

Слова и избегаемость

Анна (Эдуардовна) Фрид

ИМ им. С. Л. Соболева СО РАН, Новосибирск
anna.e.frid@gmail.com

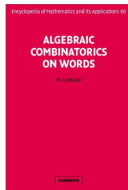
Лекция 1, 15.10.2011

Начало этой лекции в основном излагается по книге

M. Lothaire, *Algebraic Combinatorics on Words*. Cambridge Univ. Press, 2002.

Глава 3, Pattern avoidance (J. Cassaigne)

Книга полностью выложена в сеть



Материал из других источников в конце лекции будет отмечен отдельно.

$\{0, 1, \dots, k\}$, $\{a, b, \dots, z\}$, $\{X, Y, Z\}$ — алфавиты; обозначаются прописными греческими буквами;

abracadabra, *XY YX*, 01122330 — (конечные) слова; множество всех конечных слов над алфавитом Σ обозначается через Σ^* ; конечных непустых слов — через Σ^+ ;

314159265358979323846264338327950288...

abbabaabbaababbaaaaabbbbaaaaaabbbbb... —

бесконечные слова; их множество обозначается через Σ^ω .

- Формальные языки;

- Формальные языки;
- Алгебра
(Σ^* — свободный моноид относительно катенации, единица - пустое слово);

- Формальные языки;
- Алгебра
(Σ^* — свободный моноид относительно катенации, единица - пустое слово);
- Дискретная динамика
(бесконечное слово как код траектории точки. Например, бильярды);

- Формальные языки;
- Алгебра
(Σ^* — свободный моноид относительно катенации, единица - пустое слово);
- Дискретная динамика
(бесконечное слово как код траектории точки. Например, бильярды);
- Теория чисел;

- Формальные языки;
- Алгебра
(Σ^* — свободный моноид относительно катенации, единица - пустое слово);
- Дискретная динамика
(бесконечное слово как код траектории точки. Например, бильярды);
- Теория чисел;
- Дискретная геометрия.

Подсловом конечного или бесконечного слова w называется конечное слово u , такое что $w = ruw'$.

Длина слова u обозначается через $|u|$.

011010011001...

Подсловом конечного или бесконечного слова w называется конечное слово u , такое что $w = ruw'$.

Длина слова u обозначается через $|u|$.

011010011001...

Квадратом называется слово вида uu , $u \in \Sigma^+$.

пурпур, варвар, bonbon — квадраты

Избегать чего-нибудь = не содержать подслов такого вида.

Вопрос

Существует ли бесконечное слово над конечным алфавитом, избегающее квадратов?

Вопрос

Существует ли бесконечное слово над конечным алфавитом, избегающее квадратов?

Над $\{a, b\}$: НЕТ

aba — самое длинное.

Вопрос

Существует ли бесконечное слово над конечным алфавитом, избегающее квадратов?

Над $\{a, b\}$: НЕТ

aba — самое длинное.

Над $\{a, b, c\}$:

abacaba — тупик;

Вопрос

Существует ли бесконечное слово над конечным алфавитом, избегающее квадратов?

Над $\{a, b\}$: НЕТ

aba — самое длинное.

Над $\{a, b, c\}$:

$abacaba$ — тупик;

$abcacbabcbacabscab \dots$ — вроде бы нет проблем.

Theorem (Thue, 1906)

Существует бесконечное слово над трехбуквенным алфавитом, избегающее квадратов.

Theorem (Thue, 1906)

Существует бесконечное слово над трехбуквенным алфавитом, избегающее квадратов.

Конструкция Туэ была относительно громоздкой; мы докажем теорему другим способом, но с помощью другого изобретения Туэ же.

Пусть Γ, Δ — конечные алфавиты. *Морфизмом* называется отображение $\varphi : \Gamma^* \rightarrow \Delta^*$, для всех слов $x, y \in \Gamma^*$ удовлетворяющее равенству $\varphi(xy) = \varphi(x)\varphi(y)$.

Ясно, что всякий морфизм однозначно определяется образами символов $\varphi(g)$, $g \in \Gamma$.

Example

Морфизм Туэ-Морса $\varphi : \{0, 1\}^* \rightarrow \{0, 1\}^*$ $\varphi(0) = 01, \varphi(1) = 10$
Тогда, например, $\varphi(1101) = 10\ 10\ 01\ 10$.

