Лекция 9: Бесповторные слова

А. М. Шур

Институт математики и компьютерных наук (матмех) УрФУ

26 марта 2015

Определение

Экспонентой слова называется отношение его длины к его наименьшему периоду: $\exp(w) = |w|/\mathrm{per}(w)$.

Экспонента — это характеристика глобальной периодичности слова. Она показывает взаимное расположение самых длинных одинаковых подслов в слове, одно из которых — префикс, а другое — суффикс.

★ В дальнейшем пару одинаковых подслов в слове мы называем повтором.

Определение

Экспонентой слова называется отношение его длины к его наименьшему периоду: $\exp(w) = |w|/\mathrm{per}(w)$.

Экспонента — это характеристика глобальной периодичности слова. Она показывает взаимное расположение самых длинных одинаковых подслов в слове, одно из которых — префикс, а другое — суффикс.

★ В дальнейшем пару одинаковых подслов в слове мы называем повтором.

- $\exp(w) = 2$: повтор «вплотную» (квадрат), например, ТАРТАР;
- $\exp(w) > 2$: повтор «с перекрытием», например, АНТАНТА ($\exp = 7/3$);
- $\exp(w) < 2$: повтор «на расстоянии», например, КОЛОКОЛ ($\exp = 7/4$) или АБРАКАДАБРА ($\exp = 11/7$).

Определение

Экспонентой слова называется отношение его длины к его наименьшему периоду: $\exp(w) = |w|/\mathrm{per}(w)$.

Экспонента — это характеристика глобальной периодичности слова. Она показывает взаимное расположение самых длинных одинаковых подслов в слове, одно из которых — префикс, а другое — суффикс.

★ В дальнейшем пару одинаковых подслов в слове мы называем повтором.

- $\exp(w) = 2$: повтор «вплотную» (квадрат), например, ТАРТАР;
- $\exp(w) > 2$: повтор «с перекрытием», например, АНТАНТА ($\exp = 7/3$);
- $\exp(w) < 2$: повтор «на расстоянии», например, КОЛОКОЛ ($\exp = 7/4$) или АБРАКАДАБРА ($\exp = 11/7$).

Вопрос: каких повторов можно избежать в заданном алфавите?

Определение

Слово w называется

- β-свободным, если экспонента любого его подслова меньше β;
- β^+-свободным, если экспонента любого его подслова не больше β.

Далее β^+ трактуется как «число», покрывающее β в обычном порядке $\leqslant$ на $\mathbb{R}$, а соответствующее расширение $\mathbb{R}$ обозначается через $\hat{\mathbb{R}}$.

Пример:
- 2-свободные (бесквадратные) слова = слова, не содержащие повторов вплотную
- 2^+-свободные (overlap-free, сильно бескубные) слова = слова, не содержащие повторов с перекрытием

Определение

Слово w называется

- β-свободным, если экспонента любого его подслова меньше β;
- β^+-свободным, если экспонента любого его подслова не больше β.

Далее β^+ трактуется как «число», покрывающее β в обычном порядке $\leqslant$ на $\mathbb{R}$, а соответствующее расширение $\mathbb{R}$ обозначается через $\hat{\mathbb{R}}$.

Пример:
- 2-свободные (бесквадратные) слова = слова, не содержащие повторов вплотную
- 2^+-свободные (overlap-free, сильно бескубные) слова = слова, не содержащие повторов с перекрытием

Определение

Экспонента $\beta \in \hat{\mathbb{R}}$ называется

- k-избегаемой, если существует бесконечно много β-свободных k-ичных слов
- k-неизбежной в противном случае

Свобода и неизбежность - 2

Замечание

Если экспонента $\beta \in \hat{\mathbb{R}}$ k-избегаема, то любая экспонента $\beta' > \beta$ тоже k-избегаема.

Любое β-свободное слово является β'-свободным. $\square$

Замечание

Если экспонента $\beta \in \hat{\mathbb{R}}$ k-избегаема, то любая экспонента $\beta' > \beta$ тоже k-избегаема.

Любое β-свободное слово является β'-свободным. $\square$

Замечание

Экспонента $\beta \in \hat{\mathbb{R}}$ k-избегаема тогда и только тогда, когда существует β-свободное k-ичное бесконечное вправо слово (ω-слово).

$\Leftarrow$: Очевидно – у бесконечного слова бесконечно много конечных подслов.
$\Rightarrow$: Подслова (в частности, префиксы) β-свободных слов являются β-свободными. Построим дерево всех β-свободных k-ичных слов (каждое слово – узел; узел родителя получается из узла ребенка стиранием последней буквы). Поскольку дерево бесконечно, а у каждого узла не более k детей, то в нем есть бесконечная цепь. Вдоль нее читается β-свободное ω-слово, поскольку все его префиксы β-свободны. $\square$

Аксель Туэ (1863-1922) — норвежец, один из величайших математиков рубежа веков. «Дедушка» комбинаторики слов. Публиковался в норвежских журналах. Мир с удивлением узнал о его результатах примерно в 1950-х годах.

Исследование бесповторных слов: работы Туэ

Аксель Туэ (1863-1922) — норвежец, один из величайших математиков рубежа веков. «Дедушка» комбинаторики слов. Публиковался в норвежских журналах. Мир с удивлением узнал о его результатах примерно в 1950-х годах.

Теорема Туэ о бесквадратных словах (1906)

Экспонента 2 является 3-избегаемой.

Теорема Туэ о сильно бескубных словах (1912)

Экспонента 2^+ является 2-избегаемой.

Морфизм Туэ

Определение

Морфизм есть функция $f : \Sigma^* \to \Sigma^*$ такая, что
$f(\lambda) = \lambda$ и $f(w) = f(w[1]) \cdots f(w[|w|])$ для любого непустого слова w.

★ Любой морфизм определяется своими значениями на буквах (словах длины 1).

Морфизм Туэ

Определение

Морфизм есть функция $f : \Sigma^* \to \Sigma^*$ такая, что
$f(\lambda) = \lambda$ и $f(w) = f(w[1]) \cdots f(w[|w|])$ для любого непустого слова w.

