $\bar{v}^2$ строго содержит слово, являющееся квадратом. Обозначим

Рис. 1.6: Структура слова $\bar{v}^2 = (p_n w_{n+1} \bar{s}_n)^2 = p_n \bar{w}_n$



этот квадрат как $XX = X^{(1)}X^{(2)}$. Из доказательства Леммы 10 следует, что слово $p_n w_{n+1}$ встречается в $\bar{v}^2$ только два раза. То же верно и для $\bar{s}_n$: его маршрутный код состоит из подслова $\mathsf{B}[m{+}1..2m{+}1]$, вставки $\overline{C}$ и, возможно, дополнения до замкнутого маршрута согласно (1.6), а такого подслова из B мы ранее не вставляли. Следовательно, слова $p_n w_{n+1}$ и $\bar{s}_n$ не могут быть подсловами X. Значит, есть всего три варианта расположения XX:

1. $X^{(2)}$ заканчивается в $\bar{s}_n$;
2. $X^{(2)}$ начинается в $\bar{s}_n$ и заканчивается в p_n;
3. $X^{(2)}$ начинается в p_n.

Сначала докажем, что бесквадратно не только слово $s'_{n+1}\bar{s}_n p_n$, но и $s'_{n+1}\bar{s}_n p_n w_4$. Пусть, от противного, $s'_{n+1}\bar{s}_n p_n w_4$ содержит квадрат. Понятно, что этот квадрат должен захватывать часть w_4. Но перед ним находится подслово Аршона u_n. Очевидно, что X содержит какую-то часть u_n, а она является подсловом Аршона. Значит, это подслово должно «наложиться» на подслово Аршона из некоторого s_k или $\bar{s}_n$. Но за ним следует либо ещё одно подслово Аршона, которое не может совпасть с началом w_4, либо «подавляющая Аршона» вставка (из Леммы 8 или $\overline{C}$), которая также не может совпасть с началом w_4, что легко проверяется в терминах их маршрутных кодов.

Рассмотрим первый случай. $X^{(2)}$ начинается позже, чем p_n (случай, в котором $X^{(2)}$ начинается в p_n, обсуждается ниже). $X^{(2)}$ не может содержать s'_{n+1}, так как ему негде повториться раньше вместе с префиксом $\bar{s}_n$. Значит, $X^{(2)}$ начинается внутри s'_{n+1}, то есть X не содержит w_4, следовательно, $X^{(1)}$ начинается в w_4, находящемся непосредственно перед s'_{n+1}. Значит, весь квадрат XX строго содержится в подслове $w_4 s'_{n+1}\bar{s}_n$. Применяя рассуждения, аналогичные доказательству леммы 9 (суффикс w_4, входящий в X, может играть ту же роль, что и вставки, соответствующие маршрутным кодам C'' и $\overline{C}$), приходим к противоречию.

Во втором случае «подавляющий Аршона» суффикс $\bar{s}_n$ повторяется левее уже в s_{n+1} (см. условие (*)). Но тогда период квадрата больше, чем $|\bar{s}_n p_n|$, —

снова получили противоречие.

В третьем случае $X^{(2)}$ содержит w_4, следующее за p_n. Самое близкое к нему вхождение w_4 слева находится перед s'_{n+1}. Значит, X содержит $s'_{n+1}\bar{s}_n$. Но как отмечалось выше, даже $\bar{s}_n$ не может быть подсловом X. Получаем противоречие, которое и завершает доказательство.

$\square$

Доказательство Теоремы 11. Для бесконечного числа различных значений n мы построили серию слов $\bar{w}_n$ таких, что слово $\bar{w}_n$ имеет контекст u_n длины n, но не имеет никаких контекстов большей длины: $p'_n\bar{w}_n = zu_n\bar{w}_n$ — минимальный квадрат по Лемме 10, а $p_n\bar{w}_n = yu_n\bar{w}_n$ — минимальный квадрат по Лемме 11, при этом $y \neq z \neq u_n(1)$. А значит, первое утверждение теоремы доказано.

Теперь докажем утверждение б) теоремы 11. Естественная идея состоит в следующем: чтобы получить предмаксимальное слово уровня (n, k), берём слова $\bar{w}_n$ и $\overleftarrow{\bar{w}}_k$ и соединяем их с помощью некоторого буферного слова, позволяющего избежать появления квадратов. Таким буферным словом может служить, например, подслово Аршона длины не менее $7(|\bar{w}_n| + |\overleftarrow{\bar{w}}_k|)$ (из лемм 8 и 9 следует, что слово $\bar{w}_n$ может быть продолжено подсловом Аршона вправо, а слово $\overleftarrow{\bar{w}}_k$, соответственно, влево). Слово 32332 на конце $\mathsf{cwk}(\bar{w}_n)$ меняется на 32331 по замечанию 10. Здесь нет симметрии: префикс 23323 слова $\mathsf{cwk}(\overleftarrow{\bar{w}}_k)$ не меняется, так как код Пансьё мы строим слева направо. Таким образом, мы можем взять любое подслово $\mathsf{cwk}(\mathsf{A})$, представляющее из себя произведение блоков 321 и 231 и вставить его между 32331 на конце $\mathsf{cwk}(\bar{w}_n)$ и 23323 в начале $\mathsf{cwk}(\overleftarrow{\bar{w}}_k)$, чтобы получить требуемое подслово Аршона. Благодаря выбранной длине подслова Аршона никакой период не может «перейти» через него из $\mathsf{cwk}(\bar{w}_n)$ в $\mathsf{cwk}(\overleftarrow{\bar{w}}_k)$, так как слово Аршона $(7/4)^+$-свободно. Доказательство теоремы завершено. $\square$

Глава 2

Избегаемость буквенных шаблонов в бесквадратных словах

В этой главе исследуются буквенные шаблоны длины до 6 включительно в словах языка SF и доказывается

Теорема 12 *Следующие тернарные бесквадратные буквенные шаблоны избегаемы в языке* SF*:*
а) $xyxzx$*,* $xyzxy$*;*
б) $xyxzyz$ *и все шаблоны длины 6, подсловами которых являются шаблоны из а).*
Все другие шаблоны длины $\leqslant 6$ *неизбежны в* SF.

Туэ показал [80], что любое слово вида xyz неизбежно, то есть не только шаблон xyz неизбежен, но и в любом бесквадратном слове, начиная с определённой длины, встречаются все 6 подслов вида xyz.

Рассмотрим вначале четырёхбуквенные шаблоны. Заметим, что все слова, представленные в виде буквенного шаблона, имеют одинаковый код Пансьё. Коды Пансьё четырёхбуквенных шаблонов имеют длину 2, среди них бесквадратными являются $10, 01$ и 11. Легко видеть, что благодаря свойствам бесквадратных кодовых слов такие подслова неизбежны.

Покажем теперь, какие из пяти- и шестибуквенных шаблонов неизбежны, используя свойства маршрутных кодов, и далее представим идею построения тернарных бесквадратных слов, избегающих оставшиеся шаблоны.

Прежде всего, рассмотрим все бесквадратные пятибуквенные шаблоны и их коды Пансьё.

$$\begin{array}{ccccc|ccccc|ccccc} x & y & x & z & x & x & y & x & z & y & x & y & z & x & y \\ 0 & 1 & 0 & & & 0 & 1 & 1 & & & 1 & 1 & 1 & & \end{array}$$

x	y	z	x	z	x	y	z	y	x
1	1	0			1	0	1		

По Лемме 5 буквенные шаблоны $xyxzy$ и $xyzxz$ неизбежны, т.к. 11 является запрещённым подсловом в бесквадратных маршрутных кодах. Шаблон $xyzyx$ тоже неизбежен, потому что подслова 00 и 1111 в коде Пансьё соответствуют квадратам (иными словами, в коде Пансьё бесконечного бесквадратного слова после ряда единиц обязательно должен встретиться нуль, за которым может следовать только единица). Другие два буквенных шаблона длины 5 будут избегаемыми, если существуют бесконечные бесквадратные маршрутные коды, построенные с использованием только двух букв из алфавита $\{1, 2, 3\}$ ($\{1, 2\}$ в случае $xyzxy$ и $\{2, 3\}$ в случае $xyxzx$).

Допустим, что нам удалось построить такие маршрутные коды и тем самым доказать избегаемость $xyxzy$ и $xyxzx$ в языке SF. В таком случае, мы должны будем рассмотреть только те 6-буквенные шаблоны, которые не содержат избегаемых 5-буквенных шаблонов в качестве подслов:

x	y	x	z	y	z	x	y	z	x	z	y	x	y	z	y	x	z
0	1	1	0			1	1	0	1			1	0	1	1		

С помощью рассуждений, аналогичных случаю 5-буквенных шаблонов, мы заключаем, что последние два 6-буквенных шаблона неизбежны. Первый шаблон соответствует букве 2 в маршрутном коде, а значит, если существует бесконечный бесквадратный маршрутный код, построенный только из букв 1 и 3, то $xyxzyz$ — избегаемый шаблон. Обобщая вышесказанное, Теорема 12 эквивалентна существованию бесконечных бесквадратных маршрутных кодов, содержащих только две различные буквы, причем любые, из $\{1, 2, 3\}$. Легко видеть, что если такие маршрутные коды существуют, они не содержат кубов одной буквы и квадратов второй согласно Лемме 5 (например, если эти две буквы — $\{2, 3\}$, то маршрутный код не может содержать подслова 22, поскольку его нельзя продолжить ни вправо, ни влево, сохраняя бесквадратность). Среди бесконечных бинарных слов таким свойством обладает, например, ω-слово Фибоначчи F. Мы используем его для получения маршрутных кодов.

2.1 Построение бесквадратных кодов из слов Фибоначчи

Приведём свойства слов Фибоначчи, которые понадобятся далее:

1. ω-слово Фибоначчи не содержит подслов aaa и bb (слово с такими подсловами не имеет прообраза относительно φ).

2. Если подслово ω-слова Фибоначчи имеет нетривиальные периоды, то его минимальный период — число Фибоначчи [24].

3. $\operatorname{lexp}(\mathsf{F}) = 2+\rho$, где ρ — золотое сечение. Этот супремум не достигается, таким образом, локальная экспонента любого слова Фибоначчи меньше, чем $2+\rho$ [44].

4. Если u^k — подслово F, где $u \neq \lambda, k > (2+\rho)/2$, то существует $n \geqslant 1$ такое, что u — сопряжённое к F_n и, более того, каждое вхождение u^k содержится в некотором максимальном F_n^s для некоторого $s \in [2, 2+\rho)$ [58].

5. Обозначим n-е число Фибоначчи как Φ_n (полагая $\Phi_0 = 1$, $\Phi_1 = 2$). Длина подслова F с периодом Φ_n не превосходит $\Phi_{n+1} + 2\Phi_n - 2$ [44].

Рассмотрим маршрутные коды, получаемые из слов Фибоначчи при помощи следующих подстановок:

$$\sigma_{21} : a \to \mathsf{2},\ b \to \mathsf{1} \tag{2.1a}$$

$$\sigma_{31} : a \to \mathsf{3},\ b \to \mathsf{1} \tag{2.1b}$$

$$\sigma_{32} : a \to \mathsf{3},\ b \to \mathsf{2} \tag{2.1c}$$

Будем называть такие маршрутные коды *маршрутными кодами Фибоначчи*. Обозначим как $\mathcal{F}_n$ $[\mathcal{F}]$ маршрутный код, полученный из слова $\mathsf{F_n}$ [ω-слова F, соответственно] под действием одной из этих подстановок. Когда нужно уточнить, какая именно подстановка была применена, будем писать $\mathcal{F}_n^{ij}$ $[\mathcal{F}^{ij}]$ для $\sigma_{ij}(\mathsf{F_n})$ [$\sigma_{ij}(\mathsf{F})$, соответственно].