Неподвижные точки морфизмов

$\varphi(a)$ начинается с a , длины слов $\varphi^k(a)$ неограниченно возрастают $\implies$ у них существует предел, называемый *неподвижной точкой морфизма*.

$$a \rightarrow \varphi(a) \rightarrow \varphi^2(a) \rightarrow \varphi^3(a) \rightarrow \dots \rightarrow \varphi^\infty(a)$$

Example

$$\varphi(0) = 01, \varphi(1) = 10$$

$$0 \rightarrow 01 \rightarrow 01\ 10 \rightarrow 0110\ 1001 \rightarrow 01101001\ 10010110 \rightarrow \dots$$

Предел: слово Туэ-Морса

$$w_{TM} = \varphi^\infty(a) = 0110\ 1001\ 1001\ 0110\ 1001\ 0110\ 0110\ 1001 \dots$$

Почему неподвижная точка? Потому что $\varphi(\varphi^\infty(a)) = \varphi^\infty(a)$.

Перекрытием (overlap) называется слово вида $axaxa$, где a — буква, а x — слово (возможно, пустое).

Theorem (Thue, 1912)

Слово Туэ-Морса избегает перекрытий.

The Thue-Morse word is overlap-free.

$$w_{TM} = \varphi^\infty(a) = 0110\ 1001\ 1001\ 0110\ 1001\ 0110\ 0110\ 1001 \dots$$

Theorem (Thue, 1912)

Слово Туэ-Морса избегает перекрытий.

ДОКАЗАТЕЛЬСТВО.

- 000, 111 не встречаются в w_{TM} .
- Пусть $avava$ — кратчайшее перекрытие в w_{TM} ; тогда $|avava| \geq 5$, то есть $|ava| \geq 3$, то есть в $|ava|$ есть подслово 00 или 11, то есть длина $|av|$ четна.
- Возьмем символы слова ava через один, получим более короткое перекрытие, которое встречается в w_{TM} . Противоречие. $\square$

Конструкция бесквадратного слова

Рассмотрим морфизм $\mu : \{a, b, c\}^* \rightarrow \{a, b, c\}^*$:

$$\mu : \begin{cases} a \mapsto abc, \\ b \mapsto ac, \\ c \mapsto b. \end{cases}$$

Пусть $u = abc \ ac \ b \ abcb \ ac \ abcsacbac \ abcb \cdots$ — неподвижная точка морфизма μ . Тогда u не содержит квадратов.

Конструкция бесквадратного слова

Рассмотрим морфизм $\mu : \{a, b, c\}^* \rightarrow \{a, b, c\}^*$:

$$\mu : \begin{cases} a \mapsto abc, \\ b \mapsto ac, \\ c \mapsto b. \end{cases}$$

Пусть $u = abc \ ac \ b \ abcb \ ac \ abcasb \ ac \ abcb \cdots$ — неподвижная точка морфизма μ . Тогда u не содержит квадратов.

ДОКАЗАТЕЛЬСТВО. $u = \mu(u)$ по определению.

Пусть $\pi(a) = 011$, $\pi(b) = 01$, $\pi(c) = 0$. Тогда

- $\pi \circ \mu = \varphi \circ \pi$.
- $\varphi(\pi(u)) = \pi(\mu(u)) = \pi(u)$.
- $\pi(u) = w_{TM}$.

Если u содержит квадрат vv , то w_{TM} содержит перекрытие $\pi(v)\pi(v)a$.
Противоречие. □

Что вокруг?

- Избегание паттернов (pattern avoidance);

Что вокруг?

- Избегание паттернов (pattern avoidance);
- Избегание дробных степеней: гипотеза Дежан недавно доказана!

Что вокруг?

- Избегание паттернов (pattern avoidance);
- Избегание дробных степеней: гипотеза Дежан недавно доказана!
- Количество слов длины n , избегающих паттерна или степени;

- Избегание паттернов (pattern avoidance);
- Избегание дробных степеней: гипотеза Дежан недавно доказана!
- Количество слов длины n , избегающих паттерна или степени;
- Избегание степеней в сильном смысле: абелевы степени, весовые степени: избегание весовых кубов! Нерешенный вопрос о квадратах.

Пусть слова рассматриваются над алфавитом Σ . *Паттерном* называется слово над алфавитом Δ из переменных.

Квадрат можно интерпретировать как паттерн XX .