★ Любой морфизм определяется своими значениями на буквах (словах длины 1).

Морфизм Туэ $\theta : \{0,1\}^* \to \{0,1\}^*$ задается значениями

$$\theta(0) = 01, \quad \theta(1) = 10.$$

Теорема Туэ о сильно бескубных словах вытекает из следующей леммы.

Лемма о морфизме Туэ

Если w — 2^+-свободное слово, то $\theta(w)$ — также 2^+-свободное слово.

Морфизм с таким свойством называется 2^+-*свободным*.

Определение

Морфизм есть функция $f : \Sigma^* \to \Sigma^*$ такая, что
$f(\lambda) = \lambda$ и $f(w) = f(w[1]) \cdots f(w[|w|])$ для любого непустого слова w.

★ Любой морфизм определяется своими значениями на буквах (словах длины 1).

Морфизм Туэ $\theta : \{0,1\}^* \to \{0,1\}^*$ задается значениями

$$\theta(0) = 01, \quad \theta(1) = 10.$$

Теорема Туэ о сильно бескубных словах вытекает из следующей леммы.

Лемма о морфизме Туэ

Если w — 2^+-свободное слово, то $\theta(w)$ — также 2^+-свободное слово.

Морфизм с таким свойством называется 2^+-*свободным*. На самом деле, морфизм Туэ еще круче: он является β-свободным для любого $\beta \geqslant 2^+$.

Доказательство леммы о морфизме Туэ

Пусть слово $\theta(w)$ не является 2^+-свободным, т.е. имеет подслово длины $> 2k$ с периодом k для некоторого $k \geqslant 1$. Б.о.о. это подслово начинается с 0. Тогда оно начинается (или совпадает) с $0x0x0$ для некоторого слова x.

Само слово $\theta(w)$ есть произведение $|w|$ штук блоков 01 и 10. Значит, $x \neq \lambda$, а слово $0x0x0$, имеющее нечетную длину, выровнено слева или справа по границе блоков. Пусть оно выровнено слева (второй случай разбирается точно так же). Тогда $0x$ начинается с блока 01; положим $x = 1x'$. За последним 0 следует 1 из того же блока, т.е. в $\theta(w)$ есть подслово $01x'01x'01$:

$$\theta(w) = \underline{\quad |01|\ x'\ \ 01\ \ x'\ |01|\quad}$$

Пусть слово $\theta(w)$ не является 2^+-свободным, т.е. имеет подслово длины $> 2k$ с периодом k для некоторого $k \geqslant 1$. Б.о.о. это подслово начинается с 0. Тогда оно начинается (или совпадает) с $0x0x0$ для некоторого слова x.

Само слово $\theta(w)$ есть произведение $|w|$ штук блоков 01 и 10. Значит, $x \neq \lambda$, а слово $0x0x0$, имеющее нечетную длину, выровнено слева или справа по границе блоков. Пусть оно выровнено слева (второй случай разбирается точно так же). Тогда $0x$ начинается с блока 01; положим $x = 1x'$. За последним 0 следует 1 из того же блока, т.е. в $\theta(w)$ есть подслово $01x'01x'01$:

$$\theta(w) = \underline{\quad |01| \; x' \;\; 01 \;\; x' \; |01| \;}$$

- Если $|x'|$ — нечетная, то слова $x'0$ и $1x'$ являются произведениями блоков. Значит, x' начинается с 0, заканчивается на 1 и состоит из чередующихся букв (так как между любыми двумя позициями в x' проходит граница блока). Но длина слова с таким набором свойств должна быть четной, противоречие.
- Значит, $|x'|$ — четная. Тогда x' — произведение блоков, а значит, w содержит прообраз $0\theta^{-1}(x')0\theta^{-1}(x')0$ слова $01x'01x'01$ и не является 2^+-свободным. $\square$

Замечание

Если буква a и морфизм f таковы, что $f(a)$ начинается с a, то

- $f^2(a) = f(f(a))$ начинается с $f(a)$, ..., $f^{n+1}(a)$ начинается с $f^n(a)$, ...
- если $|f^n(a)| \to \infty$, то у последовательности $\{f^n(a)\}_0^\infty$ есть «предел» – ω-слово $\mathbf{w}$, для которого все $f^n(a)$ – префиксы

Этот предел называется словом, порожденным морфизмом f, и является неподвижной точкой морфизма f, т.е. $f(\mathbf{w}) = \mathbf{w}$.

Слово Туэ-Морса

Замечание

Если буква a и морфизм f таковы, что $f(a)$ начинается с a, то

- $f^2(a) = f(f(a))$ начинается с $f(a)$, ..., $f^{n+1}(a)$ начинается с $f^n(a)$, ...
- если $|f^n(a)| \to \infty$, то у последовательности $\{f^n(a)\}_0^\infty$ есть «предел» – ω-слово $\mathbf{w}$, для которого все $f^n(a)$ – префиксы

Этот предел называется словом, порожденным морфизмом f, и является неподвижной точкой морфизма f, т.е. $f(\mathbf{w}) = \mathbf{w}$.

Слово, порожденное морфизмом Туэ, называется словом Туэ-Морса, выглядит как

$$\mathbf{t} = \theta^\infty(0) = 0110\,1001\,1001\,0110\,1001\,0110\,0110\,1001\,1001\ \ldots$$

Замечание

Если буква a и морфизм f таковы, что $f(a)$ начинается с a, то

- $f^2(a) = f(f(a))$ начинается с $f(a)$, ..., $f^{n+1}(a)$ начинается с $f^n(a)$, ...
- если $|f^n(a)| \to \infty$, то у последовательности $\{f^n(a)\}_0^\infty$ есть «предел» – ω-слово $\mathbf{w}$, для которого все $f^n(a)$ – префиксы

Этот предел называется словом, порожденным морфизмом f, и является неподвижной точкой морфизма f, т.е. $f(\mathbf{w}) = \mathbf{w}$.

