Московский государственный университет имени М. В. Ломоносова
Механико-математический факультет

На правах рукописи
УДК 510.5+519.1

Притыкин Юрий Львович

Алгоритмические свойства последовательностей, близких к периодическим

специальность 01.01.06 —
математическая логика, алгебра и теория чисел

Диссертация
на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
член-корреспондент РАН,
профессор Алексей Львович Семёнов

Москва — 2009

Оглавление

Глава 1.

Введение

1.1. Актуальность темы и цель работы

В работе мы рассматриваем разного типа обобщения понятия бесконечной периодической последовательности и исследуем их свойства.

Центральное понятие работы — почти периодические последовательности. Последовательность символов конечного алфавита почти периодическая, если для каждого конечного подслова u последовательности найдётся такое натуральное число n, что в каждом подслове длины n последовательности найдётся вхождение слова u.

Впервые почти периодические последовательности были рассмотрены в связи с символической динамикой. Продолжая и ссылаясь на работы Адамара и Пуанкаре, Морс в 1921 году опубликовал работу [24], в которой среди прочего определил почти периодические последовательности (он называл их рекуррентными). В 1938 — 1940 годах вышли работы Морса и Хедлунда "Символическая динамика" [22, 23], в которых многие свойства почти периодических последовательностей были подробно исследованы.

Идея символической динамики заключается в сопоставлении траектории в непрерывной динамической системе последовательности букв конечного алфавита. Понимая траекторию как движение точки в некотором пространстве, мы, во-первых, выделяем в пространстве конечное количество областей и, во-вторых, выбираем дискретное множество моментов времени (например, через каждую фиксированную единицу времени), после чего смотрим, в каких областях оказывается точка в эти моменты. Траектории точки сопоставляется бесконечная последовательность букв конечного алфавита, где буквы соответствуют областям пространства.

Более формально, *топологическая динамическая система* — это топологическое пространство V с заданным на нём непрерывным отображением $f\colon V \to V$. Рассмотрим V, f — топологическую динамическую систему, $A_1, \ldots, A_k$ — попарно непересекающиеся открытые подмножества в V, и p — точку, орбита которой $\{f^n(p) : n \in \mathbb{N}\}$ лежит в $\bigcup_{i=1}^k A_k$. Определим последовательность $x\colon \mathbb{N} \to \{1, \ldots, k\}$ условием $f^n(p) \in A_{x(n)}$. Таким образом, (некоторым) точкам пространства V мы сопоставили последовательности букв алфавита $A = \{1, \ldots, k\}$. При этом применению отображения f соответствует операция левого сдвига L: $L(a_0a_1a_2a_3\ldots) = a_1a_2a_3\ldots$

Ясно, что периодические последовательности описывают только самые простые ситуации. Однако оказывается, что в довольно широком классе ситуаций возникающая символическая последовательность будет близка к периодической. Это наблюдение можно формализовывать разными способами. Вот один из них. Результат этой теоремы, видимо, уже является фольклором. Доказательство можно найти в [25].

Теорема 1.1 ([25]). *Если пространство V компактно, и орбита каждой точки плотна в V, то последовательность x почти периодическая.*

Более того, несложно показать, что каждая почти периодическая последовательность может быть получена таким образом, как описано в теореме 1.1. Для этого надо взять пространство всех последовательностей с топологией, порождённой метрикой d_C, для каждой буквы a в качестве области A_a взять множество последовательностей, начинающихся с a, и в качестве отображения f взять отображение L левого сдвига, а в качестве V взять замыкание x относительно действия f, после чего замкнуть получившееся множество относительно метрики d_C. Подробнее см., например, [20].

Топологическая динамическая система, удовлетворяющая свойству из условия теоремы 1.1, называется *минимальной*. Почти периодические последовательности также называют минимальными последовательностями. Они обладают следующим свойством минимальности. Обозначим через $\mathrm{Fac}(x)$ множество всех конечных подслов последовательности x. Почти периодические последовательности обладают минимальным множеством подслов среди всех последовательностей, в следующем смысле: для любых бесконечных последовательностей x и y если x почти периодическая и $\mathrm{Fac}(y) \subseteq \mathrm{Fac}(x)$, то $\mathrm{Fac}(y) = \mathrm{Fac}(x)$, и для любой последовательности y существует почти периодическая последовательность x, такая что $\mathrm{Fac}(x) \subseteq \mathrm{Fac}(y)$ (см. предложение 2.1).

Это утверждение — пример комбинаторного свойства почти периодических последовательностей. Комбинаторные результаты появляются уже в [22, 23]. Кроме того, конечные и бесконечные слова и до этого изучались с комбинаторной точки зрения. Зарождением соответствующей области — комбинаторики слов — принято считать работы Туэ [34, 33], в которых он, в частности, изучает свойства последовательности, теперь называемой последовательность Туэ — Морса (см. 2.4), и доказывает её бескубность (подробнее о работах Туэ см. в [6]). Отметим, что Туэ не имел в виду каких-то конкретных применений своих результатов и считал рассматриваемые им вопросы представляющими самостоятельный интерес (см. [2]). Сейчас комбинаторика слов — активная область, и в настоящей работе к ней можно отнести многие результаты.

Впервые рассмотренные в символической динамике, почти периодические последовательности нашли применение и в математической логике.

Рассмотрим следующие вопросы о последовательности x: входит ли в x символ a? Входит ли слово u? Входит ли слово u бесконечно много раз? Можно сформулировать и более сложные вопросы про последовательность. Когда на такие вопросы можно отвечать алгоритмически, получая на вход последовательность?

Различные, в том числе и перечисленные только что, свойства бесконечных последовательностей могут быть выражены в *логической теории первого порядка*. Под такой теорией для последовательности $x \in A^{\mathbb{N}}$ мы будем понимать следующее. Формально, в качестве структуры возьмём $\langle \mathbb{N}, S, <, X \rangle$, где $\mathbb{N}$ — множество натуральных чисел, которое пробегают индивидные переменные, S — двухместный предикат следования, $<$ — двухместный предикат порядка на натуральных числах, X — функциональный символ, интерпретируемый как последовательность $x \colon \mathbb{N} \to A$. В качестве теории берём обычную теорию первого порядка, несущественным образом расширенную атомарными формулами вида $X(n) = a$ для $n \in \mathbb{N}$ и $a \in A$; истинность формул интерпретируем естественным образом. Во всех рассматриваемых нами теориях мы подразумеваем наличие двухместного предиката равенства, интерпретируемого естественным образом.

Ясно, что у такой реализации есть много вариантов, эквивалентных между собой по выразительным способностям. Например, можно было вместо двухместного предиката следования взять в структуру одноместную функцию следования. Можно было и не брать предикат следования вообще, потому что он выразим через неравенство. Точно так же вместо одной функции X можно было взять семейство предикатов X_a по одному для каждого $a \in A$, истинных ровно там, где в x стоит буква a. Можно обойтись и меньшим — логарифмическим по отношению к размеру алфавита последовательности — количеством предикатов. Кроме того, поскольку константа 0 выразима в определённой выше структуре, ясно, что можно было, не меняя множество выразимых формул, добавить в структуру все константы. Поэтому можно переходить от одной конкретной реализации к другой, если понадобится.

Теорию, определённую выше, будем обозначать $\mathrm{T}\langle\mathbb{N}, <, x\rangle$.

Несложно видеть, что простые свойства последовательности x, типа упомянутых в начале раздела, можно выразить в теории $\mathrm{T}\langle\mathbb{N}, <, x\rangle$. Например, формула $\forall p \exists q \exists r\, q > p \wedge S(q, r) \wedge X(q) = 0 \wedge X(r) = 1$ означает свойство вхождения в x бесконечное количество раз слова 01. Тем не менее, некоторые свойства, которые, вероятно, хотелось бы выразить, в теории первого порядка выразить нельзя, например, "последовательность в алфавите $\{a, b\}$, в которой между любыми последовательными вхождениями b (такими, между которыми нет других вхождений b) входит нечётное количество букв a".

Гораздо больше можно выразить в более сильной *монадической теории второго порядка*. В такой теории, кроме обычных переменных по натуральным числам $p, q, \ldots$, разрешены также монадические переменные по множествам (или по одноместным предикатам) $P, Q, \ldots$. Вводятся также соответствующие атомарные формулы вида $P(p)$, $Q(p)$, ..., обозначающие "p принадлежит P", "p принадлежит Q"... Разрешены также кванторы по монадическим переменным. Такую теорию мы будем обозначать $\mathrm{MT}\langle\mathbb{N}, <, x\rangle$.

Теории, аналогичные вышеперечисленным, но не расширенные последовательностью, будем обозначать соответственно $\mathrm{T}\langle\mathbb{N}, <\rangle$, $\mathrm{MT}\langle\mathbb{N}, <\rangle$.

Теория $\mathrm{MT}\langle\mathbb{N}, <, x\rangle$ богаче теории $\mathrm{T}\langle\mathbb{N}, <, x\rangle$. Например, упомянутое выше не выразимое в теории первого порядка свойство в монадической теории можно выразить так: $\forall p \forall q (X(p) \wedge X(q) \wedge p < q \wedge \forall r (p < r \wedge r < q \to \neg X(r)) \to \exists Y (\forall u \forall v (S(u,v) \to (Y(u) \leftrightarrow Y(v))) \wedge Y(p) \wedge Y(q)))$.

Как и в случае с теориями первого порядка, в описанной формализации монадической теории многое можно реализовать по-другому. Например, можно отказаться от неравенства, потому что в монадической теории оно выразимо через следование. Как и в случае теорий первого порядка, здесь можно переходить от одной реализации к другой для удобства.

Для формулы ϕ в любой из вышеописанных теорий будем обозначать через $L(\phi)$ множество всех последовательностей x, для которых эта формула верна (то есть верна, если интерпретировать входящую в неё единственную свободную переменную по функциям X как x).

Естественным вопросом о любой теории является вопрос о её разрешимости. Теория *разрешима*, если существует алгоритм, который по любой замкнутой формуле определяет её истинность. Оказывается, что благодаря применению теории автоматов для достаточно широких классов последовательностей можно получить критерий разрешимости их монадических теорий.

Интересным образом оказывается, что свойства типа почти периодичности полезны и, более того, естественно возникают при решении такого типа задач. Следующее понятие было введено Семёновым в [37]. Назовём последовательность обобщённо почти периодической, если каждое конечное слово либо входит в неё бесконечное количество раз с ограниченными интервалами между соседними вхождениями, либо входит лишь конечное количество раз. Заметим, что естественное предположение о том, что такие последовательности исчерпываются почти периодическими последовательностями с приписанным к ним префиксом, оказывается неверным, см. предложение 2.4. Обобщённо почти периодическая последовательность эффективно обобщённо почти периодическая, если по каждому слову можно определить, входит ли оно в последовательность бесконечное или конечное количество раз, и в первом случае можно также определить оценку сверху на расстояние между соседними вхождениями, а во втором можно также определить оценку на длину префикса, которым ограничиваются вхождения.

Следующий результат теоремы 1.2 о теориях первого порядка был получен в [37].

Теорию $\mathrm{T}'\langle \mathbb{N}, <, x\rangle$ (незначительную модификацию теории $\mathrm{T}\langle \mathbb{N}, <, x\rangle$) определим следующим образом. Пусть P — система предикатов, задающая последовательность x. Термами теории будут выражения вида c, $x+c$, где $c \in \mathbb{Z}$, x — переменная. Атомарными формулами теории будут выражения $\tau < \tau'$, $\tau > \tau'$, $p(\tau)$, где p — предикат из системы P, τ и τ' — термы. Несмотря на то, что значения термов могут не лежать в $\mathbb{N}$, отношения, задаваемые атомарными формулами, всюду определены на $\mathbb{N}$.

Мы говорим, что теория *бескванторная*, если для каждой формулы теории найдётся эквивалентная ей бескванторная.

Теорема 1.2 ([37]). *1) Если теория* $\mathrm{T}'\langle \mathbb{N}, <, x\rangle$ *бескванторная, то* $x \in \mathcal{GAP}$.

2) Пусть $x \in \mathcal{GAP}$. *Тогда теория* $\mathrm{T}'\langle \mathbb{N}, <, x\rangle$ *бескванторная, причём она разрешима, если и только если x эффективно обобщённо почти периодична.*

Вопрос о разрешимости монадических теорий оказывается более сложным. В [8] была доказана разрешимость теории $\mathrm{MT}\langle \mathbb{N}, <\rangle$, при помощи сопоставления формул теории конечным автоматам на бесконечных последовательностях. Конечный автомат Бюхи состоит из конечного множества состояний, некоторые из которых называются принимающими, а какое-то одно — начальным, а также правила перехода, по которому из любого состояния, получив на вход букву алфавита, определяется следующее состояние. Если правило по букве и текущему состоянию определяет следующее состояние однозначно, автомат называется детерминированным, в противном случае — недетерминированным. По умолчанию мы считаем автомат Бюхи недетерминированным. Рассмотрим автомат Бюхи, которому на вход подаётся бесконечная последовательность. Автомат изначально находится в своём начальном состоянии и далее, читая буквы последовательности по одной, меняет состояние на одно из разрешённых в соответствии с правилом перехода. Получившаяся последовательность состояний называется ходом вычисления автомата на последовательности. Поскольку автомат недетерминированный, ходов может быть бесконечно много. Мы говорим, что автомат

Бюхи принимает последовательность, если, будучи запущенным на последовательности, он может побывать в одном из своих заключительных состояний бесконечно много раз, то есть существует ход, в котором одно из заключительных состояний встречается бесконечно много раз. Для автомата Бюхи M множество всех последовательностей, которые принимаются автоматом M, обозначим $L(M)$.

Теорема 1.3 ([8]). *Существует алгоритм, который по каждому автомату Бюхи M строит формулу ϕ монадического языка, такую что $L(M) = L(\phi)$, и наоборот, по любой формуле ϕ строит такой автомат Бюхи M, что $L(\phi) = L(M)$.*

Следствие 1.4 ([8]). *Теория* $\mathrm{MT}\langle\mathbb{N},<\rangle$ *разрешима.*

После этого возникает естественный вопрос о разрешимости теорий вида $\mathrm{MT}\langle\mathbb{N},<,x\rangle$, то есть $\mathrm{MT}\langle\mathbb{N},<\rangle$, расширенной некоторой последовательностью x. Несложно видеть, что выполняется следующее важное для нас следствие из теоремы 1.3.

Следствие 1.5 ([8]). *Монадическая теория последовательности x разрешима тогда и только тогда, когда существует алгоритм, который по любому автомату Бюхи может определить, принимает ли этот автомат последовательность x или нет.*

Конечно, теория $\mathrm{MT}\langle\mathbb{N},<,x\rangle$ разрешима, если последовательность x выразима в исходной теории, но запас таких последовательностей невелик — это периодические последовательности, возможно, с предпериодом. Оказывается, что можно получить критерий разрешимости теории $\mathrm{MT}\langle\mathbb{N},<,x\rangle$, если ограничиться рассмотрением обобщённо почти периодических последовательностей.