В общем случае *вхождением паттерна P* над алфавитом $\Delta = \{X, Y, \dots\}$ называется слово вида $h(P)$, где $h : \Delta \rightarrow \Sigma^+$ — морфизм.

Example

Слово $aba\ bb\ bb\ aba$ — вхождение паттерна $XY YX$.

Паттерн X неизбежаем;

Неизбегаемые паттерны

Паттерн X неизбежаем;

Паттерн XYX неизбежаем;

Паттерн X неизбегаем;

Паттерн XYX неизбегаем;

Если паттерн P неизбегаем, и X — новая переменная, не встречающаяся в P , то паттерн PXP неизбегаем.

Определим паттерны Z_n итеративно:

$$\begin{aligned}Z_0 &= X_1, \\Z_n &= Z_{n-1}X_nZ_{n-1}.\end{aligned}$$

Theorem

Все паттерны Z_n неизбежны.

Зимин (1979,1982), Bean, Ehrenfeucht, McNulty (1979).

Theorem

Всякий паттерн избегаем тогда и только тогда, когда несводим.

Сводимость: рассмотрим двудольный *граф смежности* паттерна $P \in \Delta^*$.

$(A_l, B_r) \in E \iff AB$ подслово паттерна P .

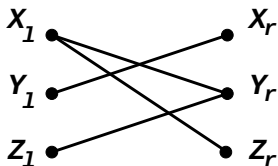


Рис.: Граф смежности паттерна $XYXZYX$

Характеризация неизбежных паттернов

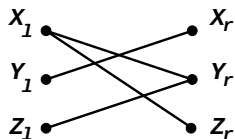


Рис.: Граф смежности паттерна XYXZYX

Свободные множества: $\forall A, B \in S$ вершины A_l и B_r принадлежат разным компонентам связности.

Здесь: $\{X\}, \{Y\}$.

Характеризация неизбежных паттернов

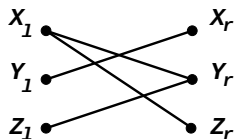


Рис.: Граф смежности паттерна XYXZYX

Свободные множества: $\forall A, B \in S$ вершины A_l и B_r принадлежат разным компонентам связности.

Здесь: $\{X\}, \{Y\}$.

$$XYXZYX \xrightarrow{\{X\}} YZY \xrightarrow{\{Y\}} Z \xrightarrow{\{Z\}} \lambda$$

Паттерн сводим — следовательно, неизбежен.

Характеризация неизбежных паттернов

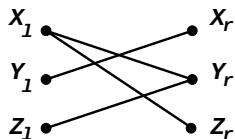


Рис.: Граф смежности паттерна XYXZYX

Свободные множества: $\forall A, B \in S$ вершины A_l и B_r принадлежат разным компонентам связности.

Здесь: $\{X\}, \{Y\}$.

$$XYXZYX \xrightarrow{\{X\}} YZY \xrightarrow{\{Y\}} Z \xrightarrow{\{Z\}} \lambda$$

Паттерн сводим — следовательно, неизбежен.

ВНИМАНИЕ! $XYXZYX \xrightarrow{\{Y\}} XXZX$

$XXZX$ избегаем — следовательно, этот путь тупиковый (но есть другой).

Иногда надо удалять сразу по два элемента (то есть использовать неодноразовые свободные множества).

Example

$$XYX \ Z \ X'YX \ W \ XYX' \ Z \ X'YX'$$

Начинать можно только со свободного множества $\{X, X'\}$

Общий вопрос

Предположим, паттерн избегаем. Какова минимальная мощность алфавита слова, избегающего его?

J. Cassaigne разобрал все бинарные паттерны;
P. Ochem продолжил.

Частный вопрос

Существует ли паттерн, избегаемый над алфавитом мощности 5, но не мощности 4?

Да, существует: R. J. Clark, 2005.

Все известные примеры бесконечных слов, избегающих паттернов, построены либо как неподвижные точки морфизмов, либо как образы неподвижной точки морфизма под действием другого морфизма:

$$w = \varphi^\infty(a) = \varphi(w),$$

$$u = \psi(w)$$

Такие слова называются *HD0L-словами*.

Нерешенный вопрос

Предположим, паттерн избегаем над данным алфавитом. Всегда ли существует HD0L-слово над этим алфавитом, избегающее его?

Слова-примеры часто строятся с помощью большого компьютерного перебора.