Слово, порожденное морфизмом Туэ, называется словом Туэ-Морса, выглядит как

$$\mathbf{t} = \theta^\infty(0) = 0110\,1001\,1001\,0110\,1001\,0110\,0110\,1001\,1001\ \ldots$$

является 2^+-свободным, переоткрыто >10 раз, а его область применения включает

решение проблемы Бернсайда в теории групп	построение геодезических линий
моделирование одномерных квазикристаллов	построение магических квадратов
исследование унимодальных функций	изменение правил шахмат
написание музыкальных произведений	...

Пара свойств слова Туэ-Морса

$$\mathbf{t} = \theta^{\infty}(0) = 0110\,1001\,1001\,0110\,1001\,0110\,0110\,1001\,1001\ \ldots$$

Если позиции в слове Туэ-Морса нумеровать, начиная с нуля, то

- i-я буква равна числу единиц в двоичной записи числа i, взятому по модулю 2

Пара свойств слова Туэ-Морса

$$\mathbf{t} = \theta^{\infty}(0) = 0110\,1001\,1001\,0110\,1001\,0110\,0110\,1001\,1001\,\ldots$$

Если позиции в слове Туэ-Морса нумеровать, начиная с нуля, то

- i-я буква равна числу единиц в двоичной записи числа i, взятому по модулю 2
- разбиение отрезка $[0..2^n-1]$ натурального ряда, задаваемое префиксом слова Туэ-Морса, обладает следующими свойствами ($n \geqslant 2$):
  - суммы элементов в классах равны
  - суммы квадратов элементов в классах равны
  - ...
  - суммы $(n-1)$-х степеней элементов в классах равны:

$$0+3+5+6 = 1+2+4+7 = 14$$
$$0^2+3^2+5^2+6^2 = 1^2+2^2+4^2+7^2 = 70$$

$$\mathbf{t} = \theta^{\infty}(0) = 0110\,1001\,1001\,0110\,1001\,0110\,0110\,1001\,1001\ \ldots$$

Если позиции в слове Туэ-Морса нумеровать, начиная с нуля, то

- i-я буква равна числу единиц в двоичной записи числа i, взятому по модулю 2
- разбиение отрезка $[0..2^n-1]$ натурального ряда, задаваемое префиксом слова Туэ-Морса, обладает следующими свойствами ($n \geqslant 2$):
  - суммы элементов в классах равны
  - суммы квадратов элементов в классах равны
  - ...
  - суммы $(n-1)$-х степеней элементов в классах равны:

$$0+3+5+6 = 1+2+4+7 = 14$$

$$0^2+3^2+5^2+6^2 = 1^2+2^2+4^2+7^2 = 70$$

Удивительно, но если позиции нумеровать, начиная с k, то второе свойство все равно выполняется! (Естественно, надо рассматривать отрезки вида $[k..2^n-k-1]$.) Например,

$$2+5+7+8 = 3+4+6+9 = 22$$

$$2^2+5^2+7^2+8^2 = 3^2+4^2+6^2+9^2 = 142$$

★ Теорема Туэ о сильно бескубных словах является оптимальным результатом в том смысле, что меньшую степень над двухбуквенным алфавитом избежать невозможно: любое слово $w \in \{0,1\}^*$ длины > 3 содержит квадрат

Определение

Границей повторяемости в k-буквенном алфавите называется действительное число $\mathrm{RT}(k) = \inf\{\beta \in \mathbb{R} \mid \beta\ k\text{-избегаема}\}$.

Граница повторяемости

★ Теорема Туэ о сильно бескубных словах является оптимальным результатом в том смысле, что меньшую степень над двухбуквенным алфавитом избежать невозможно: любое слово $w \in \{0,1\}^*$ длины > 3 содержит квадрат

Определение

Границей повторяемости в k-буквенном алфавите называется действительное число $\mathrm{RT}(k) = \inf\{\beta \in \mathbb{R} \mid \beta\ k\text{-избегаема}\}$.

★ Таким образом, $\mathrm{RT}(2) = 2$, причем сама экспонента 2 является 2-неизбежной

Остальные значения функции $\mathrm{RT}(k)$ были предсказаны в 1972 году Ф. Дежан:

Гипотеза Дежан

$$\mathrm{RT}(3) = 7/4, \mathrm{RT}(4) = 7/5, \mathrm{RT}(k) = k/(k-1) \text{ при } k \geqslant 5,$$

причем экспонента $\mathrm{RT}(k)$ является k-неизбежной.

★ Теорема Туэ о сильно бескубных словах является оптимальным результатом в том смысле, что меньшую степень над двухбуквенным алфавитом избежать невозможно: любое слово $w \in \{0,1\}^*$ длины > 3 содержит квадрат

Определение

Границей повторяемости в k-буквенном алфавите называется действительное число $\mathrm{RT}(k) = \inf\{\beta \in \mathbb{R} \mid \beta\ k\text{-избегаема}\}$.

★ Таким образом, $\mathrm{RT}(2) = 2$, причем сама экспонента 2 является 2-неизбежной

Остальные значения функции $\mathrm{RT}(k)$ были предсказаны в 1972 году Ф. Дежан:

Гипотеза Дежан

$$\mathrm{RT}(3) = 7/4, \mathrm{RT}(4) = 7/5, \mathrm{RT}(k) = k/(k-1) \text{ при } k \geqslant 5,$$

причем экспонента $\mathrm{RT}(k)$ является k-неизбежной.

★ теорема Туэ о бесквадратных словах – не оптимальный результат

Мы докажем оптимальный результат (принадлежащий Дежан).

