МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. Ломоносова

Механико-математический факультет

На правах рукописи

УДК 519.72

Румянцев Андрей Юрьевич

О колмогоровской сложности конечных подпоследовательностей в последовательности нулей и единиц.

Специальность 01.01.06 — математическая логика, алгебра и теория чисел

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва
2009

Работа выполнена на кафедре математической логики и теории алгоритмов Механико-математического факультета Московского государственного университета имени М. В. Ломоносова.

Научный руководитель:	доктор физико-математических наук, профессор Верещагин Николай Константинович.
Официальные оппоненты:	доктор физико-математических наук, профессор Вьюгин Владимир Вячеслович; кандидат физико-математических наук Фрид Анна Эдуардовна.
Ведущая организация:	Институт вычислительных технологий Сибирского Отделения РАН.

Защита диссертации состоится 6 ноября 2009 года в 16 часов 45 минут на заседании диссертационного совета Д.501.001.84 при Московском государственном университете имени М. В. Ломоносова по адресу: Р.Ф., 119991, Москва, ГСП-1, Ленинские горы, д. 1, МГУ, Механико-математический факультет, аудитория 14-08.

С диссертацией можно ознакомиться в библиотеке Механико-математического факультета МГУ (Главное здание, 14 этаж).

Автореферат разослан 6 октября 2009 года.

Ученый секретарь диссертационного
совета Д.501.001.84 при МГУ,
доктор физико-математических
наук, профессор А. О. Иванов

Общая характеристика работы

Актуальность темы.

Диссертация посвящена проблеме построения последовательностей, не нарушающих определённых запрещений; рассматриваются вероятностный, теоретико-сложностной, игровой подходы к данной проблеме.

Теория последовательностей, избегающих различные запрещения, возникла в начале XX века. А. Туэ доказал существование последовательностей, не содержащих квадратов (подслов вида xx, где x — некоторое непустое слово) в троичном алфавите и кубов (подслов вида xxx), а также частичных кубов (подслов вида $xyxyx$, где x и y — непустые слова), в двоичном алфавите (последовательность Туэ–Морса)[1,2].

Позже началось систематическое исследование последовательностей, не содержащих подслов по определённым шаблонам (например, подслов вида $xxyyzz$, где x, y и z — непустые слова). Недавно эти результаты стали обобщать на частичные последовательности (последовательности, у которых в некоторые местах стоит специальный символ пропуска, означающий, что значение в данной позиции не известно), например, была построена последовательность (без пропусков), которая остаётся последовательностью без кубов при замене любого количества символов на пропуски так, чтобы между соседними пропусками было не менее двух символов (последовательность с пропусками является последовательностью без кубов, если в ней нет подслов, получающихся заменой некоторых символов на пропуски в словах вида xxx, где x — некоторое непустое слово)[3].

Также, было введено понятие критической экспоненты последовательности — минимальной верхней грани всех показателей дробных степеней слов, входящих в последовательность (дробная степень слова x с показателем r — это слово вида $xxx\ldots xxy$, где x повторяется столько раз, какова целая часть числа r, а y — префикс слова x, длина которого равна дробной части r, умноженной на длину x). Продолжают активно изучаться последовательности с ограничениями на критическую экспоненту, т.е. с запрещениями на подслова, являющиеся степенями с определёнными показателями. В 2007 году Д. Кригер и Дж. Шаллит нашли метод построения последовательностей с заданной критической экспонентой (для любого числа, большего единицы)[4].

---

[1] Axel Thue, *Über unendliche Zeichenreihen*, Norske Vid. Skrifter I Mat.-Nat. Kl., Christiania 7 (1906) 1–22.

[2] Axel Thue, *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*, Norske Vid. Skrifter I Mat.-Nat. Kl., Christiania 1 (1912) 1–67.

[3] Florin Manea, Robert Merca, *Freeness of partial words*, Theoretical Computer Science 389, Issue 1-2 (December 2007), pp. 265–277.

[4] Dalia Krieger, Jeffrey Shallit, *Every real number greater than 1 is a critical exponent*, Theoretical Computer Science, 381 (issue 1-3, August 2007), pp. 177–182.