Пусть последовательность x обобщённо почти периодическая. Регулятором почти периодичности последовательности x называется такая функция $\mathrm{R}_x\colon \mathbb{N} \to \mathbb{N}$, которая на числе n равна минимальному такому l, что всякое слово длины n, входящее в x, может входить только в начальный отрезок x длины l и не может входить далее, а также каждое слово длины n, входящее в x бесконечно много раз, входит в каждый отрезок длины l. Назовём последовательность эффективно обобщённо почти периодической, если она является вычислимой обобщённо почти периодической последовательностью, и некоторая оценка сверху на регулятор почти периодичности такой последовательности вычислима. Другими словами, вычислимая обобщённо почти периодическая последовательность x эффективно обобщённо почти периодична, если можно эффективно по n оценивать префикс, которым ограничиваются все вхождения слов длины n, входящих в x конечное количество раз, и эффективно оценивать максимальное расстояние между вхождениями слов длины n, входящих бесконечное количество раз. Следующий критерий был получен в [38].

Теорема 1.6 ([38]). *Теория* $\mathrm{MT}\langle\mathbb{N}, <, x\rangle$ *обобщённо почти периодической последовательности* x *разрешима тогда и только тогда, когда* x *эффективно обобщённо почти периодична.*

Для доказательства этой теоремы необходимо разобраться в том, как ведёт себя конечный автомат, запущенный на обобщённо почти периодической последовательности. Этот вопрос связан со следующей серией вопросов — понять, как меняются или насколько сохраняются различные свойства последовательностей при применении к ним различного вида преобразований. Например, несложно видеть, что если стереть каждый второй символ почти периодической последовательности, то полученная последовательность будет также почти периодической. Другой пример — если закодировать каждый блок длины 10 почти периодической последовательности символом некоторого нового алфавита, причём использовать разные символы для разных блоков и одинаковые символы для одинаковых блоков, то полученная последовательность в новом алфавите будет снова почти периодической. Аналогичные свойства выполнены и для обобщённо почти периодических последовательностей. Рассмотрим менее тривиальный пример: если рассмотреть покомпонентное произведение почти периодической последовательности с периодической последовательностью, то можно доказать, что результат будет снова почти периодическим. Противоположный пример — рассмотрим операцию приписывания к последовательности символа 0. Ясно, что обобщённо почти периодические последовательности сохраняются под действием такого преобразования, но почти периодические нет — 1111... почти периодична, но 01111... нет.

Можно рассматривать и гораздо более сложные преобразования, задаваемые алгоритмами разной степени сложности устройства. Простейшие алгоритмические преобразования — конечно-автоматные, объединяющие в себе вышеперечисленные и многие другие простейшие преобразования над последовательностями. Конечно-автоматный преобразователь состоит из конечного множества состояний, одно из которых выделено и называется начальным, и правила перехода, в соответствии с которым, получив на вход символ некоторого алфавита A, преобразователь определяет следующее состояние, а также выходное слово в некотором алфавите B (это слово может иметь произвольную длину, в том числе и нулевую). Конечно-автоматный преобразователь, будучи запущенным на бесконечной последовательности, изначально находится в начальном состоянии, и получая по одному символы входной последовательности, в соответствии с правилом перехода меняет состояние и печатает слово на выходной ленте. Напечатанное в итоге на ленте бесконечное слово есть результат конечно-автоматного преобразования исходной последовательности. Основным средством доказательства теоремы 1.6 в [38] стал следующий результат, по существу доказанный в [38]. Доказательство в явном виде можно найти также в [25] или в [43]. См. также теорему 3.1.

Теорема 1.7 ([38]). *1) Множество обобщённо почти периодических последовательностей сохраняется под действием конечно-автоматных преобразований.*

2) Множество эффективно обобщённо почти периодических последовательностей сохраняется под действием конечно-автоматных преобразований.

Из теоремы 1.7 следует теорема 1.6. Однако результат теоремы 1.7 представляет и самостоятельный интерес. Известны и другие результаты о сохранении классов последовательностей под действием конечно-автоматных преобразований. В частности, в [9] доказано, что класс автоматных последовательностей, определяемый ниже, сохраняется под действием равномерных конечно-автоматных преобразований (см. также [4]). Под равномерными мы понимаем такие конечно-автоматные преобразования, которые, получив на вход букву, всегда выдают на выход ровно одну букву. Как доказано в [10], морфические последовательности (см. ниже) сохраняются под действием конечно-автоматных преобразований (см. также [4]).

Таким образом, представляется естественным более детально изучить свойства замкнутости классов последовательностей с различными свойствами типа почти периодичности относительно конечно-автоматных преобразований.

Другие важные для нас классы последовательностей, обладающих свойствами, близкими к свойствам периодических последовательностей, являются автоматные и морфические последовательности.

Любая периодическая последовательность (возможно, с предпериодом), например, $3142857142857142\ldots$ (цифры десятичной записи числа $22/7$) может быть порождена машиной с конечной памятью. Достаточно иметь информацию о предпериоде и периоде последовательности: в нашем случае 3 и 142857. И наоборот, любая машина с конечной памятью, печатающая символы конечного алфавита, порождает заключительно периодическую последовательность. Действительно, во время её работы в какой-то момент конфигурация машины полностью совпадёт с какой-то из уже встречавшихся, и так начнётся период в выходной последовательности.

Если разрешить машине обращаться к символам, написанным ранее, класс порождаемых последовательностей существенно возрастёт. При некоторых ограничениях на порядок, в котором ранее написанные символы читаются снова, мы получим класс автоматных последовательностей, введённый Кобхэмом в [9].

Например, последовательность Туэ — Морса (см. раздел 2.4) $\mathbf{t} = 0110100110010110\ldots$ удовлетворяет такому условию: если на n-м месте стоит 0, то на $2n$-м и $(2n+1)$-м будет 0 и 1 соответственно, а если на n-м месте 1, то, наоборот, на $2n$-м и $(2n+1)$-м будет 1 и 0 соответственно.

Кобхэм вводит иерархию на автоматных последовательностях, в зависимости от того, как далеко надо заглядывать назад, чтобы написать очередной символ. В k-автоматных последовательностях n-й символ определяет, что стоит на местах $kn, kn+1, kn+2, \ldots, (k+1)n-1$.

Формально определим автоматные последовательности двумя эквивалентными способами.

Рассмотрим конечный автомат, действующий на словах в алфавите $\Sigma_k = \{0, 1, \ldots, k-1\}$. Каждому состоянию автомата соответствует буква в некотором другом алфавите A (разным состояниям могут соответствовать одинаковые буквы). Автомат действует так: получает на вход слово в алфавите Σ_k, производит вычисления и выдаёт ту букву алфавита A, которая соответствует последнему состоянию в вычислении. Последовательность x букв алфавита A называется k-автоматной, если существует конечный автомат вышеуказанного вида, который, будучи запущенным на числе n, записанном в k-ичной системе счисления, выдаёт букву $x(n)$. Когда ясно или неважно, о каком k идёт речь, приставку “k-” мы будем опускать.

Теперь опишем другой подход. Пусть A, B — конечные алфавиты. Рассмотрим операцию подстановки, которая каждую букву алфавита A отображает в слово алфавита B. Результатом применения подстановки к конечному слову или бесконечной последовательности в алфавите A будет конкатенация образов букв, составляющих это слово или последовательность. Такие отображения подстановки будем называть морфизмами. Морфизм называется k-равномерным, если длины образов всех букв совпадают и равны k.

Ограничимся теперь рассмотрением ситуации, когда алфавиты A и B совпадают. Нас интересуют бесконечные последовательности, являющиеся неподвижными точками морфизмов, причём неподвижными точками достаточно общего вида. Пусть ϕ — морфизм в алфавите A, и s — буква алфавита A, такая что слово $\phi(s)$ начинается с s. Будем итерировать ϕ на s, получим слова s, $\phi(s)$, $\phi^2(s)$, $\phi^3(s)$, $\phi^4(s)$, $\ldots$, каждое из которых начинается с предыдущего. Если длина слова $\phi^n(s)$ стремится к бесконечности с ростом n, можно корректно определить бесконечную последовательность $\phi^\infty(s)$, началами которой являются слова $\phi^n(s)$ для любого n. Последовательности $x = \phi^\infty(s)$, которые можно получить таким способом, называются чисто морфическими. Несложно видеть, что чисто морфическая последовательность, порождённая морфизмом ϕ, является неподвижной точкой под действием этого морфизма. Последовательности, получающиеся из чисто морфических отождествлением некоторых символов, называются морфическими. Как доказал Кобхэм в [9], k-автоматные последовательности — это в точности морфические последовательности, полученные из порождённых k-равномерными морфизмами чисто морфических последовательностей.

Морфические последовательности, точнее, тесно связанные с ними

$D0L$-последовательности и, более общо, L-системы (системы Линденмайера), впервые появились в работах математика и биолога Линденмайера для описания процессов развития живых организмов (например, см. сборник [30]). Однако впоследствии морфические последовательности стали рассматриваться и в теории динамических систем, комбинаторике слов, информатике (см. [4, 16, 29, 30]). Автоматные последовательности неявно были впервые рассмотрены в [7], впервые систематически изучались в [9]. Хорошей монографией, посвящённой автоматным и морфическим последовательностям, является [4].

Множества морфических и почти периодических последовательностей находятся в общем положении. Действительно, существуют последовательности, являющиеся морфическими, но не почти периодическими: например, последовательность в алфавите $\{0, 1\}$, у которой каждый символ с номером 2^n равен 1, а все остальные равны 0 (мы нумеруем последовательности, начиная с 0), 2-автоматна (легко построить автомат, который будет распознавать числа вида $1000\dots00$ в двоичной системе счисления), но не является периодической. Существуют почти периодические, но не морфические последовательности; более того, почти периодических последовательностей континуум (как доказано, например, в [17]), а морфических лишь счётное число. Существуют также и одновременно морфические и почти периодические последовательности. Конечно, все периодические последовательности являются таковыми. Среди непериодических такова, например, последовательность Туэ — Морса, упомянутая выше (см. также раздел 2.4). Заметим также, что морфические последовательности можно конечно описать — для этого нужно описать алфавит, морфизм, букву, на которой итерируется морфизм, и последующий способ отождествления букв в чисто морфической последовательности. Поэтому можно ставить алгоритмические вопросы о морфических последовательностях. Довольно естественным является вопрос о существовании алгоритма для следующей проблемы.

Проблема распознавания почти периодичности морфических последовательностей (ППМ).

Дано: конечное описание морфической последовательности.

Определить: является ли эта последовательность почти периодической.

Многие алгоритмические задачи, связанные с морфизмами, являются довольно сложными и интересными. Хорошим обзором по таким задачам является [16]. Один из самых известных примеров такой задачи — проблема соответствий Поста (Post correspondence problem), которая неразрешима (см. [32], а также [16]). Про проблему ППМ до сих пор неизвестно, является ли она разрешимой, хотя гипотеза о разрешимости является довольно правдоподобной. В настоящей работе мы исследуем некоторые частные случаи проблемы ППМ, возникающие при ограничении множества входов на морфические последовательности специального вида.

Тесто связанный с проблемой ППМ вопрос или даже вариант формулировки — получение по возможности как можно более компактного и эффективно проверяемого критерия почти периодичности для морфических последовательностей. Вопрос о получении такого критерия для чисто морфических последовательностей был поставлен как открытый в [4], раздел 10.12, проблема 5.

Следующая тема диссертации — подсловная сложность, одна из простейших и наиболее естественных характеристик бесконечных последовательностей. Подсловной сложностью последовательности x называется такая функция $\mathrm{p}\colon \mathbb{N} \to \mathbb{N}$, что $\mathrm{p}(n)$ равно количеству слов длины n, входящих в последовательность x. Для последовательностей в алфавите из m символов подсловная сложность может варьироваться от 1 до m^n. Как доказано в [22] (также см. предложение 2.5), подсловная сложность последовательности ограничена тогда и только тогда, когда последовательность является заключительно периодической.

Асимптотическое поведение подсловной сложности последовательностей широко изучалось, например, см. обзоры [5, 14]. В серии работ, завершающейся работой [26], доказано, что подсловная сложность чисто морфической последовательности может удовлетворять одной из следующих пяти асимптотик: $O(1)$, $\Theta(n)$, $\Theta(n \log\log n)$, $\Theta(n \log n)$, $\Theta(n^2)$. В [27] показано, что существуют примеры морфических последовательностей с подсловной сложностью вида $\Theta(n^{1+\frac{1}{k}})$ для каждого натурального $k \geqslant 1$. После этого про подсловную сложность морфических последовательностей произвольного вида долгое время было ничего неизвестно. Наконец, в [11] было доказано, что подсловная сложность морфической последовательности имеет один из следующих видов: $O(n \log n)$ или $\Theta(n^{1+\frac{1}{k}})$ для некоторого $k \in \mathbb{N}$. Таким образом, для полного описания возможных асимптотик подсловной сложности морфической последовательности осталось разобраться подробнее со случаем $O(n \log n)$.

В случае автоматных последовательностей ситуация существенно проще. Как доказано в [9], подсловная сложность автоматной последовательности не более чем линейна.

Несложно показать (например, см. [14]), что подсловная сложность почти периодической последовательности в алфавите из m символов не может быть больше $m^{\alpha n}$ для некоторого фиксированного $\alpha < 1$. При этом можно показать, что для любого $\beta < 1$ существует почти периодическая последовательность в алфавите из m символов с подсловной сложностью не меньше $m^{\beta n}$ (например, см. [46]).

Оказывается, подсловная сложность последовательности, являющейся одновременно морфической и почти периодической, устроена гораздо проще. Как мы доказываем в настоящей работе, подсловная сложность таких последовательностей не более чем линейна. Отметим, что линейность подсловной сложности почти периодических чисто морфических последовательностей была известна ранее (например, см. [29]).

Наконец, заключительной темой настоящей работы является изучение свойств понятия меры апериодичности бесконечной последовательности.

Периодические последовательности имеют самую простую структуру среди всех последовательностей над конечным алфавитом, поэтому естественно было бы научиться измерять, насколько данная последовательность далека от любой периодической. В [47] вводится понятие меры апериодичности бесконечной последовательности. Неформально, мера апериодичности последовательности — это максимальное такое число α, что последовательность с любым своим нетривиальным сдвигом имеет хотя бы долю α различий. Насколько известно автору, систематического исследования этого понятия и его свойств ранее не проводилось, однако известны некоторые отдельные результаты. В [47] получены некоторые простейшие свойства и посчитана мера апериодичности некоторых последовательностей, предлагается список открытых вопросов для дальнейшего исследования.

1.2. Основные результаты

В главе 2 вводятся необходимые определения и обсуждаются их основные свойства. Результаты работы излагаются в главах 3, 4, 5, 6.

Глава 3 посвящена конечно-автоматным преобразованиям. Из теоремы 1.7 следует, что образ почти периодической последовательности под действием конечно-автоматного преобразования является обобщённо почти периодической последовательностью. Оказывается, верно более сильное утверждение. Назовём последовательность заключительно почти периодической, если она является конкатенацией конечного слова и почти периодической последовательности. Как уже отмечалось ранее, множество заключительно почти периодических последовательностей является собственным подмножеством множества обобщённо почти периодических последовательностей (см. предложение 2.4).