- При наличии терпения или компьютера несложно показать, что любое слово $w \in \{a, b, c\}^*$ длины $\geqslant 39$ содержит подслово с экспонентой $\geqslant 7/4$
- Для доказательства 3-избегаемости экспоненты $(7/4)^+$ Дежан построила $(7/4)^+$-свободный морфизм с длиной блока 19

Воспользуемся более изящной конструкцией, предложенной С.Е. Аршоном (1937)

- При наличии терпения или компьютера несложно показать, что любое слово $w \in \{a, b, c\}^*$ длины $\geqslant$39 содержит подслово с экспонентой $\geqslant$7/4
- Для доказательства 3-избегаемости экспоненты $(7/4)^+$ Дежан построила $(7/4)^+$-свободный морфизм с длиной блока 19

Воспользуемся более изящной конструкцией, предложенной С.Е. Аршоном (1937)

Рассмотрим две функции α_o (нечётная) и α_e (чётная) из $\{a, b, c\}$ в $\{a, b, c\}^*$:

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

Функцию Аршона $\alpha : \{a, b, c\}^* \to \{a, b, c\}^*$ определим так:
$\alpha(w) = \alpha_o(w[1])\alpha_e(w[2])\alpha_o(w[3]) \cdots$, т.е. для букв в нечетных [четных] позициях берется нечетный [соответственно, четный] образ. При этом сохраняется свойство морфизма: из того, что $\alpha(a)$ начинается с a, следует, что $\alpha^{n+1}(a)$ начинается с $\alpha^n(a)$ для любого n. Значит, функция Аршона определяет ω-слово Аршона

$$A = \alpha^\infty(a) = abc\,acb\,cab\,cba\,cab\,acb\,cab\,cba\,bca \cdots .$$

- При наличии терпения или компьютера несложно показать, что любое слово $w \in \{a, b, c\}^*$ длины $\geqslant$39 содержит подслово с экспонентой $\geqslant$7/4
- Для доказательства 3-избегаемости экспоненты $(7/4)^+$ Дежан построила $(7/4)^+$-свободный морфизм с длиной блока 19

Воспользуемся более изящной конструкцией, предложенной С.Е. Аршоном (1937)

Рассмотрим две функции α_o (нечётная) и α_e (чётная) из $\{a, b, c\}$ в $\{a, b, c\}^*$:

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

Функцию Аршона $\alpha : \{a, b, c\}^* \to \{a, b, c\}^*$ определим так: $\alpha(w) = \alpha_o(w[1])\alpha_e(w[2])\alpha_o(w[3]) \cdots$, т.е. для букв в нечетных [четных] позициях берется нечетный [соответственно, четный] образ. При этом сохраняется свойство морфизма: из того, что $\alpha(a)$ начинается с a, следует, что $\alpha^{n+1}(a)$ начинается с $\alpha^n(a)$ для любого n. Значит, функция Аршона определяет ω-слово Аршона

$$A = \alpha^\infty(a) = abc\,acb\,cab\,cba\,cab\,acb\,cab\,cba\,bca \cdots .$$

Аршон доказал, что слово A бесквадратно, но верен и более сильный результат:

- Слово Аршона является $(7/4)^+$-свободным (Клепинин, Суханов, 1999)

Доказательство существования n-значных бесконечных асимметричных последовательностей

С. Е. Аршон (Москва)

1. Под n-значной последовательностью я разумею последовательность знаков, среди которых только n различных. В этом смысле последовательность

$$1\,2\,2\,2\,3\,2\,2\,1\,3$$

будет трехзначной, ибо ее образующими являются три знака, а именно: 1, 2, 3.

2. i каких-либо последовательных знаков данной n-значной последовательности образуют ее часть. Так, например, из указанной выше последовательности можно выделить части: 1 2 2; 2 2 1 3; 3 2 и т. д.

3. Если в последовательности существует часть из px знаков ($x = 1, 2, \ldots, \infty$), которая может быть разбита на p частей, одинаковых как по знакам, так и по порядку их следования, то такую часть (из px знаков) мы будем называть p-кратным повторением, а о последовательности говорить, что она содержит повторение.

Так, например,

$$1\,2\,3\,1\,\underbrace{2\,1\,3\,1}\,\underbrace{2\,1\,1\,2}\,3\,1$$

будет последовательностью с двукратным повторением.

4. n-значную последовательность, не содержащую повторений, будем называть асимметричной. Так, например, последовательность

$$0\,1\,2\,0\,2\,1\,2\,0\,1\,2\,1\,0\,2\,0\,1\,0\,2\,1\,2\,0\,1\,2\,1\,0\,1\,2\,0$$

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

Базовые свойства слова Аршона:

- A – произведение блоков длины 3, каждый из которых – перестановка множества $\{a, b, c\}$
- нечетные (вида α_o) и четные (вида α_e) блоки в A чередуются;
- нечетные блоки – перестановки одной ориентации, их двухбуквенные подслова – ab, bc и ca;
- аналогично для четных блоков и подслов ba, cb и ac.

$$A = \alpha^\infty(a) = abc\,acb\,cab\,cba\,cab\,acb\,cab\,cba\,bca\cdots$$

Базовые свойства слова Аршона:

- A – произведение блоков длины 3, каждый из которых – перестановка множества $\{a, b, c\}$
- нечетные (вида α_o) и четные (вида α_e) блоки в A чередуются;
- нечетные блоки – перестановки одной ориентации, их двухбуквенные подслова – ab, bc и ca;
- аналогично для четных блоков и подслов ba, cb и ac.

Докажем $(7/4)^+$-свободность слова Аршона методом минимального контрпримера.

Пусть в A есть «запрещенные» подслова с экспонентой, большей 7/4,

- n — минимальное число такое, что в $\alpha^n(a)$ есть запрещенное подслово
- w — подслово с наибольшей экспонентой в $\alpha^n(a)$
- $m = \mathrm{per}(w)$.