В 60-х годах XX века было введено понятие колмогоровской сложности. Грубо говоря, колмогоровская сложность строки — это количество бит в минимальной программе, печатающей эту строку при пустом входе. Появились работы, касающиеся последовательностей, не нарушающих запрещений в каком-либо смысле малой сложности. Так теорема Левина–Шнорра утверждает, что случайная по Мартин-Лёфу последовательность не содержит префиксов малой сложности (вариант с префиксной сложностью появился в 1975-м году[5]). Позже Левин[6] в одной из своих работ по замощениям плоскости доказал лемму о том, что существует последовательность, не содержащая подслов малой сложности. В 2003 году Мучник, Семёнов и Ушаков[7] разработали метод построения почти-периодической последовательности без префиксов малой сложности.

Цель работы.

Получить достаточные условия на возможные запрещения для существования последовательности, не нарушающей все эти запрещения и применение полученных результатов для некоторых задач о построении последовательностей.

Методы исследования.

В работе применяются методы теории вероятностей и колмогоровской сложности. Используется Локальная Лемма Ловаса для доказательства совместности событий. Для доказательства некоторых результатов использован метод построения последовательности по частям.

Научная новизна.

Результаты диссертации являются новыми. В диссертации получены следующие основные результаты:

1. Приведено несколько достаточных критериев для возможности построения последовательности, удовлетворяющей определённым ограничениям.

2. Получены отрицательные результаты, ограничивающие возможность построения последовательностей: доказано, что невозможно построить несколько последовательностей по циклу, каждая из которых обладает

[5]Gregory J. Chaitin, *A Theory of Program Size Formally Identical to Information Theory*, Journal of the ACM, 22 (issue 3, July 1975), pp. 329–340

[6]Bruno Durand, Leonid Levin, Alexander Shen, *Complex tilings*, STOC Proceedings, 2001, pp. 732–739; enhanced version: `http://arXiv.org/abs/cs.CC/0107008`

[7]Andrei Muchnik, Alexei Semenov and Maxim Ushakov, *Almost periodic sequences*, Theoretical Computer Science, 304 (issue 1-3, July 2003), pp. 1–33.

условию обобщённой леммы Левина с взятой в качества оракула следующей по циклу последовательностью; доказана невозможность построения последовательности, удовлетворяющей условию обобщённой леммы Левина вместе со своими вычислимыми перестановками (из "основного" достаточного критерия, доказанного в данной работе, следует существование последовательности, удовлетворяющей условию обычной леммы Левина вместе со своими вычислимыми перестановками); доказано, что в одном из игровых вариантов "основного" достаточного критерия игрок, пытающийся построить последовательность, удовлетворяющую некоторым ограничениям, проигрывает.

3. Полученные критерии применены для обобщения теоремы Кригер и Шаллита на частичные последовательности и для построения почти периодических последовательностей, не содержащих подслов малой сложности, а также, для построения многомерного аналога почти периодических последовательностей, не содержащих многомерных подслов малой сложности (т.е. содержимое любого прямоугольного параллелепипеда должно иметь большую сложность).

Теоретическая и практическая ценность.

Работа носит теоретический характер. Результаты, полученные в диссертационной работе, являются развитием методов комбинаторики последовательностей и могут применяться для построения последовательностей с определёнными свойствами. Они могут быть полезны в теории информации, колмогоровской сложности и комбинаторике последовательностей.

Апробация работы.

Результаты диссертации докладывались на следующих семинарах и конференциях:

- На Колмогоровском семинаре кафедры математической логики и теории алгоритмов механико-математического факультета Московского Государственного Университета имени М. В. Ломоносова под руководством профессора Н. К. Верещагина, профессора А. Л. Семёнова, к.ф.м.н. А. Е. Ромащенко и к.ф.м.н. А. Шеня в 2006 году.

- На международной конференции "Симпозиум по теории и приложениям компьютерных наук" (STACS-2006), Марсель, Франция, 23–25 февраля 2006 года.

- На международной конференции “Колмогоровская сложность и приложения” (Dagstuhl Seminar 06051), Дагштуль, Германия, 29 января – 3 февраля 2006 года.
- На международной конференции “Компьютерные науки в России” (CSR-2007), Екатеринбург, Россия, 3–7 сентября 2007 года.

Публикации.
Основные результаты диссертации опубликованы в трёх работах [1–3].