Посмотрим, что за слова w^{21}, w^{31}, w^{32} получаются при декодировании из маршрутных кодов Фибоначчи $\mathcal{F}^{21}, \mathcal{F}^{31}, \mathcal{F}^{32}$ соответственно в предположении, что мы всегда начинаем декодирование с ab.

$$\begin{aligned}
&\mathsf{f} = abaababaabaab\ldots\\
(1)\ &\mathcal{F}^{21} = 2122121221221\ldots\\
&w^{21} = abacbcacbabcacbcabacabcbacabacbcabac\ldots\\
(2)\ &\mathcal{F}^{31} = 3133131331331\ldots\\
&w^{31} = abacbabcbacbcabcbabcabacabcacbacabacbabcabac\ldots\\
(3)\ &\mathcal{F}^{32} = 3233232332332\ldots\\
&w^{32} = abacbabcacbacabcacbabcabacbcabcbacbcabacbabcabacb\ldots
\end{aligned}$$

Комбинируя свойства слов Фибоначчи с Леммой 5, покажем, что маршрутные коды Фибоначчи соответствуют бесквадратным словам. Из следующей леммы и приведённых выше рассуждений о кодах Пансьё пяти- и шестибуквенных шаблонов немедленно следует утверждение Теоремы 12.

Лемма 12 *Маршрутные коды Фибоначчи бесквадратны.*

Доказательство. Покажем, что маршрутные коды Фибоначчи $\mathcal{F}^{ij}$ удовлетворяют условиям (а) и (б) Леммы 5 и, следовательно, декодируются в бесквадратные слова.

Используя свойство 1 слов Фибоначчи, легко понять, что маршрутные коды Фибоначчи не содержат запрещённых подслов из условия (а) Леммы 5. Значит, осталось проверить, что маршрутные коды Фибоначчи не содержат подслов вида UVU, где UV помечает замкнутый маршрут во взвешенном графе $K_{3,3}$, а $|U| = 2$.

Рассмотрим периодические подслова ω-слова Фибоначчи. Возможны два принципиально разных с точки зрения рассуждений случая:

Случай 1: Периоды не являются числами Фибоначчи. Благодаря свойству 4 мы знаем, что F не содержит периодических подслов экспоненты большей, чем $(2+\rho)/2$, период которых не является числом Фибоначчи. Для достаточно больших периодов p выполняется $(2p-2)/p > (2+\rho)/2$, значит, мы не встретим запрещённых подслов вида UVU, где $|UV| = p$. Проверим короткие периоды p такие, что p чётно (в случае нечётного p маршрут будет не замкнут), $(2p-2)/p \leqslant (2+\rho)/2$ и p не число Фибоначчи. Такие значения p — это 4, 6 и 10. Подслова таких длин в F, соответствующие замкнутым маршрутам, — это $baab$, $abaaba$, $baabaa$, $aabaab$, $abaababaab$, $baababaaba$, $aababaabab$, $ababaababa$, $babaababaa$ (для всех трёх подстановок σ_{ij}, т.к. граф на Рис. 1.3 симметричен). Проверим слова UVU в каждом случае. Очевидно, что $baabba$ нет среди подслов F. Рассмотрим подслово $abaaba$ и его циклические сдвиги. Так как $aba\,aba\,aba = \varphi(ababa)a$, последняя a должна быть префиксом образа a, в противном случае $\varphi^{-1}(abaabaaba\ldots) = ababab = \varphi(aaa)$, противоречие со свойством 1. Левая сторона проверяется аналогично. Для длины 10 $abaababaab = (abaab)^2 = \mathsf{F}_3^2$ и длина подслова с таким периодом не более 16 ($|\mathsf{F}_4| + 2|\mathsf{F}_3| - 2$, см. свойство 5). Другие перечисленные подслова длины 10 — сопряжённые с F_3^2, значит, такое же наблюдение справедливо и для них.

Случай 2: Периоды являются числами Фибоначчи. Покажем, что маршрутный код, полученный из слова F_n, не замкнут для любого n на примере подстановки (2.1a). Благодаря симметрии в $K_{3,3}$ аналогичное доказательство применимо в остальных двух случаях. Рассмотрим последовательность $\{S_n\}$ кратчайших маршрутных кодов (над $\{1,2\}$) таких, что $\mathcal{F}_n S_n$ замкнутый маршрутный код (см. Рис. 1.3). Мы хотим убедиться, что все кодовые маршруты S_n непусты. Так как $\mathcal{F}_n S_{n-2} S_{n-1} = \mathcal{F}_{n-1}\mathcal{F}_{n-2} S_{n-2} S_{n-1}$ — замкнутый маршрутный код, маршрутный код S_n получается из $S_{n-2}S_{n-1}$ удалением всех циклов.

$\mathcal{F}_0 = 2 \Rightarrow S_0 = 2$,
$\mathcal{F}_1 = 21 \Rightarrow S_1 = 12$,
$S_2 = 212$,
$S_3 = \cancel{1221}2 = 2$,
$S_4 = 21\cancel{22} = 21$,
$S_5 = \cancel{22}1 = 1$,
$S_6 = 2\cancel{11} = 2 = S_0$,
$S_7 = 12 = S_1$,
$S_8 = 212 = S_2, \ldots$.

Легко видеть, что S_n — последовательность с периодом 6 и все слова S_n непусты, отсюда следует, что маршрутный код, полученный из любого слова Фибоначчи, незамкнут. Теперь заметим, что любое слово, сопряжённое к замкнутому маршрутному коду, замкнуто. Следовательно, можно заключить, что маршрутные коды, полученные из всех слов, сопряжённых к словам Фибоначчи, незамкнуты. Отметим также, что если $\mathcal{F}_n S_n$ — замкнутый маршрут, то $\mathcal{F}_n\mathcal{F}_n S_n S_n$ также замкнут, а $S_n S_n$ замкнут тогда и только тогда, когда замкнут $\mathcal{F}_n\mathcal{F}_n$. Глядя на последовательность S_n, видим, что $S_n S_n$ замкнуто для $n = 6k, 2+6k, 3+6k, 5+6k, k \in \mathbb{N}_0$, а именно, в тех случаях, когда длины слов S_n нечётны. Отсюда $\mathcal{F}_n\mathcal{F}_n$ является замкнутым маршрутом тогда и только тогда, когда длина $\mathcal{F}_n$ нечётна.

Предположим, что $\mathcal{F}$ содержит подслово UVU из Леммы 5(b), где UV — сопряжённое к $\mathcal{F}_n\mathcal{F}_n$ для некоторого n. Тогда слово UVU имеет период Φ_n и длину $< (2+\rho)\Phi_n$ по свойству 3. Тогда длина слова UVU меньше, чем $(2|XY|-2)$ для всех достаточно больших n. Легко проверить, что будет достаточно взять $n = 5$, тогда $|XY| = 2\Phi_5 = 26$. Для меньших n (2 и 3) получаем периоды 6 и 10, которые были рассмотрены в предыдущем случае.

Таким образом, лемма доказана. □

2.2 Экспоненты слов, избегающих 5- и 6-буквенные шаблоны

Как уже неоднократно упоминалось, минимальная локальная экспонента тернарного бесконечного слова равна 7/4 (эта граница достигается, например, на слове Аршона A). Естественный вопрос состоит в том, как влияют на минимальную локальную экспоненту дополнительные запреты на буквенные шаблоны, рассматриваемые в этой главе. В этом разделе даются оценки минимальных экспонент тернарных бесквадратных слов, избегающих шаблоны $xyxzx$, $xyzxy$ и $xyxzyz$. Находя критические экспоненты построенных слов w^{ij} и получая некоторые нижние границы, мы доказываем следующую теорему.

Теорема 13 *Минимальная критическая экспонента тернарного бесквадратного слова, избегающего буквенный шаблон длины* $\leqslant 6$
(1) равна 15/8 для шаблона $xyxzx$,
(2) равна 11/6 для шаблона $xyzxy$, *и*
(3) не превосходит $1+\rho/2$ *для шаблона* $xyxzyz$, *где* ρ *— золотое сечение.*

Замечание 15 *Используя свойство слов Фибоначчи* $\mathsf{F}_n = \mathsf{F}_{n-1}\mathsf{F}_{n-2}$, *можно получить последовательность равенств:*

$$\begin{aligned}
&\mathsf{F}_n = \mathsf{F}_{n-1}\mathsf{F}_{n-2} = \\
&\mathsf{F}_{n-2}\mathsf{F}_{n-3}\mathsf{F}_{n-2} = \\
&\mathsf{F}_{n-3}\mathsf{F}_{n-4}\mathsf{F}_{n-3}\mathsf{F}_{n-3}\mathsf{F}_{n-4} = \\
&\mathsf{F}_{n-3}\mathsf{F}_{n-4}\mathsf{F}_{n-4}\mathsf{F}_{n-5}\mathsf{F}_{n-4}\ldots = \\
&\mathsf{F}_{n-3}\mathsf{F}_{n-4}\mathsf{F}_{n-4}\mathsf{F}_{n-5}\mathsf{F}_{n-5}\mathsf{F}_{n-6}\ldots = \\
&\mathsf{F}_{n-3}\mathsf{F}_{n-4}\mathsf{F}_{n-4}\mathsf{F}_{n-5}\mathsf{F}_{n-6}\mathsf{F}_{n-7}\ldots = \\
&\mathsf{F}_{n-3}\mathsf{F}_{n-4}\mathsf{F}_{n-4}\mathsf{F}_{n-4}\mathsf{F}_{n-7}\ldots
\end{aligned}$$

Значит, для любого $n \geqslant 2$ *куб* $\mathsf{F}_n\mathsf{F}_n\mathsf{F}_n$ *встречается в* F, *и для* $n \geqslant 3$ *этот куб — образ куба* $\mathsf{F}_{n-1}\mathsf{F}_{n-1}\mathsf{F}_{n-1}$. *Заметим, что слова Фибоначчи поочерёдно заканчиваются либо на* ab, *либо на* ba, *а значит, период* $|\mathsf{F}_n|$ *не может быть продолжен влево.*

Лемма 13 $\operatorname{lexp}(w^{21}) = 11/6$, $\operatorname{lexp}(w^{31}) = 1+\rho/2$, $\operatorname{lexp}(w^{32}) = 15/8$.

Доказательство. Для вычисления критических экспонент слов w^{21}, w^{31}, w^{32} нам нужно рассмотреть два типа подслов в $\mathcal{F}$, декодирующихся в периодические слова:
(1) подслова вида UVU, где $|UV| \geqslant 6$ и UV — метка замкнутого маршрута (случай $|UV| = 4$ невозможен, см. доказательство Леммы 12). Как мы видели ранее, максимальные по включению подслова такого вида в $\mathcal{F}$ имеют вид $\mathcal{F}_n\mathcal{F}_n\mathcal{F}_nZ$, где Z — префикс $\mathcal{F}_n$);
(2) подслова, декодирующиеся в слова с периодами меньше 11 (см. Замечание 3.3).

Для удобства будем полагать, что код Пансьё, соответствующий данному маршрутному коду, всегда начинается с 0 и заканчивается на 1. Например, маршрутный код $(121)^3$ соответствует кодовому слову $(0101101)^3$. В таком случае длина кода Пансьё всегда равна сумме длины маршрутного кода и суммы его цифр. Напомним, что код Пансьё длины n соответствует тернарному слову длины $n{+}2$.

1. Оценим экспоненты подслов типа (1) слов w^{ij}. Минимальный период слова u — результата декодирования из $\mathcal{F}_n\mathcal{F}_n\mathcal{F}_nZ$, где $\mathcal{F}_n\mathcal{F}_n$ — замкнутый

маршрут, — равен длине p кодового слова, соответствуюшего $\mathcal{F}_n\mathcal{F}_n$, в свою очередь $p = s + l$, где s — сумма цифр маршрутного кода $\mathcal{F}_n\mathcal{F}_n$, а $l = |\mathcal{F}_n\mathcal{F}_n|$. Длина слова u — это сумма цифр кода $\mathcal{F}_n\mathcal{F}_n\mathcal{F}_nZ$ плюс $|\mathcal{F}_n\mathcal{F}_n\mathcal{F}_nZ| + 2$.