Теорема 3.4. *Почти периодические последовательности под действием конечно-автоматных преобразований переходят в заключительно почти периодические.*

Часть 2 теоремы 1.7 является эффективизацией части 1. По существу, в дополнение к результату части 1 в ней требуется получить оценку на регулятор почти периодичности образа последовательности под действием конечно-автоматного преобразователя, зная преобразователь, последовательность и оценку на её регулятор почти периодичности. Именно эта эффективная версия и позволяет доказать теорему 1.6.

По аналогии рассмотрим такой вопрос об эффективизации теоремы 3.4. Допустим, нам известны конечно-автоматный преобразователь, почти периодическая последовательность и оценка на её регулятор почти периодичности. В соответствии с теоремой 3.4 от образа этой последовательности под действием этого преобразователя можно отрезать некоторый начальный отрезок, так что получится почти периодическая последовательность. Можно ли оценить сверху длину этого отрезка? Ан. А. Мучник в 2005 г. высказал гипотезу о том, что это невозможно сделать эффективно, однако результат следующей теоремы 3.5 показывает, что такая оценка существует. Обозначим через R_x регулятор почти периодичности последовательности x. Для произвольной функции

$g\colon \mathbb{N} \to \mathbb{N}$ обозначим $\underbrace{g \circ g \circ \cdots \circ g}_{m}$ через g^m.

Теорема 3.5. *Пусть M — равномерный конечно-автоматный преобразователь с m состояниями, и x — почти периодическая последовательность. Тогда последовательность $M(x)$ — образ x под действием M — становится почти периодической последовательностью при удалении префикса длины не более* $\mathrm{R}_x^m(1) + \mathrm{R}_x^{m-1}(1) + \cdots + \mathrm{R}_x(1)$.

Далее мы получаем свойства замкнутости для рекуррентных последовательностей относительно конечно-автоматных преобразований, а также для обобщённо почти периодических последовательностей относительно почти обратимых конечно-автоматных преобразований. Мы доказываем, однако, что множество обобщённо почти периодических последовательностей не замкнуто относительно более широкого класса стековых конечно-автоматных преобразований.

Глава 4 посвящена вопросам вычислимости на бесконечных последовательностях. В связи с теоремой 3.5 возникает следующий более общий вопрос: пусть нам дана заключительно почти периодическая последовательность и оценка сверху на её регулятор почти периодичности. Можно ли только на основании этих данных получить оценку на тот префикс, который достаточно отрезать от этой последовательности, чтобы получить почти периодическую последовательность? Как доказывается в теореме 4.1, такой оценки не существует. Теоремы 4.2, 4.3, 4.8 являются дальнейшими примерами результатов, в которых доказывается невозможность по обобщённо почти периодической, заключительно почти периодической или почти периодической последовательности и её регулятору почти периодичности найти некоторую характеристику или проверить некоторое свойство. Наиболее общая из них — теорема 4.3, утверждающая, что для любого счётного множества M последовательностей, содержащего все периодические последовательности, по обобщённо почти периодической последовательности и её регулятору невозможно определить, принадлежит ли эта последовательность множеству M. В качестве следствий из этих результатов доказывается, что некоторые множества последовательностей не являются регулярными, см. следствия 4.10 и 4.11.

В главе 5 исследуются свойства морфических последовательностей. Основной обсуждаемый вопрос — разрешимость проблемы ППМ (см. раздел 1.1). Вопрос по-прежнему остаётся открытым, но мы получаем следующие два более слабых результата об эффективной разрешимости почти периодичности для морфических последовательностей специального вида.

Теорема 5.6. *Существует полиномиальный по времени алгоритм, определяющий по чисто морфической последовательности, является ли она почти периодической.*

Теорема 5.7. *Существует полиномиальный по времени алгоритм, определяющий по автоматной последовательности, является ли она почти периодической.*

Перед доказательством этих результатов мы получаем критерии почти периодичности для чисто морфических и автоматных последовательностей. Отметим, что результат теоремы 5.6, а также соответствующий критерий почти периодичности были изначально получены автором в [45] для чисто морфических последовательностей, порождённых нестирающими морфизмами, то есть такими, что образ каждой буквы является непустым словом. Впоследствии в работе [48] этот результат был обобщён первым автором на произвольные чисто морфические последовательности.

В заключительном разделе 5.4 главы 5 мы получаем следующий результат.

Теорема 5.19. *Подсловная сложность почти периодических морфических последовательностей не более чем линейна.*

Глава 6 посвящена мере апериодичности бесконечных последовательностей. Мы приводим результаты, опубликованные в [47], которые получены автором диссертации. А именно, мы получаем общие верхнюю и нижнюю оценки для меры апериодичности последовательностей для алфавита фиксированного размера. Мы доказываем, что мера апериодичности последовательностей Штурма (ещё одно простейшее обобщение периодических последовательностей — см. определение ниже) равна 0. После этого мы доказываем, что мера апериодичности последовательности Туэ — Морса равна 1/3, а также получаем оценку на меру апериодичности естественного обобщения последовательности Туэ — Морса — последовательностей Пруэ.

Благодарности

Автор выражает глубокую благодарность своему научному руководителю чл.-корр. РАН, проф. Алексею Львовичу Семёнову за постановку задач, поддержку и постоянное внимание к работе, а также к. ф.-м. н. Андрею Альбертовичу Мучнику (1958 — 2007) за постановку задач, многочисленные обсуждения результатов работы. Преждевременный уход от нас Андрея Альбертовича — огромная потеря для российской и мировой математики.

Автор благодарен своим соавторам Франсуа Николя (François Nicolas), Михаилу Александровичу Раскину, Юлии Сергеевне Уляшкиной за успешное сотрудничество, а также Сергею Владимировичу Августиновичу, Сергею Ивановичу Адяну, Николаю Константиновичу Верещагину, Ростиславу Андреевичу Девятову, Брюно Дюрану (Bruno Durand), Юхани Кархумяки (Juhani Karhumäki), Жюльену Кассэню (Julien Cassaigne), Орне Купферман (Orna Kupferman), Нараду Рамперсаду (Narad Rampersad), Андрею Юрьевичу Румянцеву, Калле Саари (Kalle Saari), Александру Шеню и другим за многочисленные полезные обсуждения на темы, затронутые в работе.

Результаты работы неоднократно обсуждались на Колмогоровском семинаре кафедры математической логики и теории алгоритмов мехмата МГУ, а также на семинаре “Алгоритмические вопросы алгебры и логики” под руководством академика РАН С. И. Адяна, автор благодарен всем руководителям и участникам этих семинаров за внимание и поддержку. Автор благодарен всем сотрудникам кафедры математической логики и теории алгоритмов за прекрасную доброжелательную атмосферу.

Глава 2.

Обзор основных понятий

2.1. Основные определения

Алфавитом мы называем произвольное конечное множество. Обычно алфавит мы обозначаем буквами A, B и т. п. Его элементы мы называем *символами* или *буквами*. Стандартный бинарный алфавит из двух символов $\{0, 1\}$ обозначим $\mathbb{B}$. Алфавит $\{0, 1, \dots, n-1\}$ обозначим Σ_n. По умолчанию мы рассматриваем только конечные алфавиты, и при рассмотрении бесконечных алфавитов оговариваем это особо. Для конечного множества X через $\#X$ или $|X|$ обозначим количество элементов в нём. Множество натуральных чисел $\{0, 1, 2, \dots\}$ обозначим через $\mathbb{N}$.

Если $i \leqslant j$ натуральные, через $[i, j]$ обозначим отрезок натурального ряда с концами в i и j, то есть множество $\{i, i+1, i+2, \dots, j\}$. Отрезок $[0, n]$ будем обозначать просто $[n]$. Конечная последовательность букв в алфавите A называется *словом* над или в A. Через A^* обозначим множество всех конечных слов над A, включая пустое слово Λ. Мы отождествляем однобуквенные слова и буквы алфавита. Через $|u|$ будем обозначать длину слова u, через $|u|_a$ — количество букв a в слове u. Также под словом мы будем понимать отображение $u\colon [|u|-1] \to A$, так что $u(i)$ для $0 \leqslant i \leqslant |u|-1$ обозначает i-й по порядку символ слова u. Для слова $u \in \mathbb{B}^*$ обозначим через $\bar{u}$ слово, полученное из u заменой всех 0 на 1 и всех 1 на 0.

Будем рассматривать *последовательности* над A — отображения $x\colon \mathbb{N} \to A$. Определим *канторову метрику* d_C на последовательностях как $d_C(x, y) = 2^{-n}$, где $n = \min\{k : x(k) \neq y(k)\}$. Множество всех последовательностей, наделённое метрикой d_C, образует стандартное канторовское компактное метрическое пространство, которое мы обозначаем $A^{\mathbb{N}}$.

Таким образом, $\lim_{n\to\infty} u_n = x$, если для каждого i найдётся такое n, что для всех $m > n$ имеем $u_m(i) = x(i)$ (это определение годится и для конечных слов u_n). Последовательности называют также *бесконечными словами*.

Другой способ определить расстояние между бесконечными последовательностями — *расстояние Безиковича*: $d_B(x, y) = \liminf \frac{1}{n}\#\{i : 0 \leqslant i \leqslant n-1, x(i) \neq y(i)\}$. Такая функция уже не будет метрикой, потому что возможно $d_B(x, y) = 0$ при $x \neq y$. Однако двум другим свойствам метрики — симметричность и неравенство треугольника — расстояние Безиковича удовлетворяет.

Через $x[i, j]$ обозначим отрезок последовательности x — слово $x(i)x(i+1)\dots x(j)$. Говорят, что $[i, j]$ — *вхождение в последовательность x слова $u \in A^*$*, если $x[i, j] = u$. Вхождение $u = x[i, j]$ в x называется *k-выровненным*, если k делит i. Слово $u \neq \Lambda$ называется *фактором* или *подсловом* последовательности x, если u входит в x. Множество всевозможных факторов последовательности x будем обозначать $\mathrm{Fac}(x)$, множество всевозможных факторов длины m — $\mathrm{Fac}_m(x)$. Факторы вида $x[0, i]$ называются *префиксами* x, последовательности вида $x(i)x(i+1)x(i+2)\dots$ — *суффиксами* x и обозначаются $x[i, \infty)$. Мы представляем себе последовательность расположенной горизонтально и идущей слева направо до бесконечности, поэтому будем использовать выражения “левее” и “правее” для меньших и бóльших индексов соответственно. Операция *левого сдвига* L заключается в удалении из последовательности её начального символа: $L(a_0a_1a_2a_3\dots) = a_1a_2a_3\dots$

Подсловной сложностью последовательности x называется такая функция $\mathrm{p}_x\colon \mathbb{N} \to \mathbb{N}$, что $\mathrm{p}_x(n)$ равно количеству слов длины n, входящих в последовательность x.

2.2. Морфизмы

Пусть A, B — конечные алфавиты. Отображение $\phi\colon A^* \to B^*$ называется *морфизмом*, если для любых $u, v \in A^*$ выполнено $\phi(uv) = \phi(u)\phi(v)$. Ясно, что морфизм полностью определяется своими значениями на однобуквенных словах. Морфизм *нестирающий*, если $|\phi(a)| \geqslant 1$ для всех $a \in A$. Морфизм называется *k-равномерным*, если $|\phi(a)| = k$ для всех $a \in A$. 1-равномерный морфизм называется *кодированием*. Естественным образом морфизмы продолжаются на бесконечные последовательности: если x — последовательность букв алфавита A, по определению положим $\phi(x) = \phi(x(0))\phi(x(1))\phi(x(2))\ldots$

Пусть $\phi\colon A^* \to A^*$ — некоторый морфизм. Назовём его *неприводимым*, если для любых $a, b \in A$ существует n, такое что $\phi^n(a)$ содержит b. Назовём его *примитивным*, если существует такое n, что для всех $a, b \in A$ слово $\phi^n(a)$ содержит b. Каждый примитивный морфизм является неприводимым, но обратное, вообще говоря, неверно, простейший контрпример — морфизм ϕ, для которого $\phi(0) = 1$ и $\phi(1) = 0$.

Слово $w \in A^*$ называется *ϕ-ограниченным*, если последовательность конечных слов $(w, \phi(w), \phi^2(w), \phi^3(w), \ldots)$ периодична с некоторого места. Слово $w \in A^*$ называется *ϕ-возрастающим*, если $|\phi^n(w)| \to \infty$ при $n \to \infty$. Несложно видеть, что каждое слово из A^* либо ϕ-ограниченное, либо ϕ-возрастающее. Морфизм ϕ называется *возрастающим*, если $|\phi(a)| \geq 2$ для любого $a \in A$. Эквивалентно, морфизм $\phi\colon A^* \to A^*$ возрастающий, если в алфавите A все буквы ϕ-возрастающие.

Слово $w \in A^*$ называется *ϕ-стираемым*, если $\phi^n(w) = \Lambda$ для некоторого n. Слово ϕ-стираемо тогда и только тогда, когда оно состоит только из ϕ-стираемых символов. Пусть I_ϕ — множество всех ϕ-возрастающих символов, B_ϕ — множество всех ϕ-ограниченных символов. Определим также $E_\phi \subseteq B_\phi$ — множество всех ϕ-стираемых символов. Заметим, что $A = I_\phi \cup B_\phi$ и $I_\phi \cap B_\phi = \varnothing$.

Пусть $|A| = n$. Отождествим тогда A с Σ_n, то есть будем считать, что морфизм имеет вид $\phi\colon \Sigma_n^* \to \Sigma_n^*$. Определим *матрицу инцидентности* M_ϕ морфизма, так что $(M_\phi)_{ij}$ равно количеству вхождений символа i в слово $\phi(j)$. Несложно видеть, что для любого $l \in \mathbb{N}$ имеем $M_\phi^l = M_{\phi^l}$. Видно, что ϕ примитивный тогда и только тогда, когда для некоторого l все элементы матрицы M_ϕ^l положительные.

Пусть $\phi\colon A^* \to A^*$ — морфизм. Определим ориентированный *граф инцидентности* G_ϕ морфизма ϕ. Его множество вершин — алфавит A. В графе G_ϕ рёбра идут из любой буквы $b \in A$ во все буквы, входящие в слово $\phi(b)$. Многие свойства морфизма определяются по его графу инцидентности. Однако граф G_ϕ не содержит информации о количестве вхождений i в $\phi(j)$, то есть G_ϕ содержит меньше информации, чем M_ϕ.

2.3. Основные классы последовательностей

Последовательность x *периодическая*, если для некоторого положительного $T \in \mathbb{N}$, называемого *периодом*, имеем $x(i) = x(i+T)$ для любого $i \in \mathbb{N}$. В соответствии с обычным словоупотреблением, периодом мы называем также и слово $x[0, T-1]$. Будем называть последовательность *заключительно периодической*, если то же выполнено для всех i, начиная с некоторого K. Тогда $x[0, K-1]$ называется *предпериодом*. Предпериодом будем называть также и число K. Последовательность, не являющаяся заключительно периодической, называется *апериодической*. Множество всех периодических последовательностей обозначим $\mathcal{P}$, множество заключительно периодических обозначим $\mathcal{EP}$. Рассмотрим некоторые расширения этих классов.