Структура работы.
Работа состоит из введения, 4 глав, содержащих 14 разделов, и списка литературы. Библиография содержит 17 наименований. Текст диссертации изложен на 88 страницах и содержит 6 иллюстраций.

Краткое содержание диссертации

Предположим, что нам нужно построить бесконечную последовательность нулей и единиц, на которую наложены некоторые ограничения (не содержать подслов определённого вида или не содержать данных комбинаций битов в данных позициях и т.п.)

Если этих ограничений в каком-то смысле немного (отбраковывается мало последовательностей), то можно надеяться, что искомая последовательность существует по количественным соображениям. Например, если суммарная вероятность всех забракованных множеств меньше единицы, то заведомо есть незабракованная последовательность (вероятностное доказательство существования).

В данной работе рассматриваются обобщения этого метода, использующие лемму Ловаса, их переформулировка на языке колмогоровской сложности и применения: построение последовательностей с заданной критической экспонентой и построение почти периодических последовательностей со сложными подсловами.

В этой работе мы применяем колмогоровскую сложность для доказательства комбинаторных утверждений. Неформально говоря, колмогоровская сложность слова из нулей и единиц — это число битов в кратчайшей программе, печатающей данное слово при пустом входе. Колмогоровская сложность определена с точностью до постоянного слагаемого (см. подробнее в книге Ли и Витаньи[8]). Можно говорить и о колмогоровской сложности объ-

[8]Li M., Vitanyi P, *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd ed. N.Y.: Springer, 1997.

ектов, не являющихся двоичными словами, надо лишь вычислимо закодировать их такими словами. Например, сложность натурального числа можно определить как сложность его двоичной записи. Обычную колмогоровскую сложность объекта a в данной работе будем обозначать $\mathrm{K}(a)$, префиксную сложность будем обозначать $\mathrm{KP}(a)$.

Пусть ω — бесконечная последовательность нулей и единиц, а $X \subset \mathbb{N}$ — конечное множество индексов. Через $\omega(X)$ мы обозначим двоичное слово, составленное из битов ω с индексами из множества X (записанных слева направо). Мы будем рассматривать ограничения на последовательность ω, имеющие такой вид: для данного множества индексов X и для данного слова Y требуется, чтобы $\omega(X) \neq Y$. Таких ограничений не должно быть слишком много; переходя от количественных формулировок к сложностным, мы требуем, чтобы эти ограничения были в том или ином смысле "просты".

Во **введении** приводятся основные понятия и обозначения и даётся краткое содержание диссертации.

В **главе 1** мы приводим простую переформулировку вероятностного метода (ограничения совместны, если сумма вероятностей запрещённых объектов меньше единицы) в терминах Колмогоровской сложности.

В **разделе 1.1** приводятся необходимые определения и свойства колмогоровской и префиксной сложности, а также, случайности по Мартин-Лёфу.

В **разделе 1.2** даётся собственно переформулировка вероятностного метода доказательства существования: берём случайную последовательность (в смысле алгоритмической теории вероятностей, то есть определения Мартин-Лёфа) и проверяем, что она удовлетворяет всем ограничениям. В частности, по теореме Левина–Шнорра (в варианте для префиксной сложности) случайная последовательность не содержит префиксов с маленькой сложностью.

Теорема 1 (Левин–Шнорр)**.** *Последовательность ω случайна по Мартин-Лёфу тогда и только тогда, когда для некоторого c и для любого целого положительного числа n выполнено неравенство*

$$\mathrm{KP}(\omega([0,n))) \geqslant n - c.$$

Здесь $[0,n)$ обозначает промежуток $\{0,1,2,\ldots,n-1\}$, так что $\omega([0,n))$ есть n-битовое начало последовательности. Мы усиливаем эту теорему в одну сторону, доказывая более сильное свойство случайных последовательностей:

Теорема 2. *Пусть последовательность ω случайна по Мартин-Лёфу. Тогда для любого конечного множества $X \subset \mathbb{N}$ верно неравенство*

$$\mathrm{KP}(\omega(X), X) \geqslant |X| - O(1).$$

Если $X = [0, n)$, то $\omega(X)$ определяет X, второй член пары не нужен и мы получаем утверждение "только тогда" теоремы Левина–Шнорра в качестве частного случая. Теорема 2 даёт критерий случайности, инвариантный относительно вычислимых перестановок последовательности.