Далее, периодическое подслово в w^{ij} с периодом, обусловленным $\mathcal{F}_n\mathcal{F}_n$ в его маршрутном коде, ещё на несколько символов длиннее, чем слово u, декодированное из $\mathcal{F}_n\mathcal{F}_n\mathcal{F}_nZ$. Рассмотрим, например, $\mathcal{F}^{21}$. Предположим, что подслову $\mathcal{F}_n\mathcal{F}_n\mathcal{F}_nZ$ предшествует 1. Так как эта 1 нарушает период $|\mathcal{F}_n|$ (см. Замечание 15), то продолжить его можно, заменив 1 на 2. А значит, в терминах кодов Пансьё период будет продолжен, если мы добавим 110 слева. Так как мы имеем вместо этого 10, период продолжается влево всего на 1 символ в кодовом слове, а значит, и на один символ в w^{21}. Заметим, что смена ролей 1 и 2 не влияет на результат. Теперь посмотрим на продолжение вправо: если следующий символ после Z в маршрутном коде — 1 [2], то в коде Пансьё имеем продолжение 010 [011, соответственно]. Отсюда период в кодовом слове продолжается ровно на 2 символа. В итоге получается, что периодическое подслово в w^{21} с периодом, порождённым вхождением $\mathcal{F}_n\mathcal{F}_n$ в маршрутном коде, имеет длину, равную сумме цифр кода $\mathcal{F}_n\mathcal{F}_n\mathcal{F}_nZ$ плюс $|\mathcal{F}_n\mathcal{F}_n\mathcal{F}_nZ| + 5$. Точно такие же рассуждения для оставшихся двух слов дают такую же константу 5 для w^{31} и константу 7 для w^{32}.

Используя свойство 5, получаем $|\mathcal{F}_n\mathcal{F}_n\mathcal{F}_nZ| \leqslant \Phi_{n+1} + 2\Phi_n - 2 = \Phi_n + \Phi_{n-1} + 2\Phi_n - 2 = 3\Phi_n + \Phi_{n-1} - 2$, т.е., Z является префиксом $\mathcal{F}_n$ длины $\leqslant \Phi_{n-1} - 2$. Количество букв a в f_n равно Φ_{n-1}, количество букв b равно Φ_{n-2}. Обозначим как α и β образы a и b соответственно под действием подстановки σ_{ij}.

Помня, что суффикс длины 2 любого f_n — это ab или ba, мы можем установить верхнюю границу локальной экспоненты периодического слова w, порождённого в w^{ij} подсловом $\mathcal{F}_n\mathcal{F}_n\mathcal{F}_nZ$ в маршрутном коде:

$$\begin{aligned}
|w| &\leqslant 3(\alpha\Phi_{n-1} + \beta\Phi_{n-2} + \Phi_n) + \alpha(\Phi_{n-2} - 1) + \beta(\Phi_{n-3} - 1) + (\Phi_{n-1} - 2) + m, \\
&\qquad\qquad \text{где } m = 5 \text{ для } w^{21}, w^{31}, m = 7 \text{ для } w^{32}, \\
\operatorname{lexp}(w) &\leqslant \frac{3(\alpha\Phi_{n-1} + \beta\Phi_{n-2} + \Phi_n) + \alpha(\Phi_{n-2} - 1) + \beta(\Phi_{n-3} - 1) + (\Phi_{n-1} - 2) + m}{2(\alpha\Phi_{n-1} + \beta\Phi_{n-2} + \Phi_n)} \\
&= \frac{3}{2} + \frac{\alpha\Phi_{n-2} + \beta\Phi_{n-3} + \Phi_{n-1} - \alpha - \beta - 2 + m}{2(\alpha\Phi_{n-1} + \beta\Phi_{n-2} + \Phi_n)}\,.
\end{aligned}$$

Для $\alpha = 2, \beta = 1$ имеем

$$\operatorname{lexp}(w) \leqslant \frac{3}{2} + \frac{2\Phi_{n-2} + \Phi_{n-3} + \Phi_{n-1}}{2(2\Phi_{n-1} + \Phi_{n-2} + \Phi_n)} = \frac{3}{2} + \frac{\Phi_{n+1}}{2\Phi_{n+2}} = 1 + \frac{\Phi_{n+3}}{2\Phi_{n+2}}\,.$$

Эта оценка стремится к $1 + \rho/2$ при $n \to \infty$, а её максимум $29/16$ достигается при $n = 2$.

Для $\alpha = 3, \beta = 1$:

$$\operatorname{lexp}(w) \leqslant \frac{3}{2} + \frac{3\Phi_{n-2} + \Phi_{n-3} + \Phi_{n-1} - 1}{2(3\Phi_{n-1} + \Phi_{n-2} + \Phi_n)} = \frac{3}{2} + \frac{2\Phi_n - 1}{4\Phi_{n+1}} = 1 + \frac{\Phi_{n+2} - 1/2}{2\Phi_{n+1}}.$$

Эта оценка также стремится к $1+\rho/2$ при $n \to \infty$. Используя явную формулу

$$\Phi_n = \frac{1}{\sqrt{5}}\Big(\rho^n + (-1)^{n+1}\rho^{-n}\Big)$$

для чисел Фибоначчи, нетрудно убедиться, что наша оценка меньше $1 + \rho/2$ для любого $n \geqslant 2$.

Наконец, для $\alpha = 3, \beta = 2$:

$$\operatorname{lexp}(w) \leqslant \frac{3}{2} + \frac{3\Phi_{n-2} + 2\Phi_{n-3} + \Phi_{n-1}}{2(3\Phi_{n-1} + 2\Phi_{n-2} + \Phi_n)} = \frac{3}{2} + \frac{\Phi_{n+1} + \Phi_{n-1}}{2(\Phi_{n+2} + \Phi_n)} = 1 + \frac{\Phi_{n+3} + \Phi_{n+1}}{2(\Phi_{n+2} + \Phi_n)}.$$

Снова, предел этого выражения при $n \to \infty$ равен $1 + \rho/2$. С точностью до множителя $1/2$ эта дробь ведёт себя почти также, как отношение между $(n+1)$-м и n-м числами Фибоначчи: сходится к ρ и поочерёдно то больше, то меньше, чем ρ. Её максимум $9/11$ достигается при $n = 2$, давая нам верхнюю оценку $20/11$.

2. Короткие периодические подслова проверим напрямую. Благодаря Замечанию 3.3 достаточно проверить только периоды 6 и 8.

1. Маршрутный код Фибоначчи $\mathcal{F}$, полученный с помощью подстановки (2.1a) содержит подслово 1221, декодируемое в слово $a\,bacabc\,bacab\,a$ с локальной экспонентой $11/6 > 29/16$. w^{21} не имеет подслов с периодом 8, поскольку кодовые маршруты таких подслов должны содержать 33. Следовательно, локальная экспонента w^{21} равна $11/6$.

2. Маршрутный код Фибоначчи $\mathcal{F}$, полученный с помощью подстановки (2.1b) не содержит подслов 1221 и 2332, декодируемых в слова с локальными экспонентами больше, чем $1 + \rho/2$ (с периодами 6 и 8, соответственно), следовательно, локальная экспонента слова w^{31} равна $1 + \rho/2$. Заметим, что в отличие от двух других случаев, это значение никогда не достигается.

3. Маршрутный код Фибоначчи $\mathcal{F}$, полученный с помощью подстановки (2.1c) содержит подслово 2332, декодируемое в слово $a\,bacbcabc\,bacbc$ $ab\,a$ с локальной экспонентой $15/8 > 20/11$. Следовательно, локальная экспонента w^{32} равна $15/8$.

$\square$

Доказательство Теоремы 13. (1) Любое слово $w \in \mathsf{SF}$, избегающее шаблон $xyxzx$, имеет подслово, порождённое подсловом 2332 в маршрутном коде по Лемме 5, следовательно, $\operatorname{lexp}(w) \geqslant 15/8$, как было показано в доказательстве Леммы 13. Слово w^{32} служит примером того, что эта оценка точна.
(2) Заменив 2332 на 1221, w^{32} на w^{21} и $15/8$ на $11/6$ в предыдущем случае, получаем требуемое.
(3) Слово w^{31} избегает $xyxzyz$ и $\operatorname{lexp}(w^{31}) \leqslant 1 + \rho/2$. $\square$

2.3 Дальнейшие перспективы исследования буквенных шаблонов

Исследование избегаемости буквенных шаблонов в тернарных бесквадратных словах может быть расширено на бо́льшие длины шаблонов. Легко проверить, что единственный 7-буквенный шаблон, не содержащий избегаемых шаблонов меньших длин, — это $xyzxzyx$. Доказательство его избегаемости завершило бы классификацию буквенных шаблонов, избегаемых в языке SF. Этот шаблон имеет код Пансьё 11011. Значит, маршрутный код бесконечного тернарного слова, избегающего этот шаблон, не содержит подслов из множества 22, 23, 32, 33. Следовательно, после каждой 2 и 3 должна быть 1. Неясно, может ли такой маршрутный код быть построен, но нам снова нужна бинарная последовательность над алфавитом $\{12, 13\}$, а значит, предложенный подход может дать результат и в этом случае.

Глава 3

Структура префиксного дерева тернарного бесквадратного языка

В этой главе пойдёт речь о результатах, полученных в ходе изучения структуры префиксного дерева языка SF. А именно, будет доказано, что длина неветвящейся цепи (фиксированного правого контекста), порождённой словом в этом дереве, — логарифмическая от длины слова. Также получена нижняя оценка доли ветвящихся узлов в произвольной бесконечной цепи этого дерева. Более точно, основными результатами являются

Теорема 14 *Любой фиксированный контекст тернарного бесквадратного* w *имеет длину* $O(\log |w|)$.

Теорема 15 *Пусть* $w \in \mathsf{SF}$ *—* ω*-слово и* $b(n)$ *— количество ветвящихся узлов на пути* $(\lambda, w[1..n])$ *в дереве префиксного порядка языка* SF. *Тогда* $\liminf_{n\to\infty} \frac{b(n)}{n} \geqslant \frac{2}{9}$.

Пусть w — бесквадратное слово. Назовём букву $w[i]$ *фиксированной* p*-квадратом*, если $w[i{-}2p{+}1..i{-}1]y$ — p-квадрат, где y — буква, отличная и от $w[i]$, и от $w[i{-}1]$. С точки зрения структуры префиксного дерева это означает, что узел, помеченный словом $w[1..i-1]$, имеет только одного потомка. Заметим, что в любом фиксированном контексте бесквадратного слова каждая буква фиксирована каким-нибудь квадратом.

Доказательство Теоремы 14 производится в 2 этапа. Чтобы получить оценку длины фиксированного контекста слова, мы разделяем все квадраты на «короткие» и «длинные» и по отдельности находим верхнюю оценку для количества букв в фиксированном контексте, фиксированных короткими и длинными квадратами. Объединение полученных результатов позволяет доказать Теоремы 14 и 15.

3.1 Логарифмическая оценка длины фиксированного контекста

3.1.1 Короткие квадраты

Мы называем квадраты с периодом $\leqslant 17$ *короткими* (согласно Теореме 1, 17 — это наибольшее значение периода, для которого не существует минимального квадрата). Наша цель — получить верхнюю оценку количества букв в тернарном бесквадратном слове, фиксированных короткими квадратами. А именно, мы докажем следующую лемму.

Лемма 14 *Пусть $t, l \in \mathbb{N}$ и w — бесквадратное слово такое, что либо $|w| \geqslant t{+}l$, либо w бесконечно. Тогда из позиций в отрезке $[t{+}1..t{+}l]$ не более $\frac{2l+5}{3}$ букв фиксировано короткими квадратами.*

Для доказательства этой леммы мы используем кодировку Пансьё, описанную в подразделе 1.3.1. Будем рассматривать коды Пансьё всевозможных коротких квадратов. Мы найдём в них позиции, фиксированные другими короткими квадратами, и далее выделим диапазоны позиций, в которых все буквы фиксированы короткими квадратами и соответствующие им последовательности периодов этих квадратов. Оказывается, что длина такой последовательности периодов не превосходит 5 и по индукции любое слово может быть разбито на подслова, в каждом из которых доля букв, фиксированных короткими квадратами, не превосходит 2/3 за исключением, возможно, самого первого подслова.