Последовательность x называется *почти периодической*, если для каждого её фактора u найдётся такое натуральное l, что на каждом отрезке длины l последовательности x найдётся вхождение слова u. Тем самым, любое слово, входящее в почти периодическую последовательность, входит в неё бесконечное количество раз. Через $\mathcal{AP}$ (almost periodic) будем обозначать класс всех таких последовательностей. Ясно, что для проверки почти периодичности достаточно только убедиться в повторяемости с ограниченными интервалами всех префиксов, а не всех факторов. Почти периодические последовательности называют также *равномерно рекуррентными* и *минимальными*. Одним из вариантов того, как можно сформулировать свойство минимальности почти периодических последовательностей, является следующее утверждение.

Предложение 2.1. *1) Для любых бесконечных последовательностей x и y если x почти периодическая и $\mathrm{Fac}(y) \subseteq \mathrm{Fac}(x)$, то $\mathrm{Fac}(y) = \mathrm{Fac}(x)$.*

2) Для любой последовательности y существует почти периодическая последовательность x, такая что $\mathrm{Fac}(x) \subseteq \mathrm{Fac}(y)$.

Доказательство предложения 2.1 можно найти, например, в [19].

Будем называть последовательность x *заключительно почти периодической*, если некоторый её суффикс почти периодичен. Класс всех таких последовательностей обозначим $\mathcal{EAP}$ (eventually almost periodic).

Последовательность x называется *обобщённо почти периодической*, если для каждого её фактора u, входящего в неё бесконечное число раз, найдётся такое натуральное l, что на каждом отрезке длины l последовательности x найдётся вхождение слова u. Класс всех таких последовательностей обозначим через $\mathcal{GAP}$ (generalized almost periodic).

Если $x \in \mathcal{EAP}$, то минимальное такое n, что $x[n, \infty) \in \mathcal{AP}$, будем называть *минимальным префиксом* и обозначать $\mathrm{pr}(x)$. Заметим, что для любого $m \geqslant \mathrm{pr}(x)$ имеем $x[m, \infty) \in \mathcal{AP}$.

Назовём последовательность x *рекуррентной*, если каждое слово, которое в неё входит, обязательно входит бесконечное количество раз. Ясно, что если последовательность рекуррентная и обобщённо почти периодическая, то она почти периодическая. Класс рекуррентных последовательностей обозначим $\mathcal{R}$. Последовательность *заключительно рекуррентная*, если некоторый её суффикс рекуррентен. Класс таких последовательностей обозначим $\mathcal{ER}$.

Регулятором почти периодичности последовательности $x \in \mathcal{GAP}$ назовём функцию $\mathrm{R}_x\colon \mathbb{N} \to \mathbb{N}$, которая на числе n равна минимальному такому l, что каждое слово длины n, которое входит в x бесконечное количество раз, встретится на любом отрезке длины l последовательности x, а также любое слово длины n, которое входит в x конечное количество раз, не входит в $x[l, \infty)$ (второе важно только для обобщённо почти периодических последовательностей, не являющихся почти периодическими). Часто вместо регулятора нам будет достаточно рассматривать только какую-то верхнюю оценку на него, то есть функцию $f\colon \mathbb{N} \to \mathbb{N}$, такую что $f(n) \geqslant \mathrm{R}_x(n)$ для всех n — в этом случае мы пишем $f \geqslant \mathrm{R}_x$. Если говорить более точно, регулятор почти периодичности объединяет в себе две функции: одна следит за расстояниями между вхождениями слов, входящих бесконечное количество раз, а другая следит за тем префиксом, которым ограничивается вхождение слов, входящих конечное количество раз. Иногда стоит эти функции разделять явно, но в данной работе мы этого делать не будем.

Будем говорить, что последовательность x *эффективно обобщённо почти периодическая*, если x вычислима, обобщённо почти периодична, и некоторая оценка сверху на регулятор почти периодичности последовательности x также вычислима. Мы получим эквивалентное определение, если заменим требование вычислимости некоторой оценки сверху на регулятор требованием вычислимости регулятора в точности.

Предложение 2.2. *Пусть последовательность $x \in \mathcal{GAP}$ вычислима, и вычислима некоторая функция $f \geqslant \mathrm{R}_x$. Тогда функция R_x вычислима.*

Доказательство. Для каждого n можно найти все слова длины n, входящие в x бесконечное количество раз, и все слова длины n, входящие в x конечное ненулевое количество раз. Слова длины n, входящие в x бесконечно много раз — это слова длины n, входящие в $x[f(n), 2f(n)]$. Из оставшихся слов длины n, те, которые входят в $x[0, f(n)+n]$, входят в x конечное ненулевое количество раз, а остальные вообще не входят в x.

Пусть l_1 равно минимальному такому l, что все слова длины n, входящие в x конечное количество раз, не входят в $x[l, \infty]$. Поскольку $l_1 \leqslant f(n)$, такое l_1 можно найти перебором.

Аналогично предыдущему, можно найти множество K слов длины $f(n)$, хотя бы один раз входящих в x. Теперь достаточно найти минимальное такое l_2, что каждое подслово длины n, встречающееся в x бесконечно много раз, встречается в каждом подслове длины l_2 каждого из слов множества K. Тогда выполнено, что любое подслово x длины n, встречающееся в x бесконечно много раз, входит в любое подслово x длины l_2.

Таким образом, $\mathrm{R}_x(n) = \max(l_1, l_2)$. □

Назовём последовательность *эффективно почти периодической*, если она почти периодична и эффективно обобщённо почти периодична. Определение эффективной почти периодичности можно упростить.

Предложение 2.3. *Почти периодическая последовательность x эффективно почти периодична тогда и только тогда, когда x вычислима и множество подслов $\mathrm{Fac}(x)$ разрешимо.*

Доказательство. $\Rightarrow$. Пусть $x \in \mathcal{AP}$ вычислима, и мы умеем вычислять некоторую оценку $f \geqslant \mathrm{R}_x$. Тогда чтобы найти все слова длины n в x, достаточно взять произвольное слово длины $f(n)$ в x: множество его подслов длины n в точности является множеством подслов длины n всей последовательности x, в соответствии с определением почти периодичности.

$\Leftarrow$. Пусть $x \in \mathcal{AP}$ вычислима и множество $\mathrm{Fac}(x)$ разрешимо. Чтобы найти значение $\mathrm{R}_x(n)$, будем перебирать все натуральные числа по очереди, начиная, скажем, с n. Проверяя число m, мы смотрим, верно ли, что каждое из слов множества $\mathrm{Fac}_m(x)$ (подслова x длины m) содержит все слова из $\mathrm{Fac}_n(x)$ (подслова x длины n). Если это так, то $\mathrm{R}_x(n) \leqslant m$, а иначе $\mathrm{R}_x(n) > m$. Поскольку $x \in \mathcal{AP}$, до бесконечности мы проверять не можем и когда-нибудь найдём подходящее m. $\square$

Несложно видеть, что $\mathcal{P} \subset \mathcal{AP} \subset \mathcal{EAP} \subset \mathcal{GAP}$. Оказывается, все эти включения строгие. Например, известная последовательность Туэ — Морса $\mathbf{t} = 0110100110010110\ldots$ (см. раздел 2.4) — пример последовательности из $\mathcal{AP}$, но не из $\mathcal{P}$ (более того, класс $\mathcal{AP}$ континуален, тогда как $\mathcal{P}$ счётный, доказательство см. в [17] или [25]). Неравенство $\mathcal{AP} \subsetneq \mathcal{EAP}$ очевидно (можно взять $10000\ldots \in \mathcal{EAP} \setminus \mathcal{AP}$). Докажем неравенство $\mathcal{EAP} \subsetneq \mathcal{GAP}$ (доказано в [39]).

Предложение 2.4. *Существует последовательность $x \in \mathcal{GAP} \setminus \mathcal{EAP}$ в алфавите $\mathbb{B}$.*

Доказательство. Построим цепочку двоичных слов: $a_0 = 1$, $a_1 = 10011$, $a_2 = 1001101100011001001110011$, и так далее. Слово a_{n+1} получается из a_n по такому правилу:

$$a_{n+1} = a_n \overline{a}_n \overline{a}_n a_n a_n.$$

Введём обозначение

$$c_n = a_n a_n a_n a_n$$

и рассмотрим последовательность

$$x = c_0 c_1 c_2 c_3 \ldots$$

Докажем, что она и является искомой.

Длина a_n равна 5^n, поэтому длина начального отрезка $c_0 c_1 \ldots c_{n-1}$ последовательности x равна $4(1 + 5 + \cdots + 5^{n-1}) = 5^n - 1$. Для удобства введём обозначение

$$l_n = 5^n - 1 = |c_0 c_1 \ldots c_{n-1}|.$$

Покажем, что x обобщённо почти периодична. Пусть $u \neq \Lambda$ встречается в x бесконечное количество раз. Возьмём такое n, что $|u| < 5^n$. Пусть $[i, j]$ — вхождение слова u в x, такое что $i \geqslant l_n$. Из построения последовательности следует, что для каждого k суффикс $x[l_k, \infty)$ можно рассматривать не только как конкатенацию символов 0 и 1, но и как конкатенацию слов a_k и $\overline{a}_k$. Поэтому в силу выбора i слово u является подсловом одного из четырёх слов $a_n a_n$, $a_n \overline{a}_n$, $\overline{a}_n a_n$, $\overline{a}_n \overline{a}_n$. Заметим, что слово 10011 содержит всевозможные слова длины два 00, 01, 10, 11, поэтому слово a_{n+1} содержит каждое из слов $a_n a_n$, $a_n \overline{a}_n$, $\overline{a}_n a_n$, $\overline{a}_n \overline{a}_n$. Отсюда получаем, что слово u является подсловом слова a_{n+1}. Аналогично, u является подсловом слова $\overline{a}_{n+1}$. На каждом отрезке длины $2|a_{n+1}|$, расположенном правее позиции l_{n+1}, найдётся вхождение слова a_{n+1} или $\overline{a}_{n+1}$, поэтому для $l = (5^{n+1} - 1) + 2 \cdot 5^{n+1}$ выполнено, что на каждом отрезке длины l найдётся вхождение слова u.

Докажем теперь для каждого натурального $n > 0$, что слово c_n не встречается в последовательности x правее позиции l_{n+1}. Тем самым мы покажем, что для суффикса $x[l_n, \infty)$ найдётся слово, а именно, c_n, которое входит в неё ненулевое конечное количество раз, то есть такая последовательность не является почти периодической. Таким образом, будет доказано, что x не является заключительно почти периодической.

Пусть $\nu = x[l_{n+1}, \infty)$. Как уже было замечено, для каждого k, $1 \leqslant k \leqslant n+1$, ν можно представить как конкатенацию слов a_k и $\overline{a}_k$. Предположим, c_n входит в ν, и пусть $[i, j]$ — одно из таких вхождений. При $n > 0$ слово c_n начинается с a_1, поэтому $[i, i+4]$ является вхождением слова a_1 в ν. Заметим, что если $a_1 = 10011$ входит в $a_1 a_1 = 1001110011$, $a_1 \overline{a}_1 = 1001101100$, $\overline{a}_1 a_1 = 0110010011$ или $\overline{a}_1 \overline{a}_1 = 0110001100$, то только с нулевой или с пятой позиции. Поэтому вхождение $[i, j]$ 5-выровнено, то есть можно считать c_n и ν составленными из слов a_1 и $\overline{a}_1$, и c_n — слово из “букв” a_1 и $\bar{a}_1$ — входит в ν. Индукцией по m несложно доказать, что при $1 \leqslant m \leqslant n$ вхождение $[i, j]$ 5^m-выровнено, то есть если представить ν и c_n в виде конкатенации слов a_m и $\overline{a}_m$, то c_n входит в ν. База для $m = 1$ уже доказана. Зная это утверждение для $m = k$, мы можем представить ν и c_n составленными из “букв” a_k и $\overline{a}_k$, причём c_n входит в ν. Тогда, чтобы доказать утверждение для $m = k+1$, можно применить точно такое же рассуждение, как и в случае $m = 1$, только заменить 1 и 0 на a_k и $\overline{a}_k$, и a_1 и $\overline{a}_1$ на a_{k+1} и $\overline{a}_{k+1}$, и при этом использовать, что c_n начинается с a_{k+1}.

Таким образом, мы показали, что вхождение $[i, j]$ 5^n-выровнено, то есть если представлять себе ν и c_n составленными из “букв” a_n и $\overline{a}_n$, то мы доказали, что $c_n = a_n a_n a_n a_n$ входит в ν. Но заметим, что в любой последовательности, составленной из $a_1 = 10011$ и $\overline{a}_1 = 01100$, нет четырёх подряд идущих символов 0 или 1. Аналогично c_n не может входить в ν. Противоречие. $\square$

Кроме того, можно заметить следующие включения $\mathcal{P} \subset \mathcal{EP} \subset \mathcal{EAP}$, все из которых тоже, очевидно, строгие.

Назовём последовательность *точно почти периодической*, если каждое слово, которое в неё входит, входит в некоторой арифметической прогрессии. Более формально, x — точно почти периодическая, если для любого входящего в неё слова u найдутся такие $a, d \in \mathbb{N}$, что $x[a + id, a + id + |u| - 1] = u$ для всех $i \geqslant 0$. При этом u может входить и где угодно ещё в последовательности — мы не требуем, чтобы все вхождения слова u образовывали арифметическую прогрессию. Ясно, что любая точно почти периодическая последовательность почти периодична. Класс точно почти периодических последовательностей обозначим $\mathcal{PAP}$.

Введённые классы изображены на рис. 2.1.

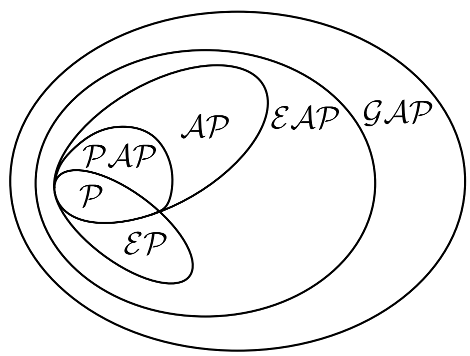

Рис. 2.1. Классы последовательностей.

Пусть $\phi\colon A^* \to B^*$ — морфизм, x обобщённо почти периодическая последовательность над алфавитом A. В [25] было показано, что если последовательность $\phi(x)$ бесконечна, то она обобщённо почти периодична. Ясно тогда, что для x — почти периодической $\phi(x)$ будет почти периодической (если бесконечна). Действительно, достаточно показать, что для любого слова $u = x[i, j]$ слово $\phi(u) = \phi(x(i)) \ldots \phi(x(j))$ встречается в $\phi(x)$ бесконечно много раз. Но это следует из определения $\phi(x)$ и из того, что x почти периодична, и значит, слово u встречается в ней бесконечно много раз. Очевидно, для x — заключительно почти периодической $\phi(x)$

будет снова заключительно почти периодической (если она бесконечна). Аналогично, для x — рекуррентной $\phi(x)$ рекуррентна (если бесконечна), для x — заключительно рекуррентной $\phi(x)$ заключительно рекуррентна (если бесконечна). Несложно видеть, что классы эффективно обобщённо почти периодических и эффективно почти периодических последовательностей также замкнуты относительно действий морфизмов.