В **главе 2** мы приводим различные более интересные результаты. В **разделе 2.1** мы приводим лемму Левина — простой пример, когда вероятностного метода недостаточно, — а затем обобщаем её.

Теорема 3 (лемма Левина)**.** *Для любого числа $0 < \alpha < 1$ существует такая последовательность ω, что для любых n и k выполнено неравенство*

$$\mathrm{K}(\omega([k, k+n))) \geqslant \alpha n - O(1).$$

Теорема 4 (обобщённая лемма Левина)**.** *Для любого числа $0 < \alpha < 1$ существует такая последовательность ω, что для любых n и k выполнено неравенство*

$$\mathrm{K}(\omega([k, k+n)) \mid \omega([0, k))) \geqslant \alpha n - O(1).$$

Обобщение леммы Левина опубликовано в [2] (как утверждение 8).

В **разделе 2.2** обсуждаются комбинаторная и игровая переформулировки леммы Левина и её обобщения и доказывается их эквивалентность сложностным вариантам (как релятивизованным, так и нерелятивизованным).

Теорема 5 (комбинаторный вариант леммы Левина)**.** *Для любого числа $0 < \alpha < 1$ существует число N со следующим свойством. Пусть A — произвольное множество слов длины не менее N, элементы которого мы называем "запрещёнными", причём среди слов длины n не больше $2^{\alpha n}$ запрещённых. Тогда найдётся последовательность ω, не содержащая запрещённых подслов.*

Эквивалентность комбинаторного и сложностного вариантов леммы Левина доказана в [2] (эквивалентность утверждений 1 и 2).

Теорема 6 (комбинаторный вариант обобщённой леммы Левина)**.** *Для любого числа $0 < \alpha < 1$ существует число N со следующим свойством. Пусть A — произвольное множество пар слов. И пусть вторые компоненты пар имеют длину не менее N, и для каждого слова x и каждого числа n существует не более $2^{\alpha n}$ слов y длины n, при которых $(x, y) \in A$. Тогда существует последовательность ω, не имеющая префиксов вида xy, где $(x, y) \in A$.*

Далее мы переформулируем утверждение теоремы 6 в терминах игры. Первый игрок строит последовательность бит за битом, второй между каждыми двумя ходами первого, а также перед первым его ходом может запретить появление некоторых подслов в некоторых позициях, если эти позиции целиком находятся в ещё не построенной части последовательности. При этом второй может запрещать лишь слова длины не менее N, и в каждой позиции длины n запретить не более $2^{\alpha n}$ слов. Первый игрок стремится построить бесконечную последовательность, не нарушающую ни одного запрещения, а цель второго — помешать ему.

Теорема 7 (игровая переформулировка обобщённой леммы Левина)**.** *Для любого числа $0 < \alpha < 1$ существует такое число N, что в описанной игре выигрывает первый игрок.*

В **разделе 2.3** мы формулируем лемму Ловаса[9] и с её помощью доказываем более интересный критерий совместности условий.

Теорема 8 (Локальная лемма Ловаса)**.** *Пусть $G = (V, E)$ — неориентированный конечный граф, V — множество вершин, а E — множество рёбер. Пусть для каждой вершины указано событие H_v в некотором вероятностном пространстве (одном и том же для всех v) и число $p_v \in (0, 1)$. Предположим, что для любого v событие H_v независимо со случайной величиной, составленной из всех событий H_u для u, не соседних с v, и выполнено неравенство:*

$$\Pr(H_v) \leqslant p_v \cdot \prod_{(v,u)\in E} (1 - p_u).$$

Тогда $\Pr\left(\bigcap\limits_{v\in V} \overline{H_v}\right) \geqslant \prod\limits_{v\in V} (1 - p_v)$*, и, следовательно, это событие не пусто.*

Применяя эту лемму для множества всех двоичных последовательностей с равномерной бернуллиевской мерой к событиям вида $\omega(X) = x$ для некоторых пар (x, X), где X — конечное множество индексов (натуральных чисел), а x — двоичное слово, мы доказываем следующий комбинаторный результат:

Теорема 9. *Для любого числа $0 < \alpha < 1$ существует число N со следующим свойством. Пусть A — некоторое множество пар вида (x, X), называемых "запрещёнными", где X — конечное множество натуральных чисел (индексов), а x — слово длины $|X|$. Пусть при этом все слова x в запрещённых парах имеют длину не менее N, и для каждого индекса t и для каждого*

[9] Rajeev Motwani, Prabhakar Raghavan, *Randomized algorithms*, Cambridge University Press, New York, NY, 1995.