Слово $y = \underbrace{x \cdots x}_k x'$, где x' — префикс слова x , называется *дробной степенью* $\frac{|y|}{|x|}$.

Слово $y = \underbrace{x \cdots x}_k x'$, где x' — префикс слова x , называется *дробной степенью* $\frac{|y|}{|x|}$.

Example

entente или *антанта* — степени $7/3$.

Слово $y = \underbrace{x \cdots x}_k x'$, где x' — префикс слова x , называется *дробной степенью* $\frac{|y|}{|x|}$.

Example

entente или *антанта* — степени $7/3$.

Критическая экспонента $RT(n)$ — максимальное число, такое что степени меньше $RT(n)$ неизбежны над алфавитом мощности n , а степени больше $RT(n)$ избегаемы.

Слово $y = \underbrace{x \cdots x}_k x'$, где x' — префикс слова x , называется *дробной степенью* $\frac{|y|}{|x|}$.

Example

entente или *антанта* — степени $7/3$.

Критическая экспонента $RT(n)$ — максимальное число, такое что степени меньше $RT(n)$ неизбежны над алфавитом мощности n , а степени больше $RT(n)$ избегаемы.

$RT(2) = 2$ (Туз-Морс); $RT(3) = 7/4$ (Dejean, 1972).

Бывшая гипотеза Дежан, 1972

$$RT(n) = \frac{n}{n-1} \text{ для всех } n \geq 5.$$

- J. Moulin Ollagnier, 1992: $5 \leq n \leq 11$
- A. Carpi, 2007: $n \geq 35$
- J. Currie, N. Rampersad, 2009; M. Rao, 2009: доделали.

Общий вопрос

Пусть $p(n)$ — количество слов над заданным алфавитом, избегающих "чего-нибудь". Как растет эта функция?

Почти всегда ответ — всего лишь оценки на асимптотику, и требует больших компьютерных вычислений.

Theorem (Kobayashi, 1988; Cassaigne, 1993)

Пусть $p(n)$ — количество бинарных слов, избегающих перекрытий.
Тогда "иногда" $C_1 n^{1,155} \leq p(n) \leq C_2 n^{1,275}$, а
"иногда" $C_3 n^{1,332} \leq p(n) \leq C_4 n^{1,587}$.

Нетривиальные вычисления с матрицами.

Количество бинарных слов, избегающих больших степеней

Theorem (Karhumäki, Shallit, 2004)

Пусть $p(\alpha, n)$ — количество бинарных слов, избегающих степени α . Тогда $p(\alpha, n)$ растёт полиномиально при $\alpha \leq 7/3$ и экспоненциально при $\alpha > 7/3$.

- Быстро развивающаяся область:
кроме вышеупомянутых, А. Шур, Р. Колпаков, Р. Оchem;

- Быстро развивающаяся область:
кроме вышеупомянутых, А. Шур, Р. Колпаков, Р. Ochem;
- Верхние оценки в целом легче получать, чем нижние (понятно, собственные числа каких матриц считать);

- Быстро развивающаяся область:
кроме вышеупомянутых, А. Шур, Р. Колпаков, Р. Ochom;
- Верхние оценки в целом легче получать, чем нижние (понятно, собственные числа каких матриц считать);
- Большие вычислительные мощности и нетривиальные алгоритмы.

Пусть $|u|_a$ — количество вхождения символа a в слово u .
Вектор состава $\psi(u) = (|u|_a, |u|_b, \dots)$.

Пусть $|u|_a$ — количество вхождения символа a в слово u .

Вектор состава $\psi(u) = (|u|_a, |u|_b, \dots)$.

Слова X и X' называются *абелево эквивалентными*, если $\psi(X) = \psi(X')$.

Слово XX' , где X и X' абелево эквивалентны, называется *абелевым квадратом*.

Пусть $|u|_a$ — количество вхождения символа a в слово u .

Вектор состава $\psi(u) = (|u|_a, |u|_b, \dots)$.

Слова X и X' называются *абелево эквивалентными*, если $\psi(X) = \psi(X')$.

Слово XX' , где X и X' абелево эквивалентны, называется *абелевым квадратом*.

Эрдёш, 1961

Существует ли бесконечное слово над конечным алфавитом, избегающее абелевых квадратов?

Эрдёш, 1961

Существует ли бесконечное слово над конечным алфавитом, избегающее абелевых квадратов?

Эрдёш, 1961

Существует ли бесконечное слово над конечным алфавитом, избегающее абелевых квадратов?