Для всех допустимых периодов $p \leqslant 17$ существует единственный циклический код Пансьё, кодирующий бесквадратное циклическое слово длины p (см. [75], а также Теорему 1); ниже приведены эти коды:

$$\begin{array}{llllll} (111) & 3 & (01110111) & 8 & (0101110110111) & 13 \\ (0101) & 4 & (01011010111) & 11 & (010110101101011) & 15 \\ (011011) & 6 & (010110111011) & 12 & (0101101101011011) & 16 \end{array} \tag{3.1}$$

Напомним, как получить код минимального квадрата из кода циклического слова (u): нужно удвоить u или любое из сопряжённых с ним слов и удалить две последние буквы. Например, получим код Пансьё минимального квадрата u' из циклического бесквадратного кода $(u) = (01110111)$. Беря сопряжённый к u код 11011101, удваивая его и удаляя две последние буквы, получаем $u' = 11011101\,110111$. Слово u' имеет длину 14 и период 8, значит, декодированное слово будет иметь длину 16 и такой же период 8 (а значит, будет минимальным 8-квадратом). Кроме того, легко видеть, что любой 2-квадрат имеет код Пансьё 00.

Пусть $w \in \mathsf{SF}$. Если некоторая его буква $w[i]$ фиксирована p-квадратом, то подслово его кода Пансьё $\mathsf{cwd}(w)[i{-}2p{+}1..i{-}2]$ отличается от кода минимального квадрата только последней буквой. Мы говорим, что буква $\mathsf{cwd}(w)[i{-}2]$ также *фиксирована p-квадратом* и называем эти бесквадратные коды *нарушенными p-квадратами*. Ниже проиллюстрировано получение из слова u' нарушенного 8-квадрата:

$$\begin{aligned} w &= \cdots a\,b\,c\,a\,c\,b\,a\,c\,a\,b\,c\,a\,c\,b\,a\,b \cdots \\ \mathsf{cwd}(w) &= \cdots 1\,1\,0\,1\,1\,1\,0\,1\,1\,1\,0\,1\,1\,0 \cdots \end{aligned}$$

Полный список нарушенных p-квадратов (см. Таблицу 3.1) может быть получен из (3.1): нужно построить коды квадратов и изменить в них последнюю букву; если полученное слово не имеет собственного суффикса, являющегося кодом квадрата, то оно — нарушенный квадрат.

Для каждого кода Пансьё u мы определяем его *фиксирующую последовательность* $\mathsf{fix}(u)$ как слово над алфавитом $\{2, 3, 4, 6, 8, 11, 12, 13, 15, 16, ?\}$длины $|u|$ такое, что $\mathsf{fix}(u)[i] = p$, если $u[i]$ фиксирована p-квадратом, где $p \leqslant 17$, и $\mathsf{fix}(u)[i] =?$ иначе (т. е., если $u[i]$ либо не фиксирована, либо фиксирована длинным квадратом). Например,

$$\begin{aligned} u &= 1\,1\,0\,1\,1\,1\,0\,1\,1\,1\,0\,1\,1\,0 \\ \mathsf{fix}(u) &= ?\,?\,?\,2\,?\,?\,3\,2\,?\,?\,3\,2\,?\,8 \end{aligned} \tag{3.2}$$

Лемма 14 говорит в точности, что

($\sharp$) для любого кода Пансьё u и $t, l \in \mathbb{N}$ подслово $\mathsf{fix}(u)[t{+}1..t{+}l]$ содержит не более $\frac{2l+5}{3}$ чисел.

Заметим, что $\mathsf{fix}(u)[1] =?$ для любого кода u. Последовательность чисел v назовём *максимальной*, если $?v?$ — подслово какой-то фиксирующей последовательности. Количество чисел в подслове v фиксируюшей последовательности обозначим как $N(v)$. Чтобы доказать ($\sharp$), найдём сначала все максимальные последовательности.

Пусть буква $w[i]$ фиксирована p-квадратом, $u = \mathsf{cwd}(w)$, $i > 2p$. Тогда $u[i{-}2p{+}1..i{-}2]$ — нарушенный квадрат и, к тому же, $u[i{-}2p] \neq u[i{-}p]$. Иначе подслово $u[i{-}2p..i{-}3]$ кодировало бы p-квадрат. Мы будем называть такой нарушенный квадрат $u[i{-}2p{+}1..i{-}2]$ *регулярным*. Заметим, что не все нарушенные квадраты регулярны: присоединение буквы слева может привести к появлению квадрата. Список коротких регулярных нарушенных квадратов приведён в первой колонке Таблицы 3.1.

Все максимальные последовательности могут быть найдены следующим образом. Если $v = \mathsf{fix}(u)[i..j]$ — максимальная последовательность, то u содержит нарушенные квадраты из списка в Таблице 3.1, заканчивающиеся в

Таблица 3.1: Все регулярные (слева) и нерегулярные (справа) короткие нарушенные квадраты. В левой колонке буква, выделенная жирным, продолжает нарушенный квадрат влево. Коды Пансьё из списка (1.1) подчёркнуты.

$\mathbf{1}\,01$	2	$\underline{011011}0111$	6_2
$\mathbf{0}\,\underline{111}0$	3	$111\underline{01110111}010$	8_2
$\mathbf{1}\,1\underline{0101}1$	4	$011\underline{01011101011}010110$	11_1
$\mathbf{0}\,1\underline{011011}010$	6_1	$010111\underline{01011010111}011$	11_2
$\mathbf{0}\,11\underline{01110111}0110$	8_1	$111\underline{01011010111}010111$	11_3
$\mathbf{1}\,110111011\underline{010110111011}1$	12_1	$1\underline{01011010111}01011011$	11_4
$\mathbf{0}\,1011\underline{010110111011}010111$	12_2	$1\underline{010110111011}01011011$	12_4
$\mathbf{1}\,11\underline{010110111011}01011010$	12_3	$0110111\underline{011010110111}010$	12_5
$\mathbf{1}\,\underline{1011101101110}10111011010$	13_1	$0111011\underline{010110111011}011$	12_6
$\mathbf{0}\,10110111\underline{0101110110111}011$	13_2	$1\underline{0101110110111}0101110111$	13_5
$\mathbf{0}\,10111\underline{0101110110111}010110$	13_3	$0111011\underline{0111010111011}0110$	13_6
$\mathbf{0}\,11\underline{0101110110111}010111010$	13_4	$01011\underline{010110101101011}01011011$	15_2
$\mathbf{1}\,11\underline{010110101101011}01011010111$	15_1	$011011\underline{0101101101011011}01011010$	16_2
$\mathbf{1}\,11011\underline{0101101101011011}010110111$	16_1	$01011011\underline{0101101101011011}010111$	16_3

каждой позиции из отрезка $[i..j]$. Мы проверяем список в порядке возрастания длины. Обрабатывая нарушенный p-квадрат u, мы ищем в нём самую последнюю букву, которая не может быть фиксирована никаким q-квадратом ($q \leqslant p$) в любом вхождении u в бесквадратное кодовое слово. После этого мы продолжаем u вправо, пока добавляемые буквы фиксированы какими-то квадратами с периодами $\leqslant p$. Оба процесса конечны для каждого нарушенного p-квадрата, а значит, поиск даёт нам полный список максимальных последовательностей:

$$\begin{array}{lllllll} 2 & 6_2\,3\,2 & 3\,2\,11_2 & 2\,4\,12_2\,3\,2 & 2\,12_6 & 3\,2\,13_4 & 2\,15_2 \\ 3\,2 & 8_1\,2 & 2\,4\,11_3\,3\,2 & 2\,12_3\,2\,4 & 2\,13_1\,2\,4 & 13_5\,3\,2 & 16_1\,3\,2 \\ 2\,4 & 3\,2\,8_2\,2\,4 & 2\,11_4 & 2\,12_4 & 3\,2\,13_2\,2\,4 & 13_6 & 2\,16_2\,2\,4 \\ 2\,6_1\,2\,4 & 2\,4\,11_1\,2 & 12_1\,3\,2 & 3\,2\,12_5\,2\,4 & 2\,4\,13_3\,2 & 2\,4\,15_1\,3\,2 & 2\,4\,16_3\,3\,2 \end{array} \tag{3.3}$$

Рассуждения похожи для всех кодов коротких регулярных нарушенных квадратов, так что мы рассмотрим только один из них — 12_2:

$$u = 1011010110111011\underline{01\overline{01}1}1 \leftarrow 01$$

Мы видим два нарушенных квадрата (выделены чертами сверху и снизу), оканчивающихся в двух предпоследних позициях u, но ни один нарушенный квадрат не соответствует позиции перед этими двумя. Далее, u имеет фиксированный контекст 01, что добавляет нам ещё два числа в максимальную последовательность, но следующая позиция после этих двух не может быть фиксирована никаким коротким квадратом. Таким образом, мы получаем максимальную последовательность $2\,4\,12_2\,3\,2$.

Теперь завершим доказательство ($\sharp$) (а значит, и Леммы 14). Рассмотрим бесквадратное кодовое слово u и фрагмент его фиксирующей последовательности $v = \mathsf{fix}(u)[t{+}1..t{+}l]$. Так как мы хотим ограничить $N(v)$ сверху, то полагаем без ограничения общности, что v начинается и заканчивается максимальными последовательностями, т. е., $v = v_1?^{j_1}v_2?^{j_2}\cdots v_k$, где каждое v_i — максимальная последовательность и каждое $j_i \geqslant 1$. Докажем оценку индукцией по k. База индукции $k = 1$. В этом случае само v максимальная последовательность, $l \leqslant 5$ по (3.3), а значит, $N(v) = l = \frac{2l+l}{3} \leq \frac{2l+5}{3}$, как и требовалось.

На шаге индукции рассмотрим v_k. Если $|v_k| \leqslant 2$, то слово $v' = v_1?^{j_1}v_2?^{j_2}\cdots v_{k-1}$ удовлетворяет $N(v') = N(v) - 2$ и $|v'| \leqslant l-3$. По предположению индукции $N(v) = 2 + N(v') \leqslant 2 + \frac{2(l-3)+5}{3} = \frac{2l+5}{3}$. Пусть $|v_k| > 2$. Девятнадцать возможных значений v_k удовлетворяют этому условию, см. (3.3). Каждое из них кодирует некоторый суффикс кода Пансьё $u[1..l]$ (или всё слово $u[1..l]$ целиком при условии нерегулярного нарушенного квадрата). Из этого подслова u можно восстановить несколько предыдущих элементов слова v, используя два правила: (1) если u содержит нарушенный p-квадрат из Таблицы 3.1, поставим p в соответствие последней позиции этого подслова; (2) если никакое слово из Таблицы 3.1 не может заканчиваться в данной позиции u, поставим ей в соответствие вопросительный знак. Цель — восстановить подслово фиксирующей последовательности, которое кончается на v_k и имеет не более $2/3$ позиций, занятых числами. Результаты собраны в таблице 3.2. Цель достигается для всех слов v_k, кроме двух; в этих двух особенных случаях восклицательные знаки обозначают позиции, где нам нужны вопросительные знаки для получения требуемой границы $2/3$, но гарантировать их мы не можем.

Если $v_k = 2\,6_1\,2\,4$ [$v_k = 2\,4\,15_1\,3\,2$], то есть два нарушенных квадрата, 11_4 и 15_2 [один 11_1, соответственно], которые могут кончаться в позиции, где стоит “!”. Все эти квадраты нерегулярны, значит, нам известно целиком все слово $u[1..l]$ в обоих случаях, и мы доказываем ($\sharp$) прямой проверкой соответствующих строк таблицы 3.2.

После исключения особых случаев, восклицательные знаки можно заменить на вопросительные. Пусть $\hat{v}$ — восстановленная часть $\mathsf{fix}(u)$. Если $\hat{v}$ содержит v, то неравенство $N(v) \leqslant \frac{2l+5}{3}$ проверяется непосредственно. Иначе $\hat{v} = ?^{j_i}v_{i+1}j^{j_{i+1}}\cdots v_k$. Пусть $v = v'\hat{v}$. Имеем $N(\hat{v}) \leqslant 2|\hat{v}|/3$ по Таблице 3.2 и $N(v') \leqslant \frac{2|v'|+5}{3}$ по предположению индукции. Следовательно, $N(v) \leqslant \frac{2l+5}{3}$. Шаг индукции доказан, а значит, и ($\sharp$) и Лемма 14.

Таблица 3.2: Восстановленные фрагменты фиксирующих последовательностей. Максимальная последовательность (1-я колонка) определяет подслово кода Пансьё (2-я колонка, верх), которое, в свою очередь, определяет подслово фиксирующей последовательности (2-я колонка, низ); отношение количества чисел в последнем к его длине (3-я колонка) меньше или равно 2/3. В двух особых случаях '!' может быть заменён либо на период нерегулярного нарушенного квадрата, либо на '?'.