Слово Аршона: маленькие периоды

$$A = \alpha^\infty(a) = abc\,acb\,cab\,cba\,cab\,acb\,cab\,cba\,bca \cdots$$

- $m \neq 1$: если есть aa, то эти a должны принадлежать разным блокам; тогда эти блоки составляют, с точностью до переобозначения букв, подслово $cba\,abc$; прообраз этого подслова равен либо cc, либо aa, противоречие с выбором n
- $m \neq 2$: внутри блока все буквы различны, поэтому в подслове $abab$ граница блоков должна проходить посередине; но соседние блоки не содержат общих двухбуквенных подслов
- $m \neq 3$: слово длины 6 должно содержать двухбуквенные подслова из блоков разной четности, а все двухбуквенные подслова слова вида $abcabc$ – из блоков одной четности
- $m \neq 4$: слова вида $abacabac$ и $abcbabcb$ не содержатся в произведении блоков
- $m \neq 6, 8$: в том же духе
- $m \neq 5, 7, 9, 10$: любое запрещенное подслово с таким периодом содержит более короткое запрещенное подслово (а выше доказано, что более коротких в A нет)

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

- $m \neq 1$: если есть aa, то эти a должны принадлежать разным блокам; тогда эти блоки составляют, с точностью до переобозначения букв, подслово $cba\,abc$; прообраз этого подслова равен либо cc, либо aa, противоречие с выбором n
- $m \neq 2$: внутри блока все буквы различны, поэтому в подслове $abab$ граница блоков должна проходить посередине; но соседние блоки не содержат общих двухбуквенных подслов
- $m \neq 3$: слово длины 6 должно содержать двухбуквенные подслова из блоков разной четности, а все двухбуквенные подслова слова вида $abcabc$ – из блоков одной четности
- $m \neq 4$: слова вида $abacabac$ и $abcbabcb$ не содержатся в произведении блоков
- $m \neq 6, 8$: в том же духе
- $m \neq 5, 7, 9, 10$: любое запрещенное подслово с таким периодом содержит более короткое запрещенное подслово (а выше доказано, что более коротких в A нет)

Для бо́льших значений m потребуется какое-нибудь общее утверждение.

Лемма о синхронизации

Пусть u — подслово длины $\geqslant 9$ в A. Тогда все вхождения u в A расположены одинаково относительно границ блоков.
(Т.е. если некоторое вхождение u начинается во второй позиции нечетного блока, то все вхождения u начинаются во вторых позициях нечетных блоков.)

Лемма о синхронизации

Пусть u — подслово длины $\geqslant 9$ в A. Тогда все вхождения u в A расположены одинаково относительно границ блоков.
(Т.е. если некоторое вхождение u начинается во второй позиции нечетного блока, то все вхождения u начинаются во вторых позициях нечетных блоков.)

Указанным свойством обладают следующие подслова (и полученные из них переименованием букв):

- $a\,bac\,a$ (образ буквы должен быть блоком)
- $a\,bca\,b$ (соседние блоки не имеют общих подслов длины 2)
- $ab\,acb\,cab$ (разбиение $a\,bac\,bca\,b$ означает, ввиду соблюдения четности, что неполный блок справа это bac, а слева – bca; полученное произведение четырех блоков есть $\alpha(bcbc)$, противоречие с выбором n)
- $abc\,acb\,ab$ (аналогично предыдущему)

Короткий перебор показывает, что любое подслово длины 9 в слове A содержит одно из разобранных подслов, что и доказывает лемму. □

Лемма о синхронизации

Пусть u — подслово длины $\geqslant 9$ в A. Тогда все вхождения u в A расположены одинаково относительно границ блоков.
(Т.е. если некоторое вхождение u начинается во второй позиции нечетного блока, то все вхождения u начинаются во вторых позициях нечетных блоков.)

Указанным свойством обладают следующие подслова (и полученные из них переименованием букв):

- $a\,bac\,a$ (образ буквы должен быть блоком)
- $a\,bca\,b$ (соседние блоки не имеют общих подслов длины 2)
- $ab\,acb\,cab$ (разбиение $a\,bac\,bca\,b$ означает, ввиду соблюдения четности, что неполный блок справа это bac, а слева – bca; полученное произведение четырех блоков есть $\alpha(bcbc)$, противоречие с выбором n)
- $abc\,acb\,ab$ (аналогично предыдущему)

Короткий перебор показывает, что любое подслово длины 9 в слове A содержит одно из разобранных подслов, что и доказывает лемму. □

Заметим, что «несинхронизированные» подслова длины 8 в слове Аршона есть, например, $abcbacab$.

Рассмотрим w — подслово максимальной экспоненты в $\alpha^n(a)$. Пусть u — длиннейший префикс w, являющийся суффиксом w. Мы знаем, что $\mathrm{per}(w) \geqslant 11$. С учетом $\exp(w) > 7/4$ имеем $|u| \geqslant 9$. По лемме о синхронизации, эти два вхождения подслова u расположены одинаково относительно блоков (с учетом четности).

Экспоненты подслов слова Аршона: окончание доказательства

Рассмотрим w — подслово максимальной экспоненты в $\alpha^n(a)$. Пусть u — длиннейший префикс w, являющийся суффиксом w. Мы знаем, что $\mathrm{per}(w) \geqslant 11$. С учетом $\exp(w) > 7/4$ имеем $|u| \geqslant 9$. По лемме о синхронизации, эти два вхождения подслова u расположены одинаково относительно блоков (с учетом четности).

- Блок, четность которого известна, однозначно определяется и своей первой буквой, и своей последней буквой
- Если u начинается не с первой позиции блока, то перед обоими вхождениями u стоит одна и та же буква (б.о.о., a); тогда слово aw имеет префикс и суффикс au, т.е. у него тот же период, что у w; получаем $\exp(aw) > \exp(w)$, противоречие с выбором w
- Аналогичное рассуждение проходит, если u заканчивается не в последней позиции блока

Экспоненты подслов слова Аршона: окончание доказательства

Рассмотрим w — подслово максимальной экспоненты в $\alpha^n(a)$. Пусть u — длиннейший префикс w, являющийся суффиксом w. Мы знаем, что $\mathrm{per}(w) \geqslant 11$. С учетом $\exp(w) > 7/4$ имеем $|u| \geqslant 9$. По лемме о синхронизации, эти два вхождения подслова u расположены одинаково относительно блоков (с учетом четности).