числа n количество запрещённых пар с множествами размера n, содержащими позицию t, не больше $2^{\alpha n}$:

$$\forall t, n \,|\{(x, X) \in A \mid \ |X| = n \quad \text{и} \quad t \in X\}| \leqslant 2^{\alpha n}.$$

Тогда существует последовательность ω, не нарушающая ни одного запрещения (для всех $(x, X) \in A$ выполнено $\omega(X) \neq x$).

Сложностной аналог этой теоремы можно сформулировать следующим образом:

Теорема 10. *Для любого $0 < \alpha < 1$ существует последовательность ω и число c с такими свойствами: в любом конечном непустом множестве $X \subset \mathbb{N}$ найдётся элемент $t \in X$, для которого*

$$\mathrm{K}(\omega(X), X \mid t) \geqslant \alpha|X| - c.$$

Эта теорема опубликована в [3] (как теорема 1).

По сравнению с теоремой 2 мы усилили утверждение, использовав условную сложность, но одновременно и ослабили его, умножив правую часть на $\alpha < 1$.

В частности из этого утверждения легко получается многомерный аналог леммы Левина, доказанный М. А. Ушаковым в [2]:

Теорема 12. *Пусть $\alpha \in (0, 1)$. Для любого $d \geqslant 1$ существует d-мерная "последовательность" (отображение $\mathbb{Z}^d \to \{0, 1\}$), в которой сложность содержимого любого (целочисленного) параллелепипеда объёма V не меньше $\alpha|V| - O(1)$.*

Из инвариантности утверждения теоремы 10 относительно вычислимых перестановок последовательности получаем такое утверждение:

Теорема 13. *Для любого числа $0 < \alpha < 1$ существует такая последовательность ω, что для любой вычислимой биекции $\sigma\colon \mathbb{N} \to \mathbb{N}$ перестановка ω с помощью σ, то есть последовательность $\nu_i = \omega_{\sigma(i)}$, обладает свойством из леммы Левина: для любых n и k выполнено неравенство*

$$\mathrm{K}(\nu([k, k + n))) \geqslant \alpha n - O(1).$$

При этом константа в $O(1)$ зависит от перестановки, но не зависит от n и k.

Далее теорема 10 обобщается на случай произвольного алфавита.

Теорема 14. *Для любого алфавита A, $2 \leqslant |A| < \infty$ и для любого $0 < \alpha < 1$ существует последовательность ω в алфавите A и число c с такими свойствами: в любом конечном непустом множестве $X \subset \mathbb{N}$ найдётся элемент $t \in X$, для которого*

$$\mathrm{K}(\omega(X), X \mid t) \geqslant \alpha |X| \log |A| - c.$$

В **главе 3** доказываются отрицательные результаты, показывающие невозможность некоторых естественных усилений результатов предыдущих разделов (невозможно совместить в одном утверждении обобщение леммы Левина и критерий, который даёт теорема 10). В **разделе 3.1** приводится доказательство того, что нельзя усилить обобщённую лемму Левина так, чтобы она стала инвариантна относительно вычислимых перестановок.

Рассмотрим (очевидно, вычислимую) перестановку

$$\sigma = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \cdots \\ 0 & 2 & 1 & 5 & 4 & 3 & 9 & 8 & 7 & 6 & \cdots \end{pmatrix},$$

которая разбивает натуральный ряд на отрезки длин $1, 2, 3, 4, \ldots$ и переворачивает каждый отрезок в обратном порядке.

Теорема 18. *Для любой двоичной последовательности ω в одной из последовательностей ω и $\sigma\omega$ найдётся сколь угодно длинное подслово сложности $O(1)$ относительно соответствующего префикса.*

Таким образом, инвариантный относительно вычислимых перестановок вариант обобщённой леммы Левина не верен (аналогичный инвариантный вариант обычной леммы Левина верен в силу теоремы 13).

Далее доказывается, что не существует пара последовательностей, каждая из которых обладала бы свойством теоремы 4 с взятой в качестве оракула другой последовательностью. А также пары последовательностей, каждая из которых обладала свойством теоремы 10 с взятой в качестве оракула другой последовательностью.