Пусть $\phi\colon A^* \to A^*$ — некоторый морфизм, и пусть $\phi(s) = su$ для некоторых $s \in A$, $u \in A^*$. Тогда для всех натуральных $m < n$ слово $\phi^n(s)$ начинается со слова $\phi^m(s)$, так что можно корректно определить $\phi^\infty(s) = \lim_{n\to\infty} \phi^n(s) = su\phi(u)\phi^2(u)\phi^3(u)\dots$. Если $\phi^n(u) \neq \Lambda$ для всех n, то $\phi^\infty(s)$ бесконечна. В этом случае говорят, что морфизм ϕ *продолжаем на* s. Последовательности вида $h(\phi^\infty(s))$, где $h\colon A \to B$ — кодирование, называются *морфическими*, вида $\phi^\infty(s)$ — *чисто морфическими*. Последовательности вида $h(\phi^\infty(s))$, где $h\colon A \to B$ — кодирование, а ϕ — k-равномерный морфизм, называются *k-автоматными*. Когда ясно или неважно, о каком k идёт речь, приставку “k-” мы будем опускать.

Заметим, что для продолжаемых морфизмов понятия примитивности и неприводимости эквивалентны.

Автоматные последовательности называются так, потому что для них можно дать следующее эквивалентное определение. Рассмотрим конечный автомат, действующий на словах в алфавите $\Sigma_k = \{0, 1, \dots, k-1\}$ (о конечных автоматах на конечных словах см., например, [32]). Каждому состоянию автомата соответствует буква в некотором другом алфавите A (разным состояниям могут соответствовать одинаковые буквы). Автомат действует так: получает на вход слово в алфавите Σ_k, производит вычисления и выдаёт ту букву алфавита A, которая соответствует последнему состоянию в вычислении. Как показал Кобхэм [9], последовательность x букв алфавита A является k-автоматной, если и только если существует конечный автомат вышеуказанного вида, который, будучи запущенным на числе n, записанном в k-ичной системе счисления, выдаёт букву $x(n)$. Подробнее об автоматных и морфических последовательностях см. [4]

2.4. Последовательность Туэ — Морса и последовательности Штурма

Пусть $\psi\colon \mathbb{B} \to \mathbb{B}$ — морфизм, такой что $\psi(0) = 01$ и $\psi(1) = 10$. Последовательность

$$\mathbf{t} = \psi^\infty(0) = 0110100110010110100101100110 1001\ldots$$

называется *последовательностью Туэ — Морса*. Видно из определения, что последовательность Туэ — Морса автоматная. Несложно видеть также, что последовательность Туэ — Морса удовлетворяет следующим соотношениям (которые можно считать и эквивалентным определением): $t_0 = 0$ и $t_{2n} = t_n$, $t_{2n+1} = \bar{t}_n$ для всех n. Несложно доказать, что последовательность Туэ — Морса почти периодическая (например, см. [4]).

Одним из самых ранних случаев появления в литературе этой последовательности можно считать работы Туэ: в [34] и [33] он строит эту последовательность как пример бескубной бесконечной последовательности. Здесь *кубами* мы называем слова вида uuu, где u — некоторое непустое слово, а *бескубной* — последовательность, в которую не входят кубы. Более того, можно доказать, что в $\mathbf{t}$ не входят слова вида $auaua$, где a — некоторый символ, а u — некоторое слово.

Другая популярная последовательность в комбинаторике слов — *последовательность Фибоначчи* — определяется как чисто морфическая

$$\mathbf{f} = \theta^\infty(0) = 0100101001001010010 10\ldots$$

для морфизма $\theta\colon \mathbb{B} \to \mathbb{B}$, такого что $\theta(0) = 01$ и $\theta(1) = 0$.

Последовательность Фибоначчи — пример последовательности Штурма. Последовательности Штурма можно определить несколькими эквивалентными способами. Мы опишем некоторые из них, вместе с сопутствующими свойствами.

Следующий результат в терминах подсловной сложности характеризует периодические последовательности.

Предложение 2.5 ([22]). *1) Если последовательность x заключительно периодическая, её подсловная сложность ограничена: существует такая константа C, что $\mathrm{p}_x(n) \leqslant C$ для всех n.*

2) Если для некоторого n верно $\mathrm{p}_x(n) \leqslant n$, то последовательность x заключительно периодическая.

Последовательность x называется *последовательностью Штурма*, если $\mathrm{p}_x(n) = n + 1$. Таким образом, последовательности Штурма — это последовательности с минимальной возможной для апериодических последовательностей подсловной сложностью.

Множество конечных слов X называется *сбалансированным*, если для любых слов $u, v \in X$ одинаковой длины имеем $||u|_1 - |v|_1| \leqslant 1$. Конечное слово или бесконечная последовательность называется *сбалансированным/ой*, если множество его/её подслов сбалансированно.

Для действительных чисел α и ρ, таких что $0 \leqslant \alpha \leqslant 1$ и $0 \leqslant \rho < 1$, определим последовательности $s_{\alpha,\rho}$ и $s'_{\alpha,\rho}$ как $s_{\alpha,\rho}(n) = \lfloor \alpha(n+1) + \rho \rfloor - \lfloor \alpha n + \rho \rfloor$, $s'_{\alpha,\rho}(n) = \lceil \alpha(n+1) + \rho \rceil - \lceil \alpha n + \rho \rceil$. Они называются, соответственно, *нижняя* и *верхняя механические* с наклоном α и сдвигом ρ. Эти термины легче понять, если посмотреть на графическую интерпретацию механических слов (рис. 2.2). Она получается следующим образом. Чтобы получить механические последовательности $s_{\alpha,\rho}$ и $s'_{\alpha,\rho}$, нужно провести прямую $y = \alpha x + \rho$, после чего на каждой вертикальной прямой $x = n$ отметить самые близкие к прямой с обеих сторон — сверху и снизу — целые точки. Далее точки над прямой нужно соединить ломаной, она определяет верхнюю механическую последовательность: горизонтальное звено ломаной соответствует 0, а наклонённое под углом 45° соответствует 1. Нижняя механическая последовательность получается аналогично по ломаной, соединяющей точки снизу от прямой.

Для действительного числа y определим дробную часть $\mathrm{fr}(y) = y - \lfloor y \rfloor$. Заметим, что $s_{\alpha,\rho}(n) = 0$, если $0 \leqslant \mathrm{fr}(\alpha n + \rho) < 1 - \alpha$, и $s_{\alpha,\rho}(n) = 1$, если $1 - \alpha \leqslant \mathrm{fr}(\alpha n + \rho) < 1$, а также $s'_{\alpha,\rho}(n) = 0$, если $0 < \mathrm{fr}(\alpha n + \rho) \leqslant 1 - \alpha$, и $s'_{\alpha,\rho}(n) = 1$, если $1 - \alpha < \mathrm{fr}(\alpha n + \rho) \leqslant 1$.

Можно показать, что $\mathbf{f} = s_{1/\gamma^2, 1/\gamma^2}$, где $\gamma = \frac{1+\sqrt{5}}{2}$ — золотое сечение, то есть последовательность Фибоначчи является последовательностью Штурма. Она изображена на рис. 2.2.

Отметим, что верхняя и нижняя механические последовательности почти совпадают, кроме, возможно, двух символов; это несовпадение происходит в случае, если найдётся такое n, что $\alpha n + \rho \in \mathbb{Z}$. Механическая последовательность называется рациональной или иррациональной в зависимости от рациональности наклона α.

Можно рассмотреть другие варианты получения символической последовательности из прямой, нарисованной на клетчатой бумаге. На-

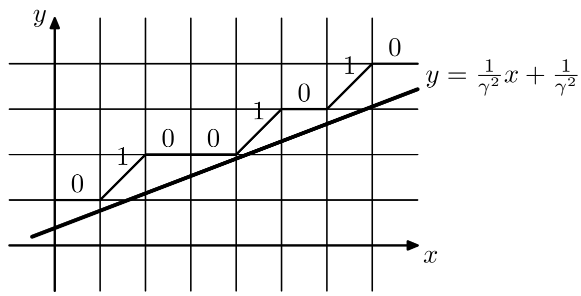

Рис. 2.2. Верхняя механическая последовательность $s'_{1/\gamma^2,1/\gamma^2} = \mathbf{f}$.

пример, можно отметить все пересечения прямой с линиями сетки, и при пересечении горизонтальной линии писать в последовательность 0, а при пересечении вертикальной писать 1. Другой вариант: можно приближать прямую сверху ломаной, составленной не из горизонтальных и наклонных отрезков, а из горизонтальных и вертикальных: при этом горизонтальный отрезок будет соответствовать 0, а вертикальный 1. Оба только что названных варианта (как и некоторые другие не названные) определяют по-прежнему класс механических последовательностей, причём от параметров этих вариантов к параметрам исходного определения можно переходить вычислимо, и наоборот.

Доказательство следующей теоремы, уже относящейся к фольклору, можно найти, например, в [20].

Теорема 2.6. *Для последовательности x следующие условия эквивалентны:*
(i) x является последовательностью Штурма;
(ii) x сбалансированная и апериодическая;
(iii) x механическая иррациональная.

Таким образом, благодаря теореме 2.6 мы имеем три эквивалентных определения последовательностей Штурма.

В заключение заметим, что любая последовательность Штурма является почти периодической (доказательство можно найти, например, в [20]).

2.5. Блочное произведение и его обобщения

Последовательность Туэ — Морса — это частный случай так называемого блочного произведения, рассмотренного в статье [18]. Пусть $u, v \in \mathbb{B}^*$.

Определим *блочное произведение* $u \otimes v$ индукцией по длине v:

$$u \otimes \Lambda = \Lambda,$$
$$u \otimes v0 = (u \otimes v)u,$$
$$u \otimes v1 = (u \otimes v)\bar{u}.$$

Легко проверить, что блочное произведение дистрибутивно справа (но не слева!) и ассоциативно, то есть $u \otimes (vw) = (u \otimes v)(u \otimes w)$ и $u \otimes (v \otimes w) = (u \otimes v) \otimes w$ для любых слов u, v, w. Пусть теперь u_k, $k = 0, 1, 2, \ldots$ — последовательность непустых слов из $\mathbb{B}^*$, таких что при $k \geqslant 1$ слово u_k начинается с 0. Тогда в последовательности $u_0, u_0 \otimes u_1, u_0 \otimes u_1 \otimes u_2, \ldots$ каждое из слов является префиксом любого из следующих, и значит, существует

$$\lim_{n \to \infty} \bigotimes_{k=0}^{n} u_k = \bigotimes_{k=0}^{\infty} u_k$$

— бесконечная последовательность в $\mathbb{B}^{\mathbb{N}}$. Например, последовательность Туэ — Морса — это блочное произведение слов $u_k = 01$.

Блочное произведение подробно изучалось в [17]. В частности, там доказано, что $\bigotimes_{k=0}^{\infty} u_k$ всегда почти периодична.

Следующий метод построения последовательностей, по существу предложенный в [38] и обсуждавшийся также в [25], в некотором смысле обобщает блочное произведение. Он является универсальным, в том смысле, что с его помощью можно получить любые обобщённо почти периодические последовательности. Мы изложим немного модифицированную версию этого метода.

Последовательность $\langle B_n, C_n, l_n \rangle$, где B_n и C_n — непустые множества непустых слов в фиксированном конечном алфавите A, l_n — натуральные числа, называется *A-$\mathcal{GAP}$-схемой*, если для неё выполняются следующие четыре условия для любого $n \in \mathbb{N}$:

(1) все слова из B_n имеют длину l_n;
(2) все слова из C_n представимы в виде $v_1 v_2$, где $v_1, v_2 \in B_n$, и каждое слово из B_n используется в качестве v_i в каком-то из слов множества C_n;
(3) каждое слово из B_{n+1} имеет вид $v_1 v_2 \ldots v_k$, где для каждого $i < k$ имеем $v_i v_{i+1} \in C_n$ (в частности, все $v_i \in B_n$), и для каждого $w \in C_n$ найдётся $i < k$, для которого $w = v_i v_{i+1}$;
(4) для каждого слова $u = v_1 v_2 \ldots v_k w_1 w_2 \ldots w_k$ из C_{n+1} (здесь $v_i, w_i \in B_n$) имеем $v_k w_1 \in C_n$.

Мы будем говорить, что последовательность $x \in A^{\mathbb{N}}$ *$\mathcal{GAP}$-порождена A-$\mathcal{GAP}$-схемой* $\langle B_n, C_n, l_n \rangle$, если для каждого $n \in \mathbb{N}$ найдётся такое k_n, что для всех $i \in \mathbb{N}$ выполнено

$$x[k_n + il_n, k_n + (i+2)l_n - 1] \in C_n.$$

Будем говорить, что последовательность *$\mathcal{GAP}$-порождена правильно*, если $l_n | k_n$.

Несложно видеть (из компактности), что каждая $\mathcal{GAP}$-схема правильно $\mathcal{GAP}$-порождает хотя бы одну последовательность. Как доказано в [25], каждая последовательность, $\mathcal{GAP}$-порождённая $\mathcal{GAP}$-схемой, является обобщённо почти периодической. Кроме того, каждая обобщённо почти периодическая последовательность $\mathcal{GAP}$-порождается некоторой $\mathcal{GAP}$-схемой, причём эту $\mathcal{GAP}$-схему можно выбрать так, чтобы последовательность порождалась правильно. Отметим также важное свойство эффективности: если обобщённо почти периодическая последовательность $\mathcal{GAP}$-порождена $\mathcal{GAP}$-схемой, то, зная параметры этой схемы (достаточно последовательностей k_n и l_n), можно вычислять оценку сверху на регулятор почти периодичности порождаемой последовательности.

Для получения почти периодических последовательностей можно также пользоваться понятием $\mathcal{GAP}$-схемы. Мы будем говорить, что последовательность $x \in A^{\mathbb{N}}$ *$\mathcal{AP}$-порождена A-$\mathcal{GAP}$-схемой* $\langle B_n, C_n, l_n \rangle$, если для всех $n \in \mathbb{N}$ и $i \in \mathbb{N}$ выполнено

$$x[il_n, (i+2)l_n - 1] \in C_n.$$

Другими словами, последовательность $\mathcal{AP}$-порождается, если она $\mathcal{GAP}$-порождается со всеми $k_n = 0$.

Аналогично предыдущему, каждая $\mathcal{GAP}$-схема $\mathcal{AP}$-порождает хотя бы одну последовательность. Каждая последовательность, $\mathcal{AP}$-порождённая $\mathcal{GAP}$-схемой, является почти периодической. Каждая почти периодическая последовательность может быть $\mathcal{AP}$-порождена какой-нибудь $\mathcal{GAP}$-схемой. Свойство эффективности сохраняется: по $\mathcal{GAP}$-схеме, порождающей последовательность, можно получить оценку сверху на регулятор почти периодичности этой последовательности.

Этот способ порождения почти периодических последовательностей можно упростить, но пожертвовав, по всей видимости, свойством эффективности.