Макс. последовательность	Код Пансьё и суффикс его фиксирующей последовательности	Доля коротких квадратов
2 6_1 2 4	0 1 0 1 1 0 1 1 0 1 0 1 1 $! \in \{11_4, 15_2, ?\}$ **!** ? 2 6 2 4	**2/3**
6_2 3 2	0 1 1 0 1 1 0 1 1 1 0 1 ? 2 ? ? 2 ? ? 2 ? 6 3 2	1/2
3 2 8_2 2 4	1 1 1 0 1 1 1 0 1 1 1 0 1 0 1 1 ? ? ? 3 2 ? ? 3 2 ? ? 3 2 8 2 4	9/16
2 4 11_1 2	0 1 1 0 1 0 1 1 1 0 1 0 1 1 0 1 0 1 1 0 1 ? 2 ? ? 2 ? 2 4 ? 3 2 ? 2 4 ? 2 ? 2 4 11 2	13/21
3 2 11_2	0 1 0 1 1 1 0 1 0 1 1 0 1 0 1 1 1 0 1 1 ? 2 ? 2 4 ? 3 2 ? 2 4 ? 2 ? 2 4 ? 3 2 11	13/22
2 4 11_3 3 2	1 1 1 0 1 0 1 1 0 1 0 1 1 1 0 1 0 1 1 1 0 1 ? ? ? 3 2 ? 2 4 ? 2 ? 2 4 ? 3 2 ? 2 4 11 3 2	7/12
12_1 3 2	1 1 1 0 1 1 1 0 1 1 0 1 0 1 1 0 1 1 1 0 1 1 1 0 1 ? ? 3 2 ? 12 3 2	5/8
2 4 12_2 3 2	0 1 0 1 1 0 1 0 1 1 0 1 1 1 0 1 1 0 1 0 1 1 1 0 1 ? ? 2 ? 2 4 12 3 2	2/3
2 12_3 2 4	1 1 1 0 1 0 1 1 0 1 1 1 0 1 1 0 1 0 1 1 0 1 0 1 1 ? ? 2 ? 2 4 ? 2 12 2 4	7/11
3 2 12_5 2 4	0 1 1 0 1 1 1 0 1 1 0 1 0 1 1 0 1 1 1 0 1 0 1 1 ? 2 ? ? 2 ? ? 3 2 ? ? 2 ? 2 4 ? 2 ? ? 3 2 12 2 4	13/24
2 13_1 2 4	1 1 0 1 1 1 0 1 1 0 1 1 1 0 1 0 1 1 1 0 1 1 0 1 0 1 1 ? ? 2 13 2 4	2/3
3 2 13_2	0 1 0 1 1 0 1 1 1 0 1 0 1 1 1 0 1 1 0 1 1 1 0 1 1 ? ? 3 2 13	3/5
2 4 13_3 2	0 1 0 1 1 1 0 1 0 1 1 1 0 1 1 0 1 1 1 0 1 0 1 1 0 1 ? ? 3 2 ? 2 4 13 2	2/3
3 2 13_4 2 4	0 1 1 0 1 0 1 1 1 0 1 1 0 1 1 1 0 1 0 1 1 1 0 1 0 1 1 ? 2 ? ? 3 2 ? 2 4 ? 3 2 13 2 4	5/8
13_5 3 2	1 0 1 0 1 1 1 0 1 1 0 1 1 1 0 1 0 1 1 1 0 1 1 1 0 1 ? ? 2 ? 2 4 ? 3 2 ? ? 2 ? ? 3 2 ? 2 4 ? 3 2 ? 13 3 2	15/26
2 4 15_1 3 2	1 1 1 0 1 0 1 1 0 1 0 1 1 0 1 0 1 1 0 1 0 1 1 0 1 0 1 1 1 0 1 $! \in \{11_1, ?\}$ **!** 2 ? 2 4 ? 2 ? 2 4 ? 2 ? 2 4 15 3 2	**2/3**
16_1 3 2	1 1 0 1 1 0 1 0 1 1 0 1 1 0 1 0 1 1 0 1 1 0 1 0 1 1 0 1 1 1 0 1 ? 2 ? 16 3 2	2/3
2 16_2 2 4	0 1 1 0 1 1 0 1 0 1 1 0 1 1 0 1 0 1 1 0 1 1 0 1 0 1 1 0 1 0 1 1 ? 2 ? ? 2 ? ? 2 ? 2 4 ? 2 ? ? 2 ? 2 4 ? 2 ? ? 2 ? 2 4 ? 2 16 2 4	17/32
2 4 16_3 3 2	0 1 0 1 1 0 1 1 0 1 0 1 1 0 1 1 0 1 0 1 1 0 1 1 0 1 0 1 1 1 0 1 ? 2 ? 2 4 ? 2 ? ? 2 ? 2 4 ? 2 ? ? 2 ? 2 4 ? 2 ? ? 2 ? 2 4 16 3 2	9/16
2 11_4	1 0 1 0 1 1 0 1 0 1 1 1 0 1 0 1 1 0 1 1 ? ? 2 ? 2 4 ? 2 ? 2 4 ? 3 2 ? 2 4 ? 2 11	3/5
2 15_2	0 1 0 1 1 0 1 0 1 1 0 1 0 1 1 0 1 0 1 1 0 1 0 1 1 0 1 1 ? 2 ? 2 4 ? 2 ? 2 4 ? 2 ? 2 4 ? 2 ? 2 4 ? 2 ? 2 4 ? 2 15	17/28

3.1.2 Длинные квадраты

Лемма 15 *Пусть $t, l \geqslant 1$, $p, q \geqslant 2$ — натуральные числа, w — слово длины $t{+}l$ такое, что подслово $w[1..t{+}l{-}1]$ бесквадратно, буква $w[t]$ фиксирована p-квадратом, и w заканчивается на q-квадрат. Тогда выполняется одно из следующих условий:*
1) $2q \geqslant 2p + l$ и $q \geqslant 2p$;

2) $p + l < 2q < 2p + l$ *и* $q \leqslant l$*;*
3) $2q \leqslant p + l$.

Доказательство. Пусть $UxUy$ — суффикс $w[1..t]$ длины $2p$, где $y = w[t] \neq w[t{-}p] = x$, а XX — суффикс w длины $2q$. Условие $2q \geqslant 2p + l$ значит, что XX начинается в w вместе с $UxUy$ или раньше. Возможны два случая в зависимости от того, где начинается правое подслово X — раньше или позже, чем Uy (см. Рис. 3.1 а,б; если второе X начинается вне $UxUy$, то требуемое неравенство $q \geqslant 2p$ выполняется тривиально). Ситуация, когда эти два слова начинаются одновременно, отнесена к случаю **б** (то есть, в обоих случаях подслово u может быть пустым).

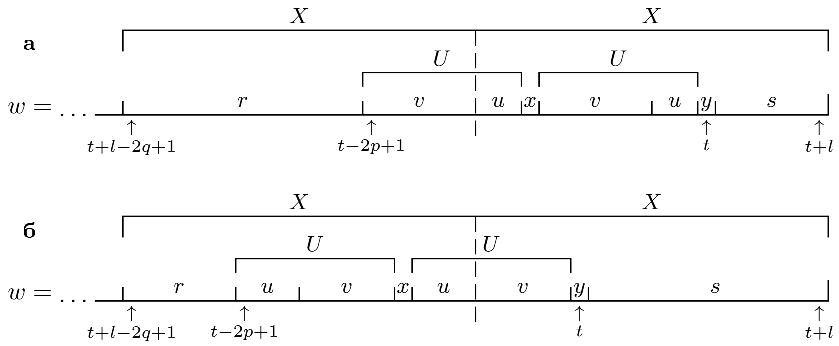


Рис. 3.1: Взаимное расположение подслов при условии $2q \geqslant 2p + l$ (Лемма 15).

В случае **а** имеем $U = vu$, $X = rv = uxvuys$. Если $|r| \leqslant |uxvu|$, то подслово U первого вхождения X находится непосредственно перед или перекрывает вхождение U в середине XX, а это противоречит бесквадратности $w[1..t{+}l{-}1]$. Следовательно, $|r| \geqslant |uxvuy|$. В итоге $q = |X| = |r| + |v| \geqslant |uxvuy| + |v| = |UxUy| = 2p$, что и требовалось.

Случай **б** подобен предыдущему: $U = uv$, $X = ruvxu = vys$, и если $|r| \leqslant |v|$, то подслово U из второго X находится непосредственно перед или перекрывает вхождение U в середине XX. Значит, $|r| \geqslant |vy|$, откуда $q = |X| = |r| + |uvxu| \geqslant |vy| + |uvxu| = |UxUy| = 2p$.

Перейдём теперь к условию 2. Нам нужно показать, что если XX начинается внутри Ux, то первое вхождение X заканчивается не раньше, чем $UxUy$. Чтобы это сделать, исключим единственную альтернативу такому расположению подслов: первое X заканчивается внутри Uy. Рассмотрим этот вариант. Если $p = q$, то $x = y$, что невозможно. Значит, нужно рассмотреть случаи, когда $q < p$ и $q > p$ (см. Рис. 3.2 а,б). В случае **а** имеем $U = ruv$, $X = vxr = uvys$. Отсюда $UxUy = ru\,vxr\,uvy = ru\,uvys\,uvy$, противоречие с бесквадратностью. В случае **б** X начинается и кончается вместе с u, а значит, XX содержит uu, т.е. не является минимальным квадратом. Это противоречие завершает доказательство. □

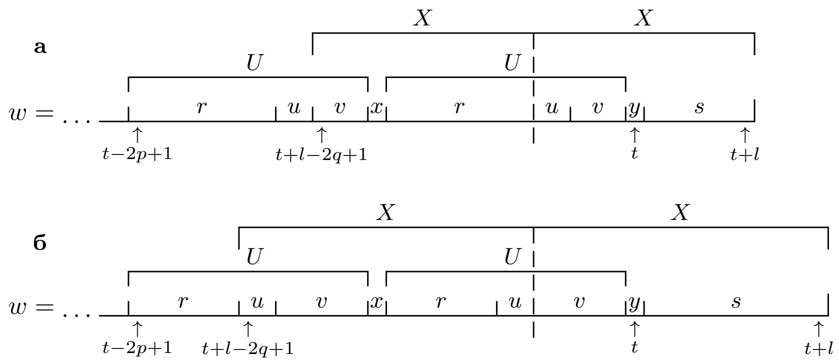


Рис. 3.2: Взаимное расположение подслов при условии $p{+}l < 2q < 2p{+}l$ (Лемма 15).

План оценки количества букв, фиксированных длинными квадратами, таков. Будем использовать жадный алгоритм: двигаясь слева направо по позициям фиксированного контекста, нужно использовать период сразу же, как только он становится доступным, т. е. перестаёт попадать в запрещённую область (корректность такого подхода доказана в Лемме 17). Из этого способа размещения фиксированных позиций выводится утверждение о том, что для любого $p > 17$ доля позиций в фиксированном контексте бесквадратного слова, зафиксированная квадратами с периодами из диапазона $[p, \ldots, 2p-1]$, — не более $1/p$. Далее нужно разбить весь диапазон возможных периодов на логарифмическое количество интервалов вида $[p, \ldots, 2p-1]$. Суммирование по интервалам даёт верхнюю оценку $l/9 + O(\log n)$ на количество позиций, фиксированных длинными квадратами в фиксированном контексте длины l слова длины n, причём можно показать, что $l \leqslant n$.

Лемма 16 *Пусть $w \in \mathsf{SF}$, $t > 0$ и буква $w[t]$ фиксирована p-квадратом. Если для некоторого $t' > t$ буква $w[t']$ фиксирована q-квадратом, то точка (t', q) на Евклидовой плоскости находится за пределами многоугольника, ограниченного прямыми:*

$$y = 2p, \quad y = x - t, \quad y = \frac{x + p - t}{2}\,, \tag{3.4}$$

где x — «временна́я» ось и y — ось периодов.

Утверждение Леммы проиллюстрировано на Рис. 3.3; недопустимые варианты для точки (t', q) помечены крестиками. В дальнейшем многоугольник, ограниченный прямыми (3.4), будем называть (t, p)-*многоугольником*, а прямые назовём (3.4) его *верхней*, *правой* и *нижней границами*, соответственно.