Теорема 20. *Для любой пары бесконечных последовательностей нулей и единиц хотя бы одна из них содержит сколь угодно длинные подслова сложности $O(1)$ относительно их позиции с взятой в качестве оракула другой последовательностью.*

В **разделе 3.2** мы формулируем игровой аналог утверждения теоремы 9 (подобно тому как теорема 7 была игровым вариантом теоремы 4).

Первый игрок, как и раньше, строит последовательность ω бит за битом. Второй игрок между каждыми двумя ходами первого (а также перед первым

его ходом) может запретить выполнение равенства $\omega(X) = x$ для некоторых пар (x, X), где X — конечное множество индексов, не пересекающееся с уже построенным началом последовательности, а x — слово длины $|X|$. При этом запрещения должны удовлетворять ограничениям, описанным в формулировке теоремы 9. Первый игрок стремится построить бесконечную последовательность, не нарушающую ни одного запрещения, а цель второго — помешать ему. Доказывается, что первый игрок не может выиграть в этой игре, даже если дополнительно ограничить права второго игрока, разрешив ему делать только запрещения со словами x некоторой фиксированной длины N и потребовав, чтобы каждый индекс t входил не более чем в два запрещения.

Теорема 21. *Для любого числа $N > 0$ и любого $\alpha < 1$ в описанной игре первый игрок проигрывает.*

В **главе 4** содержатся приложения результатов диссертации. В **разделе 4.1** мы обобщаем результат Д. Кригер и Дж. Шаллита, используя теорему 10.

Пусть a — некоторое слово (не обязательно в двоичном алфавите), а b — его префикс (начало). Тогда слово $c = aaa...aab$ называется *дробной степенью* слова a, а число $r = \frac{|c|}{|a|}$ — *показателем* этой степени. Пишут $c = a^r$. Критическая экспонента некоторой последовательности ω — это точная верхняя грань всех показателей степеней, являющихся подсловами в этой последовательности. Д. Кригер и Дж. Шаллит доказали следующую теорему:

Теорема 27 (Кригер, Шаллит)**.** *Для любого числа $r > 1$ существует последовательность ω (не обязательно в двоичном алфавите), которая имеет критическую экспоненту r.*

Мы для данного числа $r > 1$ строим последовательность, которая содержит степени слов с показателями, меньшими r, и не содержит степеней слов с показателями, большими r, и, более того, даже подслов, близких в смысле расстояния Хемминга к степеням слов с показателями, большими r.

Теорема 28. *Для любого действительного числа $r > 1$ существует число $\varepsilon > 0$ и последовательность, критическая экспонента которой равна r, которая ни для какого $s > r$ не содержит подслова, $\min\left\{\varepsilon, \frac{s-r}{s}\right\}$-близкого к степени с показателем s.*

Эта теорема усиливает результат, опубликованный в [3] (там доказывалась теорема с более слабым условием: последовательность ни для какого $s > r + \delta$

не содержит подслова, ε-близкого к степени с показателем s, где δ — положительное число, которое можно выбрать сколь угодно малым, а ε зависит как от r, так и от δ).

Аналогичное утверждение имеет место для частичных последовательностей, при этом малое расстояние Хемминга будет соответствовать малому числу пропусков. Частичная последовательность в некотором алфавите — это частичное отображение из множества натуральных чисел в этот алфавит. Частичное слово — это частичное отображение из начального отрезка натуральных чисел в алфавит (слово, у которого определены не все буквы), удобно при этом при записи вместо неопределённой буквы писать специальный символ пропуска $\diamond$. Будем говорить, что одно частичное слово включает другое, если длины этих слов одинаковы, и выполнено соответствующее включение одного частичного отображения в другое. Критическая экспонента частичной последовательности — это точная верхняя грань всех показателей степеней, включающих подслова в этой последовательности, за исключением (вырожденных) включений подслов вида $x_0x_1\dots x_{k-1}\diamond$ и $\diamond x_1\dots x_{k-1}x_0$ в степень с показателем $(k+1)/k$, где $x_0,\dots,x_{k-1}\neq\diamond$ (такие включения в степени невозможно избежать, если в последовательности есть неопределённые символы). Недавно было доказано, что существует более, чем счётное количество бесквадратных (за исключением квадратов вида $a\diamond$ и $\diamond a$, где $a\neq\diamond$) частичных последовательностей с бесконечным количеством пропусков в троичном алфавите[10], т.е. существуют частичные последовательности с критической экспонентой, не большей двух. Мы доказываем близкую теорему, про критическую экспоненту частичных слов (не ограничивая размер алфавита):