Последовательность $\langle B_n, l_n\rangle$, где B_n — непустое множество непустых слов в фиксированном конечном алфавите A, l_n — натуральные числа, называется *A-$\mathcal{AP}$-схемой*, если для неё для любого $n \in \mathbb{N}$ выполнено условие (1), и для любого $n \in \mathbb{N}$ каждое $u \in B_{n+1}$ имеет вид $u = v_1 v_2 \dots v_k$, где $v_i \in B_n$, причём для каждого $w \in B_n$ найдётся i, такое что $v_i = w$. Последовательность x *$\mathcal{AP}$-порождена A-$\mathcal{AP}$-схемой*, если для любых i и n имеем

$$x[il_n, (i+1)l_n - 1] \in B_n.$$

Как и раньше, каждая $\mathcal{AP}$-схема $\mathcal{AP}$-порождает хотя бы одну последовательность. Каждая последовательность, $\mathcal{AP}$-порождённая $\mathcal{AP}$-схемой, является почти периодической. Каждая почти периодическая последовательность может быть $\mathcal{AP}$-порождена какой-нибудь $\mathcal{AP}$-схемой. Однако теперь уже, вообще говоря, может быть неверно, что по схеме можно оценивать сверху регулятор почти периодичности порождаемой последовательности. По крайней мере, то же рассуждение, которое можно было применить раньше, здесь уже не работает. Например, неясно даже, как, зная схему, дать оценку на расстояния между соседними вхождениями двухбуквенных слов: такое слово может оказаться на стыке двух блоков — элементов схемы.

Когда ясно, о каком алфавите, о каком типе схемы или о каком типе порождения последовательности идёт речь, мы будем опускать соответствующие приставки.

2.6. Конечные автоматы и конечно-автоматные преобразователи

Назовём (недетерминированным) *автоматом Бюхи* совокупность $M = \langle A, Q, \tilde{q}, \Delta, F\rangle$, где A — входной алфавит, Q — множество состояний, $\tilde{q} \in Q$ — начальное состояние, $\Delta \subseteq Q \times A \times Q$ — множество переходов, $F \subseteq Q$ — множество заключительных состояний. Ходом автомата M на последовательности $x = x(0)x(1)x(2)\dots$ называется такая последовательность состояний $\rho = \rho(0)\rho(1)\rho(2)\dots$, что $\rho(0) = \tilde{q}$ и $\langle \rho(i), x(i), \rho(i+1)\rangle \in \Delta$ для любого i. Мы говорим, что автомат M *принимает* x, если существует хотя бы один ход ρ автомата M на x, для которого хотя бы одно состояние, встречающееся в ρ бесконечное количество раз, входит в множество заключительных состояний F. Определяя *детерминированный автомат Бюхи*, множество переходов $\Delta \subseteq Q \times A \times Q$ можно заменить на функцию переходов $\delta\colon Q \times A \to Q$ (с естественными изменениями для определения хода автомата).

Есть немного другой вариант понятия автомата на последовательности. Назовём (недетерминированным) *автоматом Мюллера* совокупность $M = \langle A, Q, \tilde{q}, \Delta, \mathcal{F} \rangle$, где A — входной алфавит, Q — множество состояний, $\tilde{q} \in Q$ — начальное состояние, $\Delta \subseteq Q \times A \times Q$ — множество переходов, $\mathcal{F} \subseteq 2^Q$ — множество заключительных макросостояний. Здесь под *макросостоянием* мы понимаем элемент множества 2^Q, то есть произвольное подмножество множества состояний Q. Ходом автомата M на последовательности $x = x(0)x(1)x(2)\ldots$ называется такая последовательность состояний $\rho = \rho(0)\rho(1)\rho(2)\ldots$, что $\rho(0) = \tilde{q}$ и $(\rho(i), x(i), \rho(i+1)) \in \Delta$ для любого i. Назовём *пределом (предельным макросостоянием)* автомата M на последовательности x вдоль хода ρ множество всех таких состояний, которые встречаются в ρ бесконечное количество раз, обозначим это множество через $\lim_\rho M$. Мы говорим, что автомат M *принимает* x, если существует хотя бы один ход ρ автомата M на x, для которого $\lim_\rho M \in \mathcal{F}$. Другими словами, слово принимается, если хотя бы на каком-нибудь ходе предельное макросостояние принадлежит множеству заключительных макросостояний $\mathcal{F}$. Аналогично предыдущему, определяя *детерминированный автомат Мюллера*, множество переходов $\Delta \subseteq Q \times A \times Q$ можно заменить на функцию переходов $\delta\colon Q \times A \to Q$ (с естественными изменениями для определения хода автомата).

Для автомата M Бюхи или Мюллера множество всех последовательностей, которые принимаются автоматом M, обозначим $L(M)$.

Для примера рассмотрим множество L последовательностей в алфавите $\{a, b, c\}$, в которые если a входит бесконечное количество раз, то и b входит бесконечное количество раз. Автоматы Мюллера и Бюхи, принимающие в точности последовательности множества L, показаны на рис. 2.3 (автомат Бюхи понадобился недетерминированный).

Это пример общей ситуации.

Теорема 2.7 ([21]). *Недетерминированные автоматы Бюхи, недетерминированные автоматы Мюллера и детерминированные автоматы Мюллера распознают один и тот же класс множеств последователь-*

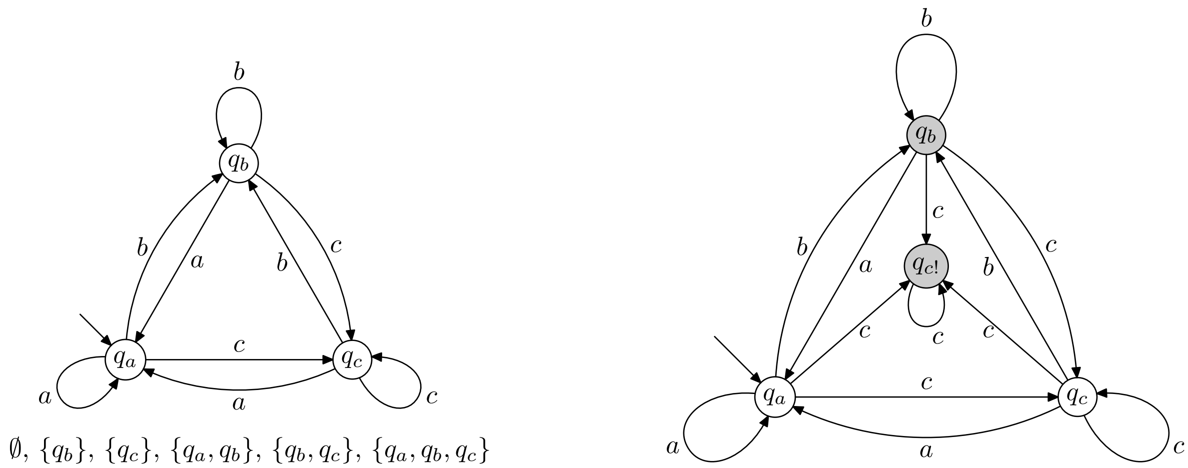

Рис. 2.3. Пример автомата Мюллера (слева) и автомата Бюхи (справа), принимающих одно и то же множество последовательностей. Множество принимающих макросостояний автомата Мюллера дано списком, заключительные состояния автомата Бюхи заштрихованы.

ностей. Более того, по автомату одного типа можно получать эквивалентный автомат другого типа алгоритмически.

Множество последовательностей, распознаваемое (недетерминированным) автоматом Мюллера или (недетерминированным) автоматом Бюхи, будем называть *регулярным*. Детерминированные автоматы Бюхи распознают меньший класс множеств.

Конечно-автоматным преобразователем назовём совокупность $M = \langle A, B, Q, \tilde{q}, \lambda, \mu \rangle$, где A и B — конечные множества, называемые соответственно входной и выходной алфавит, Q — конечное множество состояний, $\tilde{q} \in Q$ — выделенное состояние, называемое начальным, и

$$\lambda\colon Q \times A \to B^*, \qquad \mu\colon Q \times A \to Q$$

— функции переходов. Функции переходов можно естественным образом продолжить на $Q \times A^*$ по индукции: $\mu(q, ua) = \mu(\mu(q, u), a)$ и $\lambda(q, ua) = \lambda(q, u)\lambda(\mu(q, u), a)$ для $q \in Q$, $u \in A^*$, такого что $u \neq \Lambda$, и $a \in A$.

Пусть $x \in A^{\mathbb{N}}$. Последовательность $(p_n)_{n=0}^{\infty}$ элементов множества Q назовём *ходом преобразователя M на x*, если $p_0 = \tilde{q}$ и для каждого n выполняется $p_{n+1} = \mu(p_n, x(n))$. Ясно, что ход существует и единствен. Последовательность $M(x)$, определяемую как $M(x)(n) = \lambda(p_n, x(n))$, где $(p_n)_{n=0}^{\infty}$ — ход преобразователя M на x, назовём *образом последовательности x под действием M*.

Если для каждых $a \in A$, $q \in Q$ выполнено $|\lambda(q,a)| = 1$, то преобразователь M называется *равномерным.* Несложно видеть, что применение произвольного конечно-автоматного преобразователя к последовательности можно представить как последовательное применение равномерного конечно-автоматного преобразователя и некоторого морфизма.

Предложение 2.8. *Пусть $M = \langle A, B, Q, \tilde{q}, \lambda, \mu \rangle$ — конечно-автоматный преобразователь, x — последовательность. Тогда существует такой равномерный конечно-автоматный преобразователь M' и такой морфизм ϕ, что $M(x) = \phi(M'(x))$.*

Доказательство. Действительно, положим $M' = \langle A, Q \times A, Q, \tilde{q}, \lambda, \mu' \rangle$, так что $\mu'(q,a) = \langle q, a \rangle$ для любых $q \in Q$ и $a \in A$. Определим также морфизм $\phi\colon (Q \times A)^* \to B^*$, так что $\phi(\langle q, a \rangle) = \mu(q,a)$ для любых $q \in Q$ и $a \in A$. Ясно тогда, что $M(x) = \phi(M'(x))$. $\square$

Поэтому часто для упрощения ситуации мы ограничиваемся рассмотрением равномерных конечно-автоматных преобразователей. Если $[i,j]$ — вхождение слова u в последовательность x, причём $p_i = q$, где $(p_n)_{n=0}^{\infty}$ — ход преобразователя M на x, то будем говорить, что преобразователь M *подходит* к этому вхождению слова u *в состоянии* q.

Назовём конечно-автоматный преобразователь $M = \langle A, B, Q, \tilde{q}, \lambda, \mu \rangle$ *обратимым*, если для каждых $q \in Q$ и $a \in A$ существует ровно одно состояние $q' \in Q$, такое что $\mu(q', a) = q$. Другими словами, в таком преобразователе каждая буква входного алфавита осуществляет взаимно однозначное отображение множества состояний в себя. Находясь в некотором состоянии и зная последовательность предыдущих входных символов, можно восстановить и последовательность пройденных состояний (в этом и заключается свойство обратимости).

2.7. Произведение с периодической последовательностью

На последовательностях можно определить операцию $\times$, которую мы будем называть *произведением.* Для $x \in A^{\mathbb{N}}$, $y \in B^{\mathbb{N}}$ определим $x \times y \in (A \times B)^{\mathbb{N}}$, так что $(x \times y)(i) = \langle x(i), y(i) \rangle$.

Как несложно видеть, если последовательность y периодическая с периодом m, то $x \times y$ можно получить как результат равномерного конечно-автоматного преобразования последовательности x преобразователем с m состояниями. Отсюда получаем такое следствие из предложения 3.6 (этот результат сам по себе, по-видимому, уже является фольклором).

Предложение 2.9. *Если $x \in \mathcal{AP}$ и $y \in \mathcal{P}$, то $x \times y \in \mathcal{AP}$.*

Докажем, что класс точно почти периодических последовательностей замкнут относительно операции умножения на периодическую последовательность.

Предложение 2.10. *Если $x \in \mathcal{PAP}$ и $y \in \mathcal{P}$, то $x \times y \in \mathcal{PAP}$.*

Доказательство. Будем считать, что $y = 01\dots(m-1)01\dots(m-1)01\dots$, где m — период y. Случай произвольного $y \in \mathcal{P}$ получается из этого кодированием, которое, очевидно, сохраняет $\mathcal{PAP}$. В этом доказательстве для краткости вместо $u \times [i, i+|u|-1]$ будем писать $u \times [i, \cdot]$.

Пусть u входит в x, и $A = \{i : 0 \leqslant i \leqslant m-1, u \times [j, \cdot]$ входит в z для некоторого $j \equiv i \pmod m\}$. Пусть $v \times [a, \cdot]$ — такое подслово z, которое для каждого $i \in A$ содержит вхождение $u \times [j, \cdot]$ для некоторого $j \equiv i \pmod m$. Слово v входит в x, поэтому для некоторых b, d имеем $x[b+dt, b+dt+|v|-1] = v$ для каждого $t \in \mathbb{N}$. Ясно при этом, что $B = \{i : 0 \leqslant i \leqslant m-1, u \times [j, \cdot]$ входит в $v \times [b, \cdot]$ для некоторого $j \equiv i \pmod m\} = A$. Действительно, по определению A имеем $B \subseteq A$, но $B = \{(i+b-a) \mod m : i \in A\}$. Значит, $\#B = \#A$, откуда $B = A$.

Таким образом, если $u \times [s, \cdot]$ входит в z, то для некоторого $p \equiv s \pmod m$ имеем вхождение $u \times [p, \cdot]$ в $v \times [b, \cdot]$. Поэтому $u \times [p + tdm, \cdot]$ входит в z для каждого $t \in \mathbb{N}$. Следовательно, $z \in \mathcal{PAP}$. $\square$

Автору неизвестно, сохраняется ли класс точно почти периодических последовательностей под действием конечно-автоматных преобразований.

Глава 3.

Конечно-автоматные преобразования

В этой главе мы доказываем, что заключительно почти периодические последовательности замкнуты относительно конечо-автоматных преобразований (теорема 3.4, впервые доказана в [39]), а также получаем эффективный вариант этого результата (теорема 3.5, объявлена в [41], доказана в [40]). После этого мы доказываем, что заключительно рекуррентные последовательности замкнуты относительно конечно-автоматных преобразований (теорема 3.8, доказано в [43]). Мы доказываем, что обобщённо почти периодические последовательности замкнуты относительно почти обратимых конечно-автоматных преобразований, и получаем оценку на регулятор образа при таких преобразованиях (предложение 3.11, доказано в [43]). Наконец, мы доказываем, что обобщённо почти периодические последовательности не сохраняются при стековых конечно-автоматных преобразованиях (предложение 3.12, доказано в [40]).

3.1. Конечно-автоматные преобразования последовательностей со свойствами типа почти периодичности

Представляется интересным рассматривать преобразования последовательностей и пытаться понять, сохраняют ли эти преобразования свойства почти периодичности. Простейшими алгоритмическими преобразованиями можно считать конечно-автоматные.