- Блок, четность которого известна, однозначно определяется и своей первой буквой, и своей последней буквой
- Если u начинается не с первой позиции блока, то перед обоими вхождениями u стоит одна и та же буква (б.о.о., a); тогда слово aw имеет префикс и суффикс au, т.е. у него тот же период, что у w; получаем $\exp(aw) > \exp(w)$, противоречие с выбором w
- Аналогичное рассуждение проходит, если u заканчивается не в последней позиции блока
- Значит, u – произведение блоков; тогда w – тоже произведение блоков
- Тогда существует слово $\alpha^{-1}(w)$, оно является подсловом в $\alpha^{n-1}(a)$, и начинается и заканчивается на $\alpha^{-1}(u)$
- Имеем $\exp(\alpha^{-1}(w)) \geqslant \exp(w) > 7/4$, противоречие с выбором n

Экспоненты подслов слова Аршона: окончание доказательства

Рассмотрим w — подслово максимальной экспоненты в $\alpha^n(a)$. Пусть u — длиннейший префикс w, являющийся суффиксом w. Мы знаем, что $\mathrm{per}(w) \geqslant 11$. С учетом $\exp(w) > 7/4$ имеем $|u| \geqslant 9$. По лемме о синхронизации, эти два вхождения подслова u расположены одинаково относительно блоков (с учетом четности).

- Блок, четность которого известна, однозначно определяется и своей первой буквой, и своей последней буквой
- Если u начинается не с первой позиции блока, то перед обоими вхождениями u стоит одна и та же буква (б.о.о., a); тогда слово aw имеет префикс и суффикс au, т.е. у него тот же период, что у w; получаем $\exp(aw) > \exp(w)$, противоречие с выбором w
- Аналогичное рассуждение проходит, если u заканчивается не в последней позиции блока
- Значит, u – произведение блоков; тогда w – тоже произведение блоков
- Тогда существует слово $\alpha^{-1}(w)$, оно является подсловом в $\alpha^{n-1}(a)$, и начинается и заканчивается на $\alpha^{-1}(u)$
- Имеем $\exp(\alpha^{-1}(w)) \geqslant \exp(w) > 7/4$, противоречие с выбором n

Итак, A не содержит подслов с экспонентой большей $7/4$, т.е. $\mathrm{RT}(3) = 7/4$. □

При наличии очень большого терпения или компьютера можно показать перебором, что любое слово длины 122 над 4-буквенным алфавитом содержит подслово с экспонентой $\geqslant 7/5$. Для бóльших алфавитов нижняя оценка тривиальна:

Замечание о нижней оценке

$\frac{k}{k-1}$-свободное слово над k-буквенным алфавитом имеет длину не больше $k+1$.

Если $\frac{k}{k-1}$-свободное слово имеет вид axa, где a — буква, то $\frac{|axa|}{|ax|} < \frac{k}{k-1}$, откуда $|x| \geqslant k-1$. Значит, любые k последовательных букв в таком слове различны. Но любое слово длины $k+2$ над k-буквенным алфавитом, удовлетворяющее этому условию, имеет вид $a_1a_2 \cdots a_ka_1a_2$ и экспоненту $\frac{k+2}{k} \geqslant \frac{k}{k-1}$. □

При наличии очень большого терпения или компьютера можно показать перебором, что любое слово длины 122 над 4-буквенным алфавитом содержит подслово с экспонентой $\geqslant 7/5$. Для бóльших алфавитов нижняя оценка тривиальна:

Замечание о нижней оценке

$\frac{k}{k-1}$-свободное слово над k-буквенным алфавитом имеет длину не больше $k+1$.

Если $\frac{k}{k-1}$-свободное слово имеет вид axa, где a — буква, то $\frac{|axa|}{|ax|} < \frac{k}{k-1}$, откуда $|x| \geqslant k-1$. Значит, любые k последовательных букв в таком слове различны. Но любое слово длины $k+2$ над k-буквенным алфавитом, удовлетворяющее этому условию, имеет вид $a_1a_2\cdots a_ka_1a_2$ и экспоненту $\frac{k+2}{k} \geqslant \frac{k}{k-1}$. □

В то же время, с верхними оценками все сложно. Например:

Теорема (Бранденбург, 1983)

Не существует $\frac{k}{k-1}^+$-свободных морфизмов над k-буквенным алфавитом при $k \geqslant 5$.
Не существует $(7/5)^+$-свободных морфизмов над 4-буквенным алфавитом.

Кодировка Пансьё

В 1984 году Ж.-Ж. Пансьё предложил заменить k-ичные слова малой экспоненты бинарными кодами, строимыми по простому правилу, и доказывать гипотезу Дежан для любых алфавитов, работая с кодами, т.е. со словами в двоичном алфавите.

Кодировка Пансьё

В 1984 году Ж.-Ж. Пансьё предложил заменить k-ичные слова малой экспоненты бинарными кодами, строимыми по простому правилу, и доказывать гипотезу Дежан для любых алфавитов, работая с кодами, т.е. со словами в двоичном алфавите.

В каждом k-ичном $\frac{k-1}{k-2}$-свободном слове любые $k-1$ подряд идущих букв различны. Значит, каждое такое слово w можно задать его префиксом $w[1..k-1]$ и двоичным следом $\mathrm{tr}(w)$ длины $|w|-k+1$, построенным по следующему правилу:

$$\mathrm{tr}(w)[i] = \begin{cases} 0 & \text{если } w[i+k-1] = w[i], \\ 1 & \text{если } w[i+k-1] \notin \{w[i], \ldots, w[i+k-2]\}. \end{cases}$$

Пример: $w = \boldsymbol{a\,b\,c\,d}\; a\,e\,c\,b\,d\,e\,a\,b\ldots$

$\mathrm{tr}(w) = 0\,1\,0\,1\,1\,0\,1\,0\ldots$

Кодировка Пансьё

В 1984 году Ж.-Ж. Пансьё предложил заменить k-ичные слова малой экспоненты бинарными кодами, строимыми по простому правилу, и доказывать гипотезу Дежан для любых алфавитов, работая с кодами, т.е. со словами в двоичном алфавите.