Доказательство. Три границы (t, p)-многоугольника соответствуют трём ограничениям на расположение квадратов из Леммы 15 для $l = t' - t$. Рассмотрим q-квадрат XX, фиксирующий букву $w[t']$. Тогда $XX =$

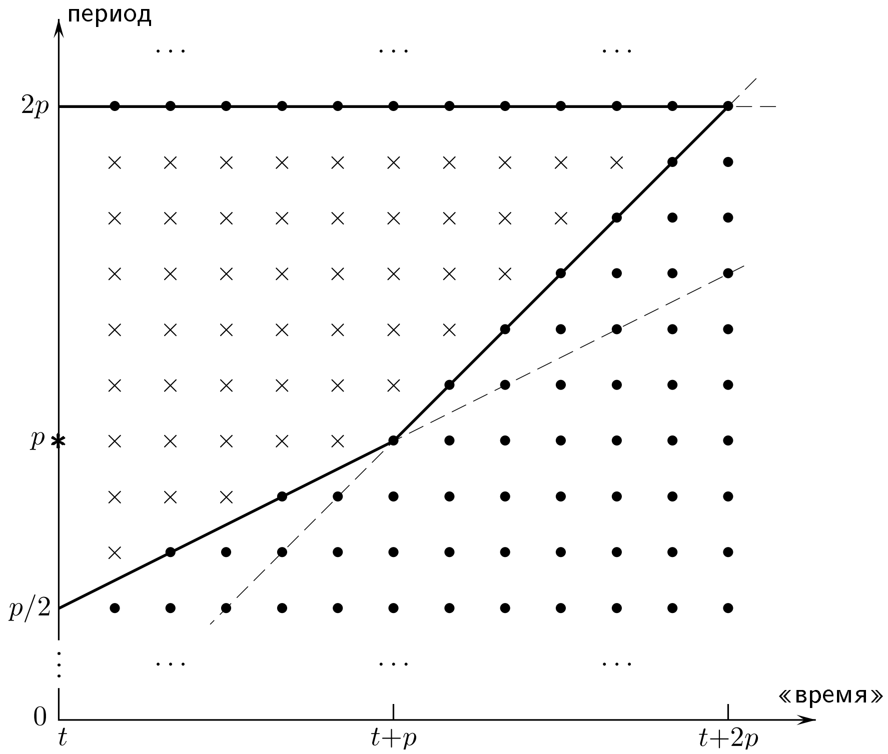


Рис. 3.3: Ограничение на фиксирование букв квадратами после того, как буква $w[t]$ была фиксирована p-квадратом. Рисунок соответствует случаю $p = 6$. Допустимые [соответственно, недопустимые] пары (**время, период**) отмечены точками [соотв., крестиками]. Границы многоугольника недопустимых фиксирующих квадратов построены по (3.4).

$(w[t'-2q+1..t'-q])^2$, а буква $w[t]$ фиксирована квадратом $(w[t-2p+1..t-p])^2$. Если $t'-2q+1 \leqslant t-2p+1$, то $q \geqslant 2p$ по Лемме 15(1), значит, точка (t', q) находится не ниже верхней границы. Если $t-2p+1 < t'-2q+1 < t-p+1$, то $t'-q \geqslant t$ по Лемме 15(2), значит, (t', q) не может быть левее правой границы. Наконец, если $t'-2q+1 \geqslant t-p+1$, (случай 3 Леммы 15), то $q \leqslant \frac{t'+p-t}{2}$, т.е., (t', q) не выше нижней границы. Таким образом, точка (t', q) не может находится внутри (t, p)-многоугольника во всех трёх случаях. □

Пусть $p \in \mathbb{N}$ и $p \geqslant 2$. Последовательность $(t_0, q_0), (t_1, q_1), \dots, (t_s, q_s)$ пар натуральных чисел назовём *p-расписанием*, если

- $t_0 < t_1 < \dots < t_s$, $p \leqslant q_i < 2p$ для всех $i = 0, \dots, s$;

- для любого $j < i$ точка (t_i, q_i) находится вне (t_j, q_j)-многоугольника.

Для удобства мы по-прежнему будем называть t и q «временем» и «периодом», соответственно. Заметим, что любая подпоследовательность p-расписания также является p-расписанием.

Лемма 17 *Пусть* $(t_0, q_0), \ldots, (t_s, q_s)$ *— p-расписание. Тогда* $t_s - t_0 > (s-1)p$.

Доказательство. Мы называем p-расписание $(t_0, q_0), \ldots, (y_s, q_s)$ *плотным*, если для любого $i > 0$ точка (t_i-1, q_i) находится внутри (t_j, q_j)-многоугольника для некоторого $j < i$. Следовательно, в плотном расписании каждый период q_i используется сразу, как только он становится доступным. Если расписание $(t_0, q_0), \ldots, (t_s, q_s)$ не плотно, то для некоторого i точка (t_i-1, q_i) находится снаружи многоугольников, порождённых предыдущими точками. Следовательно,

$$(t_0, q_0), \ldots, (t_{i-1}, q_{i-1}), (t_i-1, q_i), (t_{i+1}, q_{i+1}), \ldots, (t_s, q_s)$$

является расписанием. Действительно, для любого $j > i$ точка (t_j, q_j) находится вне многоугольников, соответствующих предыдущим точкам, так как один из этих многоугольников сдвинут на единицу влево, в то время как все остальные многоугольники остаются на месте. Используя эту процедуру «уплотнения» многократно, получаем следующее утверждение:

(∗) для любого p-расписания $(t_0, q_0), (t_1, q_1), \ldots, (t_s, q_s)$ существует плотное p-расписание $(t_0, q_0), (t'_1, q_1), \ldots, (t'_s, q_s)$ с такой же последовательностью периодов и такое, что $t'_s \leqslant t_s$.

Согласно (∗), достаточно проверить утверждение леммы для плотных p-расписаний, и далее все рассматриваемые p-расписания полагаются плотными. Заметим, что верхняя граница любого (t_i, q_i)-многоугольника находится вне отрезка $[p..2p-1]$, так что мы рассматриваем только две оставшиеся границы. Более пристальный взгляд на (3.4) и Рис. 3.3 позволяет сделать важное наблюдение: границы (t_{i+1}, q_{i+1})-многоугольника никогда не пересекают границ (t_i, q_i)-многоугольника. В самом деле, правая граница смещается вправо со временем независимо от периода, и нижняя граница также смещается вправо: для (t_i, q_i)-многоугольника прямая, образующая нижнюю границу, принимает значение $y = p$ при некотором $x \leqslant t_{i+1}$, а для (t_{i+1}, q_{i+1})-многоугольника такая прямая принимает это значение при $x > t_{i+1}$. Из этого наблюдения следует ещё одно полезное свойство: для любого $i > 0$ точка (t_{i+1}, q_{i+1}) плотного p-расписания лежит *на границе* (t_i, q_i)*-многоугольника*. Если $q_{i+1} \geqslant q_i$ $[q_{i+1} \leqslant q_i]$, то это правая [соответственно, нижняя] граница. Значит, по (3.4) имеем

$$t_{i+1} = \begin{cases} t_i + q_{i+1}, & \text{если } q_{i+1} \geqslant q_i, \\ t_i + 2q_{i+1} - q_i, & \text{если } q_{i+1} \leqslant q_i. \end{cases} \tag{3.5}$$

Теперь рассмотрим три последовательные точки p-расписания: $(t_i, q_i), (t_{i+1}, q_{i+1})$, (t_{i+2}, q_{i+2}). Покажем, что точку (t_{i+1}, q_{i+1}) можно заменить на точку (t', p), где $t' = t_i + 2p - q_i$, и новая последовательность останется p-расписанием (не обязательно плотным, но его всегда можно уплотнить). По (3.5) точка (t', p) лежит на границе (t_i, q_i)-многоугольника. Благодаря тому, что границы многоугольников, соответствующих разным точкам p-расписания, не пересекаются, достаточно показать, что точка (t'', q_{i+2}) на границе (t', p)-многоугольника удовлетворяет неравенству $t'' \leqslant t_{i+2}$. Согласно (3.5), $t'' = t' + q_{i+2} = t_i + q_{i+2} - q_i + 2p$. Для t_{i+2} требуется рассмотреть 4 случая:

$$\begin{array}{lll}
q_i \leqslant q_{i+1} \leqslant q_{i+2}: & t_{i+2} = t_i + q_{i+2} + q_{i+1} \geqslant t'', & \text{т.к. } q_i + q_{i+1} \geqslant 2p; \\
q_i \leqslant q_{i+1} \geqslant q_{i+2}: & t_{i+2} = t_i + 2q_{i+2} \geqslant t'', & \text{т.к. } q_i + q_{i+2} \geqslant 2p; \\
q_i \geqslant q_{i+1} \leqslant q_{i+2}: & t_{i+2} = t_i + q_{i+2} + 2q_{i+1} - q_i \geqslant t'', & \text{т.к. } q_{i+1} \geqslant p; \\
q_i \geqslant q_{i+1} \geqslant q_{i+2}: & t_{i+2} = t_i + 2q_{i+2} + q_{i+1} - q_i \geqslant t'', & \text{т.к. } q_{i+1} + q_{i+2} \geqslant 2p.
\end{array}$$

В результате мы доказали, что предложенная замена действительно снова даёт p-расписание. Многократно используя такие замены и уплотняя p-расписание после каждой замены, мы в конечном счёте заменим все периоды $q_1, \ldots, q_{s-1}$ на p. После этого мы также заменим q_s на p (это переместит точку на границе многоугольника ниже, то есть за пределы многоугольника) и уплотним получившееся p-расписание. В итоге, имеем плотное p-расписание вида

$$(t_0, q_0), (\hat{t}_1, p), \ldots, (\hat{t}_s, p),$$

где $\hat{t}_s \leqslant t_s$. По (3.5) $t_{i+1} = t_i + p$ для любого $i \geqslant 1$. Значит, $t_s - t_0 > \hat{t}_s - \hat{t}_1 = (s-1)p$. □

Из Лемм 16 и 17 сразу получаем оценку на количество позиций, фиксированных квадратами из данного отрезка:

Лемма 18 *Пусть $t, l \geqslant 1$, $p \geqslant 2$ — натуральные числа, $w \in$ SF такое, что либо $|w| \geqslant t{+}l$, либо w бесконечно. Тогда отрезок $[t{+}1..t{+}l]$ содержит не более $1 + \lceil (l{-}1)/p \rceil$ позиций, в которых буквы w фиксированы квадратами с периодами из отрезка $[p..2p{-}1]$.*

Доказательство. Возьмём произвольное слово w, удовлетворяющее условию леммы. Рассмотрим множество позиций $\{t_0, t_1, \ldots, t_k\}$ подслова $w[t{+}1..t{+}l]$, фиксированных квадратами с периодами $\{q_0, \ldots, q_k\}$, где $q_i \in [p..2p{-}1]$ для любого $i \in \{0, \ldots, k\}$. Очевидно, что $(t_0, q_0), \ldots, (t_k, q_k)$ — p-расписание. Тогда, по Лемме 17, $l > t_k - t_0 > (k-1)p$, откуда $l - 2 \geqslant (k-1)p$. Тогда $k - 1 \leqslant \frac{l-1}{p} - \frac{1}{p}$, а всего позиций, фиксированных квадратами с периодами из отрезка $[p..2p{-}1]$ в рассматриваемом подслове — $k+1$. Следовательно, их не больше, чем $\frac{l-1}{p} - \frac{1}{p} + 2$. Так как $k+1$ — целое, а $1 - \frac{1}{p} < 1$, имеем $k+1 \leqslant \lceil (l-1)/p \rceil + 1$. Лемма доказана. □