Евдокимов, 1962

Да, существует, над алфавитом мощности 25.

Эрдёш, 1961

Существует ли бесконечное слово над конечным алфавитом, избегающее абелевых квадратов?

Евдокимов, 1962

Да, существует, над алфавитом мощности 25.

Трех символов точно мало; пяти достаточно (Pleasants, 1965). Что с четырьмя?

V. Keränen, 1992: пусть w — неподвижная точка морфизма φ , где

$$\begin{aligned}\varphi(a) = & \textit{abcacdcbcadcdbdabacabadbabcbdbcbacbcdcacbab} \\ & \textit{dabacadcdbcacdbcbacbcdcacdcdbdcdadbdcba},\end{aligned}$$

а образы остальных символов получаются из $\varphi(a)$ циклическим сдвигом символов.

Тогда слово $w = \varphi(w)$ избегает абелевых квадратов.

Абелевы четвертые степени избегаемы над алфавитом мощности 2, а кубы — над алфавитом мощности 3 (Dekking, 1979).

2-4

3-3

4-2

Рассмотрим алфавит $\{0, 1, \dots, n\}$ и назовем весом слова над этим алфавитом сумму его букв.

Рассмотрим алфавит $\{0, 1, \dots, n\}$ и назовем весом слова над этим алфавитом сумму его букв.

Вопрос

Можно ли избежать — для некоторого $k > 0$ — k идущих подряд слов одного веса?

Рассмотрим алфавит $\{0, 1, \dots, n\}$ и назовем весом слова над этим алфавитом сумму его букв.

Вопрос

Можно ли избежать — для некоторого $k > 0$ — k идущих подряд слов одного веса?

Ответ

НЕТ (теорема Семереди).

Рассмотрим алфавит $\{0, 1, \dots, n\}$ и назовем весом слова над этим алфавитом сумму его букв.

Рассмотрим алфавит $\{0, 1, \dots, n\}$ и назовем весом слова над этим алфавитом сумму его букв.

Вопрос

Можно ли избежать — для некоторого $k > 0$ — k идущих подряд слов одного веса **и одной длины**?

Рассмотрим алфавит $\{0, 1, \dots, n\}$ и назовем весом слова над этим алфавитом сумму его букв.

Вопрос

Можно ли избежать — для некоторого $k > 0$ — k идущих подряд слов одного веса **и одной длины**?

Ответ

ДА для $k = 4$: бинарное слово, избегающее 4-х абелевых степеней, подходит.

Рассмотрим алфавит $\{0, 1, \dots, n\}$ и назовем весом слова над этим алфавитом сумму его букв.

Вопрос

Можно ли избежать — для некоторого $k > 0$ — k идущих подряд слов одного веса **и одной длины**?

Ответ

ДА для $k = 4$: бинарное слово, избегающее 4-х абелевых степеней, подходит.

Как насчет $k = 3$? $k = 2$?

Theorem (J. Cassaigne, J. Currie, L. Schaeffler, J. Shallit, 2011)

Существует бесконечное слово над алфавитом $\{0, 1, 3, 4\}$, избегающее весовых кубов.

Theorem (J. Cassaigne, J. Currie, L. Schaeffler, J. Shallit, 2011)

Существует бесконечное слово над алфавитом $\{0, 1, 3, 4\}$, избегающее весовых кубов.

Это неподвижная точка морфизма

$$\varphi : \begin{cases} 0 \mapsto 03, \\ 1 \mapsto 43, \\ 3 \mapsto 1, \\ 4 \mapsto 01. \end{cases}$$

Theorem (J. Cassaigne, J. Currie, L. Schaeffler, J. Shallit, 2011)

Существует бесконечное слово над алфавитом $\{0, 1, 3, 4\}$, избегающее весовых кубов.

Это неподвижная точка морфизма

$$\varphi : \begin{cases} 0 \mapsto 03, \\ 1 \mapsto 43, \\ 3 \mapsto 1, \\ 4 \mapsto 01. \end{cases}$$

03143011034343031011011...

Существует ли слово над конечным алфавитом из натуральных чисел, избегающее весовых квадратов?

Существует ли слово над конечным алфавитом из натуральных чисел, избегающее весовых квадратов?

Известно, что имеющиеся примеры слов, избегающих абелевы квадраты, не подойдут.

- Автоматные слова

Начало в 19.05