Теорема 29. *Для любого действительного числа $r>1$ существует число $M\in\mathbb{N}$ и (всюду определённая) последовательность, критическая экспонента которой равна r и не изменяется при замене любого количества символов в последовательности на пропуски так, чтобы между соседними пропусками было не меньше M символов.*

Эта теорема не является прямым следствием предыдущей, однако, её доказательство основано на тех же идеях.

В **разделе 4.2** мы приводим конструкцию почти периодической последовательности, не имеющей простых подслов (то есть подслов x, для которых $\mathrm{K}(x)<\alpha|x|-O(1)$ для некоторой константы $\alpha<1$) и её многомерный аналог.

[10] Vesa Halava, Tero Harju and Tomi Kärki, *Square-free partial words*, Information Processing Letters, Volume 108, Issue 5 (15 November 2008), pp. 290–292.

Мы называем d-мерными двоичными последовательностями, отображения вида $v : \mathbb{Z}^d \to \mathbb{B}$, d-мерным словом — прямоугольный параллелепипед из символов (без фиксированной позиции его в пространстве), размером d-мерного слова — число символов в нём, а подсловами d-мерной последовательности — d-мерные слова, взятые из прямоугольных параллелепипедов в последовательности.

Бесконечную d-мерную последовательность v из нулей и единиц мы называем почти периодичной, если для любого подслова в этой последовательности найдётся такое число k, что это подслово входит в любой куб со стороной k в этой последовательности (это определение соответствует сильной почти-периодичности в одномерном случае).

Теорема 30. *Для любого $\alpha \in (0, 1)$ существует почти периодическая d-мерная последовательность $v : \mathbb{Z}^d \to \mathbb{B}$, у которой сложность любого подслова размера m не меньше $\alpha m - O(1)$.*

Эта теорема усиливает результат, опубликованный в [2] (там она была доказана только для кубов, а не для произвольных прямоугольных параллелепипедов).

В **разделе 4.3** изучается величина[11] $R(a, l)$, где $a \geqslant 2$ и $l \geqslant 1$, определяемая как точная нижняя грань всех таких чисел r, для которых существует последовательность в a-символьном алфавите, не содержащая дробных степеней слов x^p для $|x| \geqslant l$ и $p \geqslant r$. Теперь мы рассматриваем дробные степени с ограничением на длину слова x.

Сначала доказывается простая нижняя оценка:

$$R(a, l) \geqslant 1 + \frac{1}{la}.$$

Затем доказывается обобщение на случай произвольного алфавита известной теоремы[12].

Теорема 31. *Для любого размера алфавита $a \geqslant 2$ и любого положительного действительного числа $b < a$ существует число N и последовательность ω в алфавите из a символов такая, что для любого $n \geqslant N$ расстояние между любыми двумя различными вхождениями одного и того же слова длины n в эту последовательность будет не меньше b^n.*

[11] Lucian Ilie, Pascal Ochem, Jeffrey Shallit, *A Generalization of Repetition Threshold*, Mathematical foundations of computer science, 345, Issue 2-3 (November 2005), pp. 359–369.

[12] J. Berk, An application of Lovász local lemma: there exists an infinite 01-sequence containing no near identical intervals. In A. Hajnal, L. Lovász, and V. T. Sós, editors, *Finite and Finite Sets*, Vol. 37 of Colloq. Math. Soc. János Bolyai, 1981, pp. 103–107.

В частности, из этой теоремы получается такая верхняя оценка: для любого целого a и действительного b таких, что $1 < b < a$, для достаточно больших l выполнено

$$R(a,l) < 1 + \frac{\log_b l}{l}.$$

После этого даётся следующая усиленная оценка для функции R.

Теорема 33. *Существует такая константа c, что для любых $a \geqslant 2$ и $l \geqslant 1$ выполнено*

$$1 + \frac{1}{al} \leqslant R(a,l) \leqslant 1 + \frac{c}{al}.$$

В **разделе 4.4** приводится ещё один пример применения колмогоровской сложности для доказательства комбинаторных утверждений. Мы формулируем на языке колмогоровской сложности результаты из теории кодирования.