Доказательство Теоремы 14. Пусть v — фиксированный контекст w, $|w| = n$, $|v| = l$. Тогда бесквадратное слово wv кончается на l фиксированных букв. По Лемме 14 не более $\frac{2l+5}{3}$ из этих l позиций фиксированы короткими квадратами (с периодами $\leqslant 17$). Разобьём отрезок $[18..\lfloor \frac{n+l}{2} \rfloor]$ всех возможных длинных периодов на отрезки вида $[p..2p-1]$, начиная слева (последний может быть неполным). Количество отрезков в разбиении — минимальное k такое, что $2^k \cdot 18 - 1 \geq \lfloor \frac{n+l}{2} \rfloor$, т.е. $k = \lceil \log \frac{1}{18} \lfloor \frac{n+l+2}{2} \rfloor \rceil$. Согласно Лемме 18 квадраты с периодами из i-го отрезка фиксируют не более $2 + \frac{l-2}{18 \cdot 2^{i-1}}$ букв. Так как количество всех фиксированных букв — l, имеем

$$l \leqslant \frac{2l}{3} + \frac{5}{3} + 2k + \frac{l-2}{18}\Big(\frac{1-1/2^k}{1-1/2}\Big) = \frac{7l}{9} + \frac{13}{9} + 2k - \frac{l-2}{18 \cdot 2^{k-1}}\,. \tag{3.6}$$

Заметим, что $k \leqslant \lceil \log \frac{n+l+2}{36} \rceil$, $2^{k-1} < \frac{n+l}{36}$. Тогда из (3.6) получаем

$$\frac{2l}{9} < 2\Big\lceil \log \frac{n+l+2}{36} \Big\rceil + \frac{13}{9} - \frac{2(l-2)}{n+l} \quad \text{, т.е.} \quad l < 9\Big\lceil \log \frac{n+l+2}{36} \Big\rceil + \frac{13}{2} - \frac{9(l-2)}{n+l}. \tag{3.7}$$

Если $l \geqslant n$, то из (3.7) следует $l < 9\lceil \log \frac{l+1}{18} \rceil + 2 + \frac{9}{l}$, но это неравенство очевидно не имеет решений в натуральных числах. Значит, $l < n$ для любого n, и из (3.7) имеем

$$l < 9\Big\lceil \log \frac{n}{18} \Big\rceil + \frac{13}{2} < 9 \log \frac{n}{9} + \frac{13}{2} < 9 \log n - \frac{57}{2} + \frac{13}{2} = 9 \log n - 22,$$

откуда следует утверждение Теоремы. [1]. □

3.2 Частота ветвления префиксного дерева языка SF

Доказательство Теоремы 15. Предположим противное: $\liminf_{n\to\infty} \frac{b(n)}{n} < \frac{2}{9}$. Тогда существует $\alpha < 2/9$, $\{l_i\}_1^\infty$ такое, что $b(l_i) \leqslant \alpha l_i$. Рассмотрим $(\lambda, w[1..l_i])$-путь в дереве. Буквы , помечающие его, разбиваются на три группы: фиксированные короткими квадратами, фиксированные длинными квадратами и не фиксированные. Лемма 14, Лемма 18 и наше предположение дают верхние

[1] Для больших n оценка может быть уменьшена до $8.5 \log n - 28$ с помощью некоторых дополнительных рассуждений.

оценки на количество букв в каждой группе. Подобно (3.6), (3.7), получаем следующую формулу (напомним, что теперь $n = 0$):

$$l_i < \frac{2l_i + 5}{3} + 2\left\lceil \log \frac{l_i + 2}{36} \right\rceil + \frac{l_i - 2}{9} + \alpha l_i \,.$$

Тогда $\big(2/9 - \alpha\big) l_i < 2 \log l_i - O(1)$. Это неравенство должно выполняться для всех l_i, но имеет лишь конечное число решений в натуральных числах. Полученное противоречие завершает доказательство. □

3.3 О возможном усилении Теоремы 14

Из Теоремы 9 [70] можно получить важное следствие: высота любого конечного поддерева индекса n — $O(n^{2/3})$. Мы предполагаем, что для конечных поддеревьев справедливо гораздо более сильное утверждение.

Гипотеза 1 *В префиксном (суффиксном) дереве языка* SF *любое конечное поддерево индекса n имеет $O(\log n)$ узлов.*

В Гипотезе 1 речь идёт уже о *размере* (количестве узлов) поддерева, а не о его высоте. Результат Шелтона и Сони даёт лишь суперполиномиальную верхнюю оценку размера конечных поддеревьев. Утверждение гипотезы справедливо для конечных поддеревьев–цепей как следствие Теоремы 14. Методом полного перебора получены нижние оценки на индексы конечных поддеревьев, содержащих не более трёх узлов, см. Рис. 3.4. Эти оценки также подтверждают гипотезу: даже очень маленькие поддеревья имеют весьма большой индекс.

Назовём поддерево S с корнем w *регулярным*, если префиксное (суффиксное) дерево языка SF содержит бесконечно много поддеревьев, изоморфных S с корнями вида vw. Нерегулярные поддеревья не учитываются в асимптотических оценках, так как они возникают только близко к корню дерева всего языка, не могут быть частями бо́льших конечных поддеревьев и т.д. Например, слово $abacaba$ является корнем нерегулярного конечного поддерева с одним узлом, которое появляется в дереве языка только один раз. С другой стороны, поддерево с одним узлом, порождённое словом $abcacbabcac$, регулярно, так как его корню может предшествовать бесконечно много слов, оканчивающихся на bac, и каждое из этих слов порождает такое поддерево.

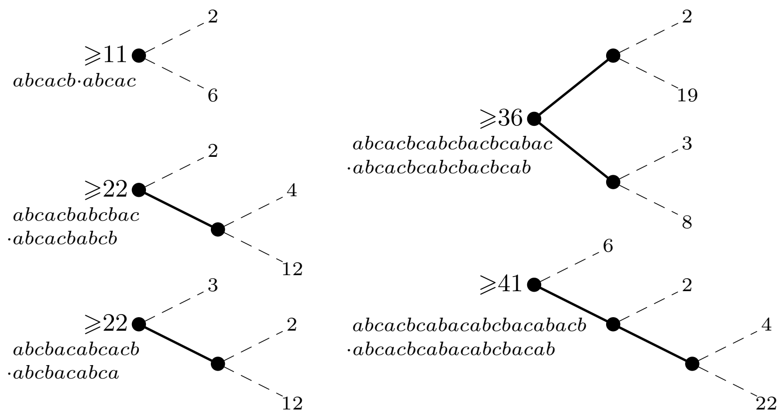


Рис. 3.4: Минимальные индексы маленьких регулярных поддеревьев. Около корней поддеревьев указаны нижняя оценка индекса и самый короткий возможный корень. Для удобства длинные корни записаны в две строки. Числа на концах пунктирных линий показывают периоды квадратов, возникающих при присоединении буквы к слову.

Заключение

Кратко подведём итоги работы по каждой из глав диссертации.

Предмаксимальные слова

Доказано существование конечных поддеревьев произвольной высоты в суффиксном (префиксном) дереве, а также конечных подграфов с цепями из суффиксных и префиксных ребёр произвольной длины в ациклическом подсловном графе языков бинарных бескубных слов CF и тернарных бесквадратных слов SF. Доказательство произведено путём построения соответствующих серий предмаксимальных слева (справа) слов для любого натурального уровня n и двусторонних предмаксимальных слов для любого уровня $(n, k) \in \mathbb{N}^2$. Таким образом, для языка SF решена открытая проблема, сформулированная Рестиво и Салеми в [61] и Аллушем и Шаллитом в [5]. При построении предмаксимальных слов использованы такие известные комбинаторные объекты, как слова Туэ–Морса, слова Аршона, а также применены свойства маршрутных кодов тернарных бесквадратных слов. Серии предмаксимальных слов построены таким образом, что длина слова растёт экспоненциально от длины его фиксированного левого (правого) контекста. Это позволяет сделать вывод о том, что для слова длины n в языках CF и SF $\log(n)$ является нижней оценкой длины его фиксированного контекста.

Избегаемость буквенных шаблонов в бесквадратных словах

Доказано, что любой буквенный шаблон длины 4 и менее неизбежен в тернарном бесквадратном языке. Основной результат главы — классификация буквенных шаблонов длины 5 и 6 по избегаемости. Для избегаемых шаблонов указан способ построения слов, избегающих эти шаблоны, основанный на маршрутных кодах, полученных из бинарных слов Фибоначчи, а также нижние оценки критических экспонент любых слов с таким свойством. Эта задача является логическим продолжением и обобщением результатов, полученных Туэ в [80]. Для завершения полной классификации буквенных шаблонов в SF остаётся доказать избегаемость одного семибуквенного шаблона. В заключении главы предложен способ, которым это можно сделать, также основанный на маршрутных кодах из слов Фибоначчи.

Структура префиксного дерева тернарного бесквадратного языка

Доказана логарифмическая оценка на длину фиксированного контекста слова w в зависимости от $|w|$ в случае тернарного бесквадратного языка SF, которая является асимптотически неулучшаемой (это следует из результатов, полученных в Главе 1). Это является существенным усилением результата, полученного Сони и Шелтоном в [70]. Как следствие получена нижняя оценка на частоту ветвления префиксного (суффиксного) дерева тернарного бесквадратного языка — $2/9$, на основании которой, учитывая индекс роста языка SF, можно сделать вывод, что дерево языка SF устроено достаточно регулярно. В заключении главы выдвинута гипотеза о том, что логарифмическая оценка справедлива не только для поддеревьев–цепей, но и для произвольных конечных поддеревьев, причём даже в смысле количества узлов в поддереве, а не его высоты. Эта гипотеза подтверждается вычислениями индексов поддеревьев небольшого размера. Ёе доказательство дало бы существенное усиление известного на данный момент результата о том, что высота любого конечного поддерева индекса $n - O(n^{2/3})$ (следствие из теоремы, доказанной в [70]). По всей видимости, такую же гипотезу можно выдвинуть и для бинарного бескубного языка.

Список литературы

[1] Аршон С. Е. Доказательство существования бесконечных n-значных асимметричных последовательностей // Мат. сб. — 1937. — Vol. 2. — P. 769–779.

[2] Клепинин А. В., Суханов Е. В. О комбинаторных свойствах последовательности Аршона // Дискретный анализ и исследование операций. Серия 1. — 1999. — Vol. 6, no. 2. — P. 23–40.

[3] Aberkane A., Currie J. D. Attainable lengths for circular binary words avoiding k-powers // Bull. Belg. Math. Soc. Simon Stevin. — 2005. — Vol. 12, no. 4. — P. 525–534.

[4] Aberkane A., Currie J. D. The Thue-Morse word contains circular $(5/2)^+$-power-free words of every length // Theoret. Comput. Sci. — 2005. — Vol. 332. — P. 573–581.

[5] Allouche J.-P., Shallit J. Automatic Sequences: Theory, Applications, Generalizations. — Cambridge University Press, 2003.

[6] Badkobeh G., Chairungsee S., Crochemore M. Hunting Redundancies in Strings // Proc. 15th Developments in Language Theory. DLT 2011. — Vol. 6795 of LNCS. — Berlin : Springer, 2011. — P. 1–14.

[7] Badkobeh G., Crochemore M. Finite-Repetition threshold for infinite ternary words // Proc. 8th Internat. Conf. Words (WORDS 2011). — Vol. 63 of EPTCS. — 2011. — P. 37–43.

[8] Badkobeh G., Crochemore M. Fewest Repetitions in Infinite Binary Words // RAIRO Inform. Théor. App. — 2012. — Vol. 46, no. 1. — P. 1–31.

[9] Badkobeh G., Crochemore M., Rao M. Finite Repetition Threshold for Large Alphabets // Proc. 14th Mons Days of Theoretical Computer Science, 2012.

[10] Bean D. A., Ehrenfeucht A., McNulty G. Avoidable patterns in strings of symbols // Pacific J. Math. — 1979. — Vol. 85. — P. 261–294.

[11] Berstel J., Séébold P. A characterization of overlap-free morphisms // Discrete Appl. Math. — 1993. — Vol. 46. — P. 275–281.

[12] Brandenburg F.-J. Uniformly growing k-th power-free homomorphisms // Theoret. Comput. Sci. — 1983. — Vol. 23. — P. 69–82.

[13] Brinkhuis J. Non-repetitive sequences on three symbols // Quart. J. Math. Oxford. — 1983. — Vol. 34. — P. 145–149.

[14] Carpi A. On Dejean's conjecture over large alphabets // Theoret. Comput. Sci. — 2007. — Vol. 385. — P. 137–151.