В каждом k-ичном $\frac{k-1}{k-2}$-свободном слове любые $k-1$ подряд идущих букв различны. Значит, каждое такое слово w можно задать его префиксом $w[1..k-1]$ и двоичным следом $\mathrm{tr}(w)$ длины $|w|-k+1$, построенным по следующему правилу:

$$\mathrm{tr}(w)[i] = \begin{cases} 0 & \text{если } w[i+k-1] = w[i], \\ 1 & \text{если } w[i+k-1] \notin \{w[i], \ldots, w[i+k-2]\}. \end{cases}$$

Пример:
$$\begin{aligned} w &= \boldsymbol{a\,b\,c\,d}\; a\,e\,c\,b\,d\,e\,a\,b\ldots \\ \mathrm{tr}(w) &= \; 0\,1\,0\,1\,1\,0\,1\,0\ldots \end{aligned}$$

k-ичные слова с равными следами получаются друг из друга переименованием букв, т.е. ничем не отличаются друг от друга с точки зрения экспонент подслов. Значит, доказательство гипотезы Дежан сводится к построению двоичных слов:

- Для любого $k \geqslant 4$ построить двоичное ω-слово, которое является следом k-ичного RT^{+}-свободного ω-слова.

В 1984 году Ж.-Ж. Пансьё предложил заменить k-ичные слова малой экспоненты бинарными кодами, строимыми по простому правилу, и доказывать гипотезу Дежан для любых алфавитов, работая с кодами, т.е. со словами в двоичном алфавите.

В каждом k-ичном $\frac{k-1}{k-2}$-свободном слове любые $k-1$ подряд идущих букв различны. Значит, каждое такое слово w можно задать его префиксом $w[1..k-1]$ и двоичным следом $\mathrm{tr}(w)$ длины $|w|-k+1$, построенным по следующему правилу:

$$\mathrm{tr}(w)[i] = \begin{cases} 0 & \text{если } w[i+k-1] = w[i], \\ 1 & \text{если } w[i+k-1] \notin \{w[i], \ldots, w[i+k-2]\}. \end{cases}$$

Пример: $$w = \boldsymbol{a\,b\,c\,d}\; a\,e\,c\,b\,d\,e\,a\,b \ldots$$
$$\mathrm{tr}(w) = 0\,1\,0\,1\,1\,0\,1\,0 \ldots$$

k-ичные слова с равными следами получаются друг из друга переименованием букв, т.е. ничем не отличаются друг от друга с точки зрения экспонент подслов. Значит, доказательство гипотезы Дежан сводится к построению двоичных слов:

- Для любого $k \geqslant 4$ построить двоичное ω-слово, которое является следом k-ичного RT^+-свободного ω-слова.

Разбив на много случаев, к 2009 году гипотезу Дежан доказали.

Примерно с 2000 года большую популярность в теории графов набрала задача о бесповторной раскраске.

Определение

Правильная k-раскраска обыкновенного графа $G = (V, E)$ есть функция $\alpha : V \to \{1, \ldots, k\}$ такая, что $\alpha(u) \neq \alpha(v)$ для любого ребра $(u, v) \in E$. Минимальное число вершин в правильной k-раскраске G называется хроматическим числом G (обозначается $\chi(G)$).

Правильность раскраски можно переформулировать так:

- ★ цвета вершин вдоль любой простой цепи образуют слово, не имеющее подслов aa, где a – буква

Естественно обобщить:

Бесповторные раскраски графов

Примерно с 2000 года большую популярность в теории графов набрала задача о бесповторной раскраске.

Определение

Правильная k-раскраска обыкновенного графа $G = (V, E)$ есть функция $\alpha : V \to \{1, \ldots, k\}$ такая, что $\alpha(u) \neq \alpha(v)$ для любого ребра $(u, v) \in E$. Минимальное число вершин в правильной k-раскраске G называется хроматическим числом G (обозначается $\chi(G)$).

Правильность раскраски можно переформулировать так:

- ★ цвета вершин вдоль любой простой цепи образуют слово, не имеющее подслов aa, где a – буква

Естественно обобщить:

Определение (Алон, Грищак (?))

Бесповторная k-раскраска обыкновенного графа $G = (V, E)$ есть функция $\alpha : V \to \{1, \ldots, k\}$ такая, что цвета вершин вдоль любой простой цепи образуют бесквадратное слово. Минимальное число вершин в бесповторной k-раскраске G называется числом Туэ графа G (обозначается $\pi(G)$).

Бесповторные раскраски путей и циклов

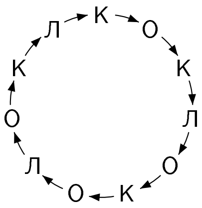
Очевидно, что $\pi(G) \geqslant \chi(G)$. Что можно доказать про $\pi(G)$?

- Пусть P_n – n-вершинная простая цепь; тогда $\pi(G) = 3$ при $n > 3$ (это в точности теорема Туэ о бесквадратных словах – надо раскрасить цепь бесквадратным словом)
- Пусть C_n – n-вершинный простой цикл; тогда $\pi(G) = 4$ при $n = 5, 7, 9, 10, 14, 17$ и $\pi(G) = 3$ в остальных случаях (Карри, 2002 – компьютерное доказательство, Шур, 2010 – бескомпьютерное)

Очевидно, что $\pi(G) \geqslant \chi(G)$. Что можно доказать про $\pi(G)$?

- Пусть P_n – n-вершинная простая цепь; тогда $\pi(G) = 3$ при $n > 3$ (это в точности теорема Туэ о бесквадратных словах – надо раскрасить цепь бесквадратным словом)
- Пусть C_n – n-вершинный простой цикл; тогда $\pi(G) = 4$ при $n = 5, 7, 9, 10, 14, 17$ и $\pi(G) = 3$ в остальных случаях (Карри, 2002 – компьютерное доказательство, Шур, 2010 – бескомпьютерное)

Задача о бесповторной раскраске циклов – это задача о существовании бесквадратных циклических слов над трехбуквенным алфавитом. Циклические слова можно представлять себе как обычные, у которых “склеены” начало и конец.