Одним из важных понятий теории кодирования является декодирование списком. Пусть f — функция из $\{0,1\}^k$ в $\{0,1\}^n$, называемая фукнцией кодирования. Тогда декодирование списком, исправляющее pn ошибок, где $0 < p < 1$, это отображение, сопоставляющее каждому двоичному слову x длины n список всех слов длины k, для которых образ под действием функции f лежит на расстоянии (по Хеммунгу) не более pn от слова x. Ясно, что если существуют такие 2^k слов длины n, что любой шар Хемминга радиуса pn содержит не более L из них, то существует функция кодирования f позволяющая декодировать списком длины не более L.

Говоря о колмогоровской сложности, мы рассматриваем "устойчивые" относительно изменения не более чем в pn битах слова, т.е. такие слова x, для которых $\mathrm{K}(x \mid y)$ мало для любого слова y, отличающегося от x не более чем в pn позициях. Связь существования таких слов с возможностью декодирования списком (ограниченной длины) показывается такой теоремой.

Теорема 35. *Пусть фиксированы $p < \frac{1}{2}$, $k < n$ и $L < 2^n$. Тогда справедливы следующие утверждения.*

А. *Если существуют такие 2^k слов длины n, что любой шар Хемминга радиуса pn содержит не более L из них, то существует такое слово x длины n и сложности не менее $k - O(\log n)$, что для любого слова y, отличающегося от x не более чем в pn битах, имеет место оценка $\mathrm{K}(x \mid y) \leq \log L + O(\log n)$.*

Б. *Если существует такое слово x длины n и сложности не менее k, что для любого слова y, отличающегося от x не более чем в pn битах, $\mathrm{K}(x \mid y) \leq \log L$, то существуют такие $2^{k - O(\log n)}$ слов длины n, что любой шар Хемминга содержит не более $L \cdot 2^{O(\log n)}$ из них.*

Эта теорема опубликована в [1] (как теорема 2).

Из теории кодирования известно[13] существование таких множеств (как в пункте Б теоремы) с определёнными значениями констант. Мы, однако, приводим другой вариант доказательства этого утверждения, основанный на колмогоровской сложности.

Теорема 36. *Пусть фиксировано число $p < 1/2$ и $\alpha = 1 - H(p)$. Тогда для любого n можно найти слово x длины n, для которого*

(а) $\mathrm{K}(x) = \alpha n + O(\log n)$*;*

(б) $\mathrm{K}(x \mid y) = O(\log n)$ *для любого слова y, отличающего от x не более чем в pn позициях.*

(Здесь константа в обозначении $O(\log n)$ может зависеть от p, но не от n и y.) Эта теорема опубликована в [1] (как теорема 3).

Благодарности.

Автор благодарит своих научных руководителей д.ф.м.н., профессора Н. К. Верещагина и к.ф.м.н. А. Шеня за постановку задач и постоянную помощь. Автор благодарен руководителям и участникам Колмогоровского семинара, а также сотрудникам кафедры математической логики и теории алгоритмов за полезные обсуждения работы.

Список литературы

[1] А. Ю. Румянцев, *Передача информации по каналу с ошибками с точки зрения колмогоровской сложности*, Вестник Московского университета, Серия 1, Математика, Механика, 2006, 1.С., стр. 54–56.

[2] Andrey Yu. Rumyantsev, Maxim A. Ushakov, *Forbidden Substrings, Kolmogorov Complexity and Almost Periodic Sequences*, Springer, Lecture Notes in Computer Science, Volume 3884 / 2006, STACS 2006, pp. 396–407.

[3] Andrey Yu. Rumyantsev, *Kolmogorov Complexity, Lovász Local Lemma and Critical Exponents*, Springer, Lecture Notes in Computer Science, Volume 4649 / 2007, CSR 2007, pp. 349–355.

В работе [2] А. Ю. Румянцеву принадлежат утверждения 6–9, а также лемма и следствие в конце раздела 4; М. А. Ушакову принадлежит утверждение 4.

[13] Мак-Вильямс Ф. Дж., Слоэн Н. Дж. А., *Теория кодов, исправляющих ошибки*, М.: Радио и связь, 1979.