[15] Cassaigne J. Unavoidable binary patterns // Acta Informatica. — 1993. — Vol. 30. — P. 385–395.

[16] Cassaigne J. Counting overlap-free binary words // STACS 93, Proc. 10th Symp. Theoretical Aspects of Comp. Sci. / Ed. by P. Enjalbert, A. Finkel, K. W. Wagner. — Springer-Verlag, 1993. — Vol. 665 of LNCS. — P. 216–225.

[17] Crochemore M. Sharp characterizations of squarefree morphisms // Theoret. Comput. Sci. — 1982. — Vol. 18. — P. 221–226.

[18] Currie J. D. On the structure and extendibility of k-power free words // European J. Combinatorics. — 1995. — Vol. 16. — P. 111–124.

[19] Currie J. D. There are ternary circular square-free words of length n for $n \geq 18$ // Electronic J. Combinatorics. — 2002. — Vol. 9, no. #N10.

[20] Currie J. D., Rampersad N. Dejean's conjecture holds for $n \geq 27$ // RAIRO Inform. Théor. App. — 2009. — Vol. 43. — P. 775–778.

[21] Currie J. D., Rampersad N. Cubefree words with many squares // Discrete Math. & Theoret. Comput. Sci. — 2010. — Vol. 12, no. 3. — P. 29–34.

[22] Currie J. D., Rampersad N. Infinite words containing squares at every position // RAIRO Inform. Théor. App. — 2010. — Vol. 44. — P. 113–124.

[23] Currie J. D., Rampersad N. A proof of Dejean's conjecture // Math. Comp. — 2011. — Vol. 80. — P. 1063–1070.

[24] Currie J. D., Saari K. Least periods of factors of infinite words // RAIRO Inform. Théor. App. — 2009. — Vol. 43. — P. 165–178.

[25] Currie J. D., Shelton R. O. The set of k-power free words over Σ is empty or perfect // European J. Combinatorics. — 2003. — Vol. 24. — P. 573–580.

[26] Dejean F. Sur un théorème de Thue // J. Combin. Theory. Ser. A. — 1972. — Vol. 13. — P. 90–99.

[27] Fine N. J., Wilf H. S. Uniqueness theorems for periodic functions // Proc. Amer. Math. Soc. — 1965. — Vol. 16. — P. 109–114.

[28] Fiorenzi F., Ochem P., Vaslet E. Bounds for the generalized repetition threshold // Theoret. Comput. Sci. — 2011. — Vol. 412. — P. 2955–2963.

[29] Goralcik P., Vanicek T. Binary patterns in binary words // Internat. J. Algebra Comput. — 1991. — Vol. 1. — P. 387–391.

[30] Grimm U. Improved bounds on the number of ternary square-free words // J. Integer Sequences. — 2001. — Vol. 4, no. 01.2.7.

[31] Guglielmi N., Protasov V. Exact Computation of Joint Spectral Characteristics of Linear Operators // Foundations of Computational Mathematics. — 2013. — Vol. 13, no. 1. — P. 37–97.

[32] Ilie L., Ochem P., Shallit J. A generalization of repetition threshold // Theoret. Comput. Sci. — 2005. — Vol. 345. — P. 359–369.

[33] Jungers R. M., Protasov V. Y., Blondel V. D. Overlap-free words and spectra of matrices // Theoret. Comput. Sci. — 2009. — Vol. 410. — P. 3670–3684.

[34] Karhumäki J. On cube-free ω-words generated by binary morphisms // Discrete Appl. Math. — 1983. — Vol. 5. — P. 279–297.

[35] Keränen V. Abelian squares are avoidable on 4 letters // Proc. 19th Int'l Conf. on Automata, Languages, and Programming (ICALP) / Ed. by W. Kuich. — Springer-Verlag, 1992. — Vol. 623 of LNCS. — P. 41–52.

[36] Kobayashi Y. Enumeration of irreducible binary words // Discrete Appl. Math. — 1988. — Vol. 20. — P. 221–232.

[37] Kolpakov R. Efficient lower bounds on the number of repetition-free words // J. Integer Sequences. — 2007. — Vol. 10. — P. 07.3.2.

[38] Kolpakov R., Rao M. On the number of Dejean words over alphabets of 5, 6, 7, 8, 9 and 10 letters // Theoret. Comput. Sci. — 2011. — Vol. 412. — P. 6507–6516.

[39] Kolpakov R. M. On the number of repetition-free words // J. Appl. Ind. Math. — 2007. — Vol. 1, no. 4. — P. 453–462.

[40] Lothaire M. Combinatorics on Words. — Addison-Wesley, 1983. — Vol. 17 of Encyclopedia of Mathematics and Its Applications.

[41] Lothaire M. Algebraic Combinatorics on Words. — Cambridge University Press, 2002. — Vol. 90 of Encyclopedia of Mathematics and Its Applications.

[42] Mercaş R., Ochem P., Samsonov A., Shur A. Binary patterns in binary cube-free words: Avoidability and growth // RAIRO Inform.Théor. App. — 2014. — Vol. 48, no. 4. — P. 369–389.

[43] Mercaş R., Saarela A. 3-abelian cubes are avoidable on binary alphabets // Proc. 17th Developments in Language Theory(DLT 2013). — Vol. 7907 of LNCS. — Springer, 2013. — P. 374–383.

[44] Mignosi F., Pirillo G. Repetitions in the Fibonacci infinite word // RAIRO Inform. Théor. App. — 1992. — Vol. 26. — P. 199–204.

[45] Mohammad-Noori M., Currie J. D. Dejean's conjecture and Sturmian words // European J. Combinatorics. — 2007. — Vol. 28. — P. 876–890.

[46] Morse M. Recurrent geodesics on a surface of negative curvature // Trans. Amer. Math. Soc. — 1921. — Vol. 22. — P. 84–100.

[47] Morse M., Hedlund G. A. Symbolic dynamics // Amer. J. Math. — 1938. — Vol. 60. — P. 815–866.

[48] Moulin-Ollagnier J. Proof of Dejean's conjecture for alphabets with $5, 6, 7, 8, 9, 10$ and 11 letters // Theoret. Comput. Sci. — 1992. — Vol. 95. — P. 187–205.

[49] Ochem P. A generator of morphisms for infinite words // RAIRO Inform. Théor. App. — 2006. — Vol. 40. — P. 427–441.

[50] Pansiot J.-J. A propos d'une conjecture de F. Dejean sur les répétitions dans les mots // Discrete Appl. Math. — 1984. — Vol. 7. — P. 297–311.

[51] Petrova E. A. Avoiding letter patterns in ternary square-free words // Electronic J. Combinatorics. — 2016. — Vol. 23, no. 1. — P. #P1.18.

[52] Petrova E. A., Shur A. M. Constructing premaximal binary cube-free words of any level // Proc. 8th Internat. Conf. Words (WORDS 2011). — Vol. 63 of EPTCS. — 2011. — P. 168–178.

[53] Petrova E. A., Shur A. M. Constructing premaximal binary cube-free words of any level // Proc. 1st Russian Finnish Symp. on Discrete Mathematics, Saint-Petersburg, 2011. — 2011. — P. 67–68.

[54] Petrova E. A., Shur A. M. Constructing premaximal binary cube-free words of any level // Internat. J. Found. Comp. Sci. — 2012. — Vol. 23, no. 8. — P. 1595–1609.

[55] Petrova E. A., Shur A. M. Constructing premaximal ternary square-free words of any level // Proc. 37th Internat. Conf. on Mathematical Foundations of Computer Science. MFCS 2012. — Vol. 7464 of LNCS. — 2012. — P. 752–763.

[56] Petrova E. A., Shur A. M. Constructing premaximal ternary square-free words of any level // Proc. 2nd Russian Finnish Symp. on Discrete Mathematics, Turku, 2012. — TUCS Lecture Notes. — 2012. — P. 146–148.

[57] Petrova E. A., Shur A. M. On the tree of ternary square-free words // Proc. 10th Internat. Conf. Words (WORDS 2015). — Vol. 9304 of LNCS. — 2015. — P. 223–236.

[58] Pirillo G. Fibonacci numbers and words // Discrete Math. — 1997. — Vol. 173. — P. 197–207.

[59] Rampersad N., Shallit J., Wang M. Avoiding large squares in infinite binary words // Theoret. Comput. Sci. — 2005. — Vol. 339. — P. 19–34.

[60] Rao M. Last Cases of Dejean's Conjecture // Theoret. Comput. Sci. — 2011. — Vol. 412. — P. 3010–3018.

[61] Restivo A., Salemi S. Overlap free words on two symbols // Automata on Infinite Words. Ecole de Printemps d'Informatique Theorique, Le Mont Dore, 1984 / Ed. by M. Nivat, D. Perrin. — Vol. 192 of LNCS. — Springer-Verlag, 1985. — P. 198–206.

[62] Restivo A., Salemi S. Some decision results on non-repetitive words // Combinatorial algorithms on words / Ed. by A. Apostolico, Z. Galil. — Vol. F12 of NATO ASI series. — Springer-Verlag, 1985. — P. 289–295.

[63] Richomme G., Séébold P. Characterization of test-sets for overlap-free morphisms // Discrete Appl. Math. — 1999. — Vol. 98. — P. 151–157.

[64] Richomme G., Wlazinski F. Some results on k-power-free morphisms // Theoret. Comput. Sci. — 2002. — Vol. 273. — P. 119–142.

[65] Roth P. Every binary pattern of length six is avoidable on the two-letter alphabet // Acta Informatica. — 1992. — Vol. 29. — P. 95–107.

[66] Samsonov A. V., Shur A. M. On Abelian repetition threshold // RAIRO Inform. Théor. App. — 2012. — Vol. 46. — P. 147–163.

[67] Sapir M. V. Combinatorics on words with applications. — LITP report, 32. — 1995.

[68] Shallit J. Simultaneous avoidance of large squares and fractional powers in infinite binary words // Internat. J. Found. Comp. Sci. — 2004. — Vol. 15. — P. 317–327.

[69] Shelton R. Aperiodic words on three symbols. II // J. Reine Angew. Math. — 1981. — Vol. 327. — P. 1–11.

[70] Shelton R. O., Soni R. P. Aperiodic words on three symbols. III // J. Reine Angew. Math. — 1982. — Vol. 330. — P. 44–52.

[71] Shur A. M. Combinatorial complexity of regular languages // Proc. 3rd International Computer Science Symposium in Russia. CSR 2008. — Vol. 5010 of LNCS. — Berlin : Springer, 2008. — P. 289–301.

[72] Shur A. M. Two-sided bounds for the growth rates of power-free languages // Proc. 13th Int. Conf. on Developments in Language Theory. DLT 2009. — Vol. 5583 of LNCS. — Springer, 2009. — P. 466–477.

[73] Shur A. M. Growth of power-free languages over large alphabets // Proc. 5th International Computer Science Symposium in Russia. CSR 2010. — Vol. 6072 of LNCS. — Springer, 2010. — P. 350–361.

[74] Shur A. M. Growth rates of complexity of power-free languages // Theoret. Comput. Sci. — 2010. — Vol. 411. — P. 3209–3223.

[75] Shur A. M. On ternary square-free circular words // Electronic J. Combinatorics. — 2010. — Vol. 17, no. # R140.

[76] Shur A. M. On the existence of minimal β-powers // Proc. 14th Int. Conf. on Developments in Language Theory. DLT 2010. — Vol. 6224 of LNCS. — Springer, 2010. — P. 411–422.

[77] Shur A. M. Deciding context equivalence of binary overlap-free words in linear time // Semigroup Forum. — 2012. — Vol. 84. — P. 447–471.

[78] Shur A. M. Growth properties of power-free languages // Computer Science Review. — 2012. — Vol. 6. — P. 187–208.

[79] Thue A. Über unendliche Zeichenreihen // Norske vid. Selsk. Skr. Mat. Nat. Kl. — 1906. — Vol. 7. — P. 1–22.

[80] Thue A. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen // Norske vid. Selsk. Skr. Mat. Nat. Kl. — 1912. — Vol. 1. — P. 1–67.

[81] Tunev I. N., Shur A. M. On two stronger versions of Dejean's conjecture // Proc. 37th Internat. Conf. on Mathematical Foundations of Computer Science. MFCS 2012. — Vol. 7464 of LNCS. — 2012. — P. 801–813.