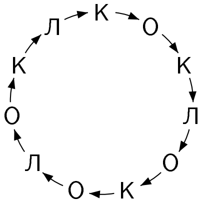


Бесповторные раскраски путей и циклов

Очевидно, что $\pi(G) \geqslant \chi(G)$. Что можно доказать про $\pi(G)$?

- Пусть P_n – n-вершинная простая цепь; тогда $\pi(G) = 3$ при $n > 3$ (это в точности теорема Туэ о бесквадратных словах – надо раскрасить цепь бесквадратным словом)
- Пусть C_n – n-вершинный простой цикл; тогда $\pi(G) = 4$ при $n = 5, 7, 9, 10, 14, 17$ и $\pi(G) = 3$ в остальных случаях (Карри, 2002 – компьютерное доказательство, Шур, 2010 – бескомпьютерное)

Задача о бесповторной раскраске циклов – это задача о существовании бесквадратных циклических слов над трехбуквенным алфавитом. Циклические слова можно представлять себе как обычные, у которых "склеены" начало и конец.



Подсловами циклического слова являются обычные слова, которые можно из него «вырезать», в том числе слова, полученные одним разрезом (равные по длине самому циклическому слову), например

$$\text{КОК}, \text{КЛОК} \text{ или } \text{КОКЛОКОЛОКЛ}, \text{ но не } \text{КОЛОКОЛ}.$$

Понятие бесквадратности переносится на циклические слова без изменений, но строить бесквадратные циклические слова сложнее, чем обычные.

Теорема (Алон, Грищак, Риордан, 2002)

Если G – дерево, то $\pi(G) \leqslant 4$.

Теорема (Алон, Грищак, Риордан, 2002)

Если G – дерево, то $\pi(G) \leqslant 4$.

- Возьмем дерево G, выберем корень v_0 и расположим все вершины по уровням (номер уровня равен расстоянию до корня)
- Пусть всего уровней n и пусть $w = a_0 a_1 \cdots a_{n-1}$ – слово длины n, в котором нет квадратов, а все палиндромы – однобуквенные (построим его позже)
- Раскрасим G так, что все вершины i-го уровня получают цвет a_i

Теорема (Алон, Грищак, Риордан, 2002)

Если G – дерево, то $\pi(G) \leqslant 4$.

- Возьмем дерево G, выберем корень v_0 и расположим все вершины по уровням (номер уровня равен расстоянию до корня)
- Пусть всего уровней n и пусть $w = a_0a_1 \cdots a_{n-1}$ – слово длины n, в котором нет квадратов, а все палиндромы – однобуквенные (построим его позже)
- Раскрасим G так, что все вершины i-го уровня получают цвет a_i
- ♣ Возьмем любую цепь $v_1, \ldots, v_{2r}$ и покажем, что $\alpha(v_1)\cdots\alpha(v_r) \neq \alpha(v_{r+1})\cdots\alpha(v_{2r})$
- ♣ Пусть цепь «перегибается» – у нее есть верхняя точка v_h уровня i, оба соседа которой находятся уровнем ниже (иначе она помечена подсловом w)
- ♣ Б.о.о., $h \leqslant r$; тогда цепь помечена словом $u = a_{h+i-1}\cdots a_{i+1}a_ia_{i+1}\cdots a_{2r-h+i}$
- ♣ Если $h < r$, то в левой половине u есть палиндром $a_{i+1}a_ia_{i+1}$; поскольку правая половина u есть подслово w, там нет этого палиндрома, а значит, u – не квадрат
- ♣ Если $h = r$, то левая половина u начинается с $a_{r+i-1}\cdots a_{i+1}$, а правая – с $a_{i+1}\cdots a_{r+i-1}$; это подслово w, т.е. не палиндром, откуда u – не квадрат

Лемма

Существует ω-слово $\mathbf{w}$ над четырехбуквенным алфавитом, в котором нет квадратов, а все палиндромы – однобуквенные.

Слово без квадратов и палиндромов

Лемма

Существует ω-слово $\mathbf{w}$ над четырехбуквенным алфавитом, в котором нет квадратов, а все палиндромы – однобуквенные.

Пусть $\mathbf{v}$ – бесквадратное ω-слово над алфавитом $\{a, b, c\}$. Тогда можно взять

$$\mathbf{w} = \mathbf{v}[1]\mathbf{v}[2]d\mathbf{v}[3]\mathbf{v}[4]d\cdots$$

- В $\mathbf{w}$ нет квадратов, так как в двух соседних подсловах равной длины
  - не совпадают вхождения d, если длина не кратна 3
  - можно вычеркнуть вхождения d, если длина кратна 3, но то, что останется – не квадрат, так как солержится в $\mathbf{v}$
- В $\mathbf{w}$ нет палиндромов длины > 1, так как каждый такой палиндром содержит палиндром вида aa или aba, где a и b – буквы, а в $\mathbf{w}$ одинаковые буквы находятся на расстоянии не меньше 3

Слово без квадратов и палиндромов

Лемма

Существует ω-слово $\mathbf{w}$ над четырехбуквенным алфавитом, в котором нет квадратов, а все палиндромы – однобуквенные.

Пусть $\mathbf{v}$ – бесквадратное ω-слово над алфавитом $\{a, b, c\}$. Тогда можно взять

$$\mathbf{w} = \mathbf{v}[1]\mathbf{v}[2]d\mathbf{v}[3]\mathbf{v}[4]d \cdots$$

- В $\mathbf{w}$ нет квадратов, так как в двух соседних подсловах равной длины
  - не совпадают вхождения d, если длина не кратна 3
  - можно вычеркнуть вхождения d, если длина кратна 3, но то, что останется – не квадрат, так как солержится в $\mathbf{v}$
- В $\mathbf{w}$ нет палиндромов длины > 1, так как каждый такой палиндром содержит палиндром вида aa или aba, где a и b – буквы, а в $\mathbf{w}$ одинаковые буквы находятся на расстоянии не меньше 3

Взяв $w = \mathbf{w}[1..n]$, построим бесповторную раскраску дерева G в 4 цвета, как на предыдущем слайде.
Трех цветов для бесповторной раскраски любого дерева недостаточно (упражнение: построить контрпример). □