Для различных классов последовательностей известны результаты о замкнутости относительно конечно-автоматных преобразований: класс автоматных последовательностей сохраняется при равномерных конечно-автоматных преобразованиях (см. [9, 4]), класс морфических последовательностей сохраняется при конечно-автоматных преобразованиях (см. [10, 4]). В настоящем разделе мы обсуждаем замкнутость относительно таких преобразований последовательностей с различными свойствами типа почти периодичности.

По существу, следующая теорема доказана в [38] (см. также [25, 43]). Для произвольной функции g обозначим $\underbrace{g \circ g \circ \cdots \circ g}_{m}$ через g^m.

Теорема 3.1 ([38, 25, 43]). *Пусть M — конечно-автоматный преобразователь с m состояниями, и $x \in \mathcal{GAP}$. Тогда верно следующее.*

1) *$M(x) \in \mathcal{GAP}$.*

2) *Пусть M — равномерный преобразователь. Тогда $\mathrm{R}_{M(x)}(n) \leqslant h(h(n))$ для всех n, где $h(n) = g^m(n) - 1$ и $g(n) = \mathrm{R}_x(n) + 1$.*

3) *Если x эффективно обобщённо почти периодическая, то $M(x)$ также эффективно обобщённо почти периодическая.*

Ключевой в доказательстве теоремы 3.1 является следующая лемма. Доказательство именно для этой формулировки в явном виде можно найти в [43]. Эта лемма понадобится нам в дальнейшем.

Лемма 3.2. *Пусть M — равномерный конечно-автоматный преобразователь с m состояниями, и $x \in \mathcal{GAP}$. Пусть $v = M(x)[i, j]$ — вхождение слова длины n в $M(x)$, такое что $i \geqslant h(n)$, где $h(t) = g^m(t) - 1$, $g(t) = \mathrm{R}_x(t) + 1$. Тогда найдётся такое r, что $j - h(n) \leqslant r \leqslant i - 1$ и $M(x)[r, r + n - 1] = v$.*

Следующий результат следует из теоремы 3.1 и по существу является переформулировкой (см. следствие 1.5) теоремы 1.6 — критерия разрешимости монадической теории первого порядка обобщённо почти периодической последовательности, доказанного в [38].

Теорема 3.3 ([38]). *Существует алгоритм, который, получив на вход оракул для последовательности $x \in \mathcal{GAP}$, оракул для некоторой функции $g \geqslant \mathrm{R}_x$ и автомат Бюхи F, определяет, верно ли, что F принимает x.*

Но если результат теоремы 3.1 в [38] можно считать вспомогательным для доказательства критерия разрешимости монадической теории обобщённо почти периодической последовательности, то в настоящей работе мы считаем его представляющим самостоятельный интерес и получаем результаты аналогичного типа.

Из теоремы 3.1 сразу следует, что образ при конечно-автоматном преобразовании заключительно почти периодической последовательности обобщённо почти периодичен. Но оказывается, можно доказать следующее более сильное утверждение — тот факт, что оно более сильное, следует из предложения 2.4. Для полноты изложения мы приводим его с доказательством по [39], несмотря на то, что оно является следствием из доказанной ниже теоремы 3.5.

Теорема 3.4. *Пусть M — конечно-автоматный преобразователь. Тогда если $x \in \mathcal{EAP}$, то $M(x) \in \mathcal{EAP}$.*

Доказательство. Очевидно, что достаточно доказать утверждение для $x \in \mathcal{AP}$, так как заключительно почти периодическая последовательность остаётся заключительно почти периодической при приписывании к ней произвольного префикса.

Заметим также, что в виду предложения 2.8 достаточно доказать теорему только для равномерных конечно-автоматных преобразователей, поскольку морфизмы сохраняют почти периодичность. Поэтому будем считать, что M равномерный.

Итак, пусть $x \in \mathcal{AP}$. По теореме 3.1 последовательность $M(x)$ обобщённо почти периодическая. Предположим, что она не является заключительно почти периодической. Это означает, что для любого натурального N можно найти такое слово, которое входит в $M(x)$ правее позиции N и после этого уже не входит. Действительно, отрезав от $M(x)$ начальный отрезок $[0, N]$, мы не получим почти периодическую последовательность, значит, найдётся слово, входящее в неё ненулевое конечное количество раз. Тогда возьмём самое правое вхождение этого слова.

Пусть $[i_0, j_0]$ — самое правое вхождение некоторого слова y_0 в последовательность $M(x)$. Для некоторого натурального l_0 слово $u_0 = x[i_0, j_0]$ входит в любой отрезок длины l_0 последовательности x (в силу почти периодичности). При этом если q_0 — состояние, в котором преобразователь M подошёл к позиции i_0, то ко всем дальнейшим вхождениям слова

u_0 в x преобразователь не может подойти в состоянии q_0, иначе бы он выдал полностью слово y_0.

Пусть теперь $[r, s]$ — самое правое вхождение некоторого слова a в последовательность $M(x)$, причём $r > i_0 + l_0$. На отрезке $x[r - l_0, r]$ найдётся вхождение $[r', s']$ слова u_0, причём в силу выбора r будет выполнено $r' > i_0$. Положим тогда

$$i_1 = r',\ j_1 = s,\ u_1 = x[i_1, j_1],\ y_1 = M(x)[i_1, j_1].$$

Поскольку слово a не входит в $M(x)$ правее позиции r, то и слово y_1, которое содержит a в качестве подслова, не входит в $M(x)$ правее позиции i_1. Это значит, что если q_1 — состояние, в котором преобразователь подошёл к позиции i_1, то больше никогда в состоянии q_1 к вхождению слова u_1 в последовательность x преобразователь не подойдёт. Поскольку u_1 начинается со слова $x[r', s'] = u_0$, и $r' > i_0$, то $q_1 \neq q_0$. Таким образом, мы нашли слово u_1, ко всем вхождениям которого, расположенным правее i_1, преобразователь не может подойти, находясь в состояниях q_0 или q_1.

Пусть $m = |Q|$ — количество возможных состояний конечно-автоматного преобразователя M. Продолжая так дальше рассуждать по индукции, мы построим цепочку слов $u_k = x[i_k, j_k]$ и соответствующих различных состояний q_k, где $k < m$, так что ко всем вхождениям u_k в x правее i_k преобразователь не может подойти в состояниях $q_0, q_1, \ldots, q_k$. При $k = m$ получаем противоречие. □

Доказательство теоремы 3.4, приведённое выше, не эффективно в следующем смысле. Допустим, мы знаем $x \in \mathcal{AP}$ и регулятор R_x. Тогда по теореме 3.4 существует оценка сверху на $\mathrm{pr}(M(x))$, так как $M(x) \in \mathcal{EAP}$, но приведённое доказательство не позволяет по имеющимся данным найти никакую такую оценку эффективно. Ан. А. Мучник на основе теоремы 3.4 изначально высказал гипотезу о том, что такой эффективной оценки не существует.

Однако оказывается, что можно доказать следующий эффективный вариант теоремы 3.4.

Теорема 3.5. *Пусть M — равномерный конечно-автоматный преобразователь с m состояниями, и $x \in \mathcal{AP}$. Тогда $M(x) \in \mathcal{EAP}$ и*

$$\mathrm{pr}(M(x)) \leqslant \mathrm{R}_x^m(1) + \mathrm{R}_x^{m-1}(1) + \cdots + \mathrm{R}_x(1).$$

Доказательство проходит индукцией по количеству состояний преобразователя. База индукции — обратимый конечно-автоматный преобразователь.

Предложение 3.6. *Пусть M — обратимый равномерный конечно-автоматный преобразователь. Тогда если $x \in \mathcal{AP}$, то $M(x) \in \mathcal{AP}$.*

Доказательство. Пусть $v = M(x)[i, j]$ — вхождение слова v длины n в $M(x)$, и $u_1 = x[i, j]$ — прообраз v в x, к которому M подходит в состоянии q_1. Тогда найдётся вхождение $u_1 = x[i_2, j_2]$, такое что $i_2 > i$, но $j_2 \leqslant i + \mathrm{R}_x(n)$. Если M подходит к i_2 в состоянии q_1, то $M(x)[i_2, j_2] = v$. Иначе M подходит к i_2 в каком-то состоянии $q_2 \neq q_1$.

Положим $u_2 = x[i, j_2]$. Имеем $|u_2| \leqslant \mathrm{R}_x(n) + 1$. Тогда найдётся вхождение $u_2 = x[i_3, j_3]$, такое что $i_3 > i$, но $j_3 \leqslant i + \mathrm{R}_x(\mathrm{R}_x(n) + 1)$. Если M подходит к $j_3 - n + 1$ в состоянии q_1, то $M(x)[j_3 - i + 1, j_3] = v$, так как u_2 заканчивается словом u_1. Если M подходит к $j_3 - n + 1$ в состоянии q_2, то в силу обратимости M подходит к i_2 в состоянии q_1, и тогда $M(x)[i_2, i_2 + n - 1] = v$, так как u_2 начинается с u_1. В худшем случае M подходит к $j_3 - n + 1$ в состоянии q_3, таком что $q_3 \neq q_1$, $q_3 \neq q_2$ (рис. 3.1).

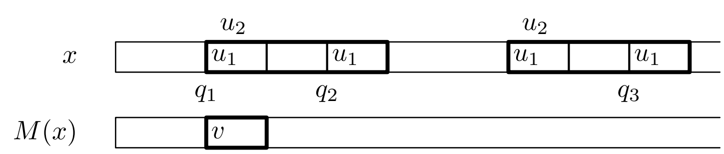

Рис. 3.1. Иллюстрация к доказательству предложения 3.6.

Пусть m — количество состояний автомата. Рассуждая так дальше, построим такие $j_2, \ldots, j_{m+1}$, что хотя бы к каким-то двум из $i = j - n + 1 = j_1 - n + 1, j_2 - n + 1, \ldots, j_{m+1} - n + 1$ преобразователь подходит в одинаковых состояниях. При этом $j_{m+1} \leqslant i + g^m(n) - 1$, где $g = \mathrm{R}_x + 1$.

Таким образом, каждое слово длины n, входящее в $M(x)$ бесконечно много раз, входит на любом отрезке длины $g^m(n) - 1$. Следовательно, $M(x) \in \mathcal{AP}$. $\square$

Для индукционного перехода рассмотрим следующую конструкцию. Пусть $x \in A^{\mathbb{N}}$, и символ $a \in A$ входит в x бесконечное количество раз. Проведём разрез в x после каждого вхождения символа a. Тогда x разрежется на блоки вида ua, где $u \in (A \setminus \{a\})^*$, то есть на слова, содержащие ровно один символ a на последнем месте. Если символ a встречается в x с ограниченными интервалами, то количество всевозможных таких блоков конечно (например, если $x \in \mathcal{GAP}$, то их длины не больше $\mathrm{R}_x(1)$). Закодируем однозначно эти блоки буквами некоторого конечного алфавита, обозначим этот алфавит $\mathrm{block}_{a,x}$. Морфизм декодирования обозначим d: если $c \in b_{a,x}$ — буква нового алфавита, то $d(c) \in A^*$ — блок, заканчивающийся буквой a, из которого эта буква получилась. Таким образом, мы из x получили новую последовательность в алфавите $\mathrm{block}_{a,x}$. Последовательность, полученную из неё отбрасыванием первого символа, назовём *a-разбиением* последовательности x и обозначим $\mathrm{split}_{a,x}$. Отброшенный первый символ, соответствующий в x префиксу до первого вхождения a включительно, обозначим $\mathrm{cut}_{a,x}$. Таким образом, имеем: $x = d(\mathrm{cut}_{a,x}\,\mathrm{split}_{a,x}) = d(\mathrm{cut}_{a,x})d(\mathrm{split}_{a,x}(0))d(\mathrm{split}_{a,x}(1))d(\mathrm{split}_{a,x}(2))\dots$ Например, 0-разбиение последовательности $x = 3200122403100110\dots$ — это $\mathrm{split}_{0,x} = (0)(12240)(310)(0)(110)\dots$, причём $\mathrm{cut}_{a,x} = (320)$ и (320), (0), $(12240), \dots \in \mathrm{block}_{a,x}$.

Лемма 3.7. *Пусть $x \in \mathcal{AP}$, и символ a входит в x. Тогда*

1) $\mathrm{split}_{a,x} \in \mathcal{AP}$ *и*

2) $|d(c)| \leqslant \mathrm{R}_x(1)$ *для любого* $c \in \mathrm{block}_{a,x}$.

Доказательство. 1) Пусть $u = \mathrm{split}_{a,x}[0, n-1]$. Положим $v = a\,d(u)$. Существует l, такое что v встречается в каждом подслове длины l в x. Пусть $w = w(0)\dots w(l-1)$ — подслово длины l в $\mathrm{split}_{a,x}$. Тогда $d(w)$ — подслово длины не менее l в x, и значит, в него входит v. Слово v начинается и заканчивается буквой a, при этом a входит только в конце слов $d(w(0))$, $d(w(1))$, $\dots$, $d(w(l-1))$. Отсюда следует, что u входит в w.

2) Если $d(c) \geqslant \mathrm{R}_x(1) + 1$ для некоторого $c \in \mathrm{block}_{a,x}$, то $d(c)[0, \mathrm{R}_x(1) - 1]$ — подслово длины $\mathrm{R}_x(1)$ в x без вхождений слова a длины 1 — противоречие с определением регулятора. $\square$

Теперь мы готовы провести доказательство теоремы 3.5.

Доказательство теоремы 3.5. Пусть $M_0 = M = \langle A, B, Q, \tilde{q}, \lambda, \mu \rangle$ — равномерный преобразователь, $x_0 = x$ — почти периодическая последовательность в алфавите A. Будем считать без ограничения общности, что $B = Q \times A$ и для всех $q \in Q$, $a \in A$ выполнено $\lambda(q, a) = \langle q, a \rangle$. Действительно, общий случай произвольного B и произвольного λ получается из рассматриваемого отождествлением каких-то пар $\langle q, a \rangle$ между собой. Аналогичное предположение будет выполнено и для всех других преобразователей, которые встретятся нам в доказательстве.

Как уже говорилось выше, доказательство утверждения теоремы можно провести формальной индукцией по количеству состояний преобразователя. Это позволит доказать $M(x) \in \mathcal{EAP}$, но даст худшую оценку на $\mathrm{pr}(M(x))$, чем заявлено в утверждении теоремы. Поэтому мы проведём рассуждения явно, "развернём" индукцию.

Итак, возможны два случая.

1) Преобразователь M — обратимый. Тогда утверждение теоремы следует из предложения 3.6, при этом $\mathrm{pr}(M(x)) = 0$.

2) Преобразователь M не является обратимым. В этом случае найдётся такой символ $a \in A$, который осуществляет с помощью μ на Q не взаимно однозначное отображение, так что имеем $Q' = \{q : \exists q' \; \mu(q', a) = q\} \subsetneq Q$, то есть $|Q'| < |Q|$. Заметим тогда, что преобразователь M, запущенный на x, после каждого вхождения символа a оказывается в одном из состояний множества Q', строго меньшего, чем Q.

Определим конечно-автоматный преобразователь $M' =$ $= \langle \mathrm{block}_{a,x}, B', Q', q', \lambda', \mu' \rangle$, так что $q' = \mu(\tilde{q}, d(\mathrm{cut}_{a,x}))$, $\mu'(q, c) = \mu(q, d(c))$ и $\lambda'(q, d(c))$ для любых $q \in Q$ и $c \in \mathrm{block}_{a,x}$. Поскольку все слова $d(c)$ для $c \in \mathrm{block}_{a,x}$ оканчиваются на a, функции переходов μ' и λ' корректно определены, и их значения не выходят за рамки множества состояний Q'. Положим $B' = Q' \times \mathrm{block}_{a,x}$.

Таким образом, мы имеем последовательность $x_2 = \mathrm{split}_{a,x}$ и преобразователь $M_2 = M'$, действующий на x_2 и имеющий $|Q'| < m$ состояний. При этом $x_2 \in \mathcal{AP}$ по лемме 3.7, п. 1. Будем итерировать эту конструкцию и строить почти периодические последовательности $x_3, x_4, \ldots$ и соответствующие конечно-автоматные преобразователи $M_3, M_4, \ldots$ со строго уменьшающимся количеством состояний, пока не получим последовательность x_k, на которой действует обратимый конечно-автоматный преобразователь M_k. При этом поскольку количество состояний при каждой итерации строго уменьшается, и любой преобразователь с одним состоянием обратим, имеем $k \leqslant m$.

Обозначим через Block алфавит последовательности x_k, через D — композицию всех операций декодирования, соответствующих каждой итерации, и через Cut — конкатенацию отброшенных на каждой итерации префиксов, так что $x = \mathrm{Cut}\; D(x_k)$. Из построения по лемме 3.7, п. 2 следует, что $|\,\mathrm{Cut}\,| \leqslant \mathrm{R}_x^k(1) + \ldots + \mathrm{R}_x(1)$. Кроме того, из предложения 3.6 следует, что $M_k(x_k)$ — почти периодическая.

Пусть $M(x) = uy$, где $|u| = |\,\mathrm{Cut}\,|$. Видно тогда, что y получается из $M_k(x_k)$ последовательным раскодированием, то есть применением некоторого морфизма. Действительно, посмотрим на одну итерацию — переход от x_i и M_i к x_{i+1} и M_{i+1}. Алфавит последовательности $M_i(x_i)$ — $Q_i \times B$, где B — алфавит последовательности x_i, Q_i — множество состояний M_i, алфавит последовательности $M_{i+1}(x_{i+1})$ — $Q_{i+1} \times \mathrm{block}_{a,x_i}$, где Q_{i+1} — множество состояний M_{i+1}, a — некоторая буква алфавита B. Видно тогда, что буквы последовательности $M_i(x_i)$ получаются из букв $q \times t$ последовательности $M_{i+1}(x_{i+1})$ следующим образом: к компоненте t применяем декодирование, соответствующее переходу от алфавита block_{a,x_i} последовательности x_{i+1} к алфавиту B последовательности x_i, после чего запускаем M_i на получившемся слове и получаем слово в алфавите $Q_i \times B$. Нужный морфизм получается из описанных морфизмов композицией по всем итерациям.

Поскольку морфизмы сохраняют почти периодичность, получаем, что $y \in \mathcal{AP}$ и $|u| \leqslant \mathrm{R}_x^k(1) + \ldots + \mathrm{R}_x(1) \leqslant \mathrm{R}_x^m(1) + \ldots + \mathrm{R}_x(1)$. □

Примечательно, что, таким образом, мы имеем два различных доказательства теоремы 3.4. Отметим также, что нижняя оценка, показывающая, что верхнюю оценку в теореме 3.5 нельзя существенно улучшить, получена в [44].

3.2. Конечно-автоматные преобразования рекуррентных последовательностей

Заметим, что, формально говоря, все введённые нами до настоящего момента классы последовательностей — это классы последовательностей над конечным алфавитом. Однако все те же самые определения (почти периодической последовательности, обобщённо почти периодической последовательности и т. д.) имеет смысл рассматривать и для бесконечного алфавита. На бесконечный алфавит можно обобщить и понятие конечно-автоматного преобразователя. Оно остаётся точно таким же, как приведённое на с. 39 для конечных алфавитов, кроме упоминания того, что входной и выходной алфавиты конечны. При этом от функций переходов не требуется никакой эффективности — они могут быть совершенно произвольными.

Идея рассмотрения бесконечного алфавита приводит к новому результату в том числе и для класса последовательностей над конечным алфавитом.

Теорема 3.8. *Пусть M — конечно-автоматный преобразователь. Тогда если $x \in \mathcal{ER}$, то $M(x) \in \mathcal{ER}$.*

Доказательство этого утверждения следует схеме доказательства теоремы 3.5. Единственное важное дополнительное замечание — утверждение нужно доказывать над бесконечным алфавитом, по-прежнему индукцией по количеству состояний преобразователя. Таким образом, всюду в этом разделе мы будем считать алфавиты счётными множествами. Заметим также, что рассуждение, приведённое выше при доказательтве теоремы 3.4, не удаётся обобщить на интересующий нас случай рекуррентных последовательностей.

Итак, первый шаг — рассмотрение обратимого преобразователя. Доказательство предложения 3.6 повторяется почти дословно, но становится даже проще.

Предложение 3.9. *Пусть M — обратимый равномерный конечно-автоматный преобразователь. Тогда если $x \in \mathcal{R}$, то $M(x) \in \mathcal{R}$.*

Доказательство. Пусть $v = M(x)[i,j]$ — вхождение слова v длины n в $M(x)$, и $u_1 = x[i,j]$ — прообраз v в x, к которому M подходит в состоянии q_1. Тогда найдётся вхождение $u_1 = x[i_2, j_2]$, такое что $i_2 > i$, поскольку $x \in \mathcal{R}$. Если M подходит к i_2 в состоянии q_1, то $M(x)[i_2, j_2] = v$. Иначе M подходит к i_2 в каком-то состоянии $q_2 \neq q_1$.

Положим $u_2 = x[i, j_2]$. Найдётся вхождение $u_2 = x[i_3, j_3]$, такое что $i_3 > i$, поскольку $x \in \mathcal{R}$. Если M подходит к $j_3 - n + 1$ в состоянии q_1, то $M(x)[j_3 - i + 1, j_3] = v$, так как u_2 заканчивается словом u_1. Если M подходит к $j_3 - n + 1$ в состоянии q_2, то в силу обратимости M подходит к i_2 в состоянии q_1, и тогда $M(x)[i_2, i_2 + n - 1] = v$, так как u_2 начинается с u_1. В худшем случае M подходит к $j_3 - n + 1$ в состоянии q_3, таком что $q_3 \neq q_1$, $q_3 \neq q_2$ (рис. 3.1).

Пусть m — количество состояний автомата. Рассуждая так дальше, построим такие $j_2, \ldots, j_{m+1}$, что хотя бы к каким-то двум из $i = j - n +$

$1 = j_1 - n + 1, j_2 - n + 1, \ldots, j_{m+1} - n + 1$ преобразователь подходит в одинаковых состояниях. Таким образом, для каждого слова, входящего в $M(x)$, мы нашли ещё одно его вхождение в $M(x)$ справа. Следовательно, $M(x) \in \mathcal{R}$. $\square$

Так же как и в случае заключительной почти периодичности, мы рассматриваем разбиение последовательности. Заметим при этом, что a-разбиение рекуррентной последовательности даже и над конечным алфавитом может быть последовательностью над бесконечным алфавитом, поскольку различных блоков между последовательными вхождениями a может быть бесконечное количество. Однако это никак не отражается на следующей лемме, полученной вместе с доказательством почти без изменений из леммы 3.7.

Лемма 3.10. *Пусть $x \in \mathcal{R}$, и символ a входит в x. Тогда* $\mathrm{split}_{a,x} \in \mathcal{R}$.

Доказательство. Пусть u — подслово в x. Положим $v = a\,d(u)$. Слово v входит в x бесконечно много раз. Оно начинается и заканчивается буквой a, при этом a входит только в конце слов $d(u(i))$, $i \in \mathbb{N}$. Отсюда следует, что каждому вхождению v в x соответствует вхождение u в $\mathrm{split}_{a,x}$. $\square$

Теперь мы готовы к доказательству теоремы.

Доказательство теоремы 3.8. Ясно, что достаточно доказать утверждение теоремы для равномерного конечно-автоматного преобразователя и рекуррентной входной последовательности.

Пусть $M_0 = M = \langle A, B, Q, \tilde{q}, \lambda, \mu \rangle$ — равномерный преобразователь с m состояниями, $x_0 = x$ — рекуррентная последовательность в алфавите A. Будем считать без ограничения общности, что $B = Q \times A$ и для всех $q \in Q$, $a \in A$ выполнено $\lambda(q, a) = \langle q, a \rangle$. Общий случай произвольного B и произвольного λ получается из рассматриваемого отождествлением каких-то пар $\langle q, a \rangle$ между собой.

Возможны два случая.

1) Преобразователь M — обратимый. Тогда утверждение теоремы следует из предложения 3.6.

2) Преобразователь M не является обратимым.

В этом случае точно так же, как в доказательстве теоремы 3.5, с помощью итеративной процедуры построим последовательности

$x_2, x_3, x_4, \ldots$ и соответствующие конечно-автоматные преобразователи $M_2, M_3, M_4, \ldots$ со строго уменьшающимся количеством состояний, пока не получим последовательность x_k, на которой действует обратимый конечно-автоматный преобразователь M_k. При этом все x_i рекуррентны по лемме 3.10. По построению $k \leqslant m$.

Обозначим через Cut конкатенацию отброшенных на каждой итерации префиксов, так что $x = \mathrm{Cut}\ D(x_k)$. По предложению 3.9 $M_k(x_k) \in \mathcal{R}$. Пусть $M(x) = uy$, где $|u| = |\mathrm{Cut}|$. Аналогично доказательству теоремы 3.5, несложно видеть, что y получается из $M_k(x_k)$ применением некоторого морфизма. Поскольку морфизмы сохраняют рекуррентность, получаем $y \in \mathcal{R}$ и, следовательно, $M(x) \in \mathcal{ER}$. □

3.3. Почти обратимые конечно-автоматные преобразования обобщённо почти периодических последовательностей

Теорема 3.1, п. 2 даёт оценку сверху порядка R_x^{2m} для регулятора выходной последовательности $M(x)$ при применении равномерного преобразователя M с m состояниями к последовательности x. Если воспользоваться общим планом доказательства теоремы 3.5 применительно к случаю обобщённо почти периодических последовательностей, то, по всей видимости, эту оценку можно улучшить. Однако, к сожалению, компактной формулы, удобной для формулировки и изложения доказательства, нам найти не удалось, поэтому полностью этот план для обобщённо почти периодических последовательностей мы здесь реализовывать не будем.

Тем не менее обобщение ключевой части этого плана — теоремы 3.6 — представляет и самостоятельный интерес. Назовём конечно-автоматный преобразователь M *почти обратимым относительно последовательности* x, если каждая буква a, встречающаяся в x бесконечное количество раз, осуществляет взаимно однозначное отображение на множестве состояний преобразователя M.

Предложение 3.11. *Пусть M — почти обратимый относительно последовательности x равномерный конечно-автоматный преобразователь с m состояниями, и $x \in \mathcal{GAP}$. Тогда $M(x) \in \mathcal{GAP}$, причём $\mathrm{R}_{M(x)}(n) \leqslant g^m(n) - 1$, где $g(n) = \mathrm{R}_x(n) + 1$.*

Доказательство. Положим $h(n) = g^m(n) - 1$ для всех n, где $g(n) = \mathrm{R}_x(n) + 1$ для всех n.

Пусть слово v длины n входит в $M(x)$, и $v = M(x)[i, j]$ — одно из таких вхождений при $i \geqslant h(n)$.

Заметим, что вхождения символов, встречающихся в x конечное количество раз, ограничены префиксом длины $\mathrm{R}_x(1)$, поэтому, начиная с позиции $\mathrm{R}_x(1)$, преобразователь M действует на x как обратимый. При этом $i \geqslant h(n) \geqslant \mathrm{R}_x(1)$.

Докажем, что найдётся вхождение $[r, s]$ слова v в $M(x)$ при $i < r \leqslant i + h(n)$.

Итак, пусть $v = M(x)[i, j]$, и $u_1 = x[i, j]$ — прообраз v в x, к которому M подходит, находясь в состоянии q_1. Если бы u_1 входило в x конечное количество раз, то u_1 не могло бы входить в $x[t, \infty)$ для $t \geqslant \mathrm{R}_x(n)$ по определению регулятора. Но $i \geqslant \mathrm{R}_x(n)$, поэтому u_1 входит в x бесконечно много раз. Тогда найдётся вхождение $u_1 = x[i_2, j_2]$, такое что $i_2 > i$, но $j_2 \leqslant i + \mathrm{R}_x(n)$. Если M подходит к i_2 в состоянии q_1, то $M(x)[i_2, j_2] = v$. Иначе M подходит к i_2 в некотором состоянии $q_2 \neq q_1$ (в этом случае $m \geqslant 2$).

Положим $u_2 = x[i, j_2]$. Имеем $|u_2| \leqslant \mathrm{R}_x(n) + 1$. Тогда, поскольку $i \geqslant h(n) \geqslant \mathrm{R}_x(\mathrm{R}_x(n) + 1)$ (при $m \geqslant 2$), u_2 входит в x, бесконечно много раз. Поэтому найдётся вхождение $u_2 = x[i_3, j_3]$, такое что $i_3 > i$, но $j_3 \leqslant i + \mathrm{R}_x(\mathrm{R}_x(n) + 1)$. Если M подходит к $j_3 - n + 1$ в состоянии q_1, то $M(x)[j_3 - j + 1, j_3] = v$, так как u_2 заканчивается словом u_1. Если M подходит к $j_3 - n + 1$ в состоянии q_2, то в силу обратимости на $x[i, \infty)$ преобразователь M подходит к i_2 в состоянии q_1, и тогда $M(x)[i_2, i_2 + n - 1] = v$, так как u_2 начинается с u_1. В худшем случае M подходит к $j_3 - n + 1$ в некотором состоянии q_3, таком что $q_3 \neq q_1$ и $q_3 \neq q_2$ (рис. 3.1).

Рассуждая так дальше, найдём такие $j_2, \ldots, j_{m+1}$, что хотя бы к каким-то двум из позиций $i = j - n + 1$, $j_2 - n + 1$, $\ldots$, $j_{m+1} - n + 1$ преобразователь подходит в одинаковых состояниях. При этом $j_{m+1} \leqslant i + h(n)$.

Таким образом, мы доказали, что если слово длины n входит в $M(x)[h(n), \infty)$, то оно входит в $M(x)$ бесконечно много раз. По лемме 3.2 мы знаем тогда, что это слово входит в каждый отрезок длины $h(n)$ в $M(x)$. Отсюда $M(x) \in \mathcal{GAP}$ и $\mathrm{R}_x(n) \leqslant h(n)$. □

Рассмотрев теперь разбиение $\mathrm{split}_a(x)$ последовательности x для некоторой буквы a, входящей в x бесконечно много раз, можно провести индукцию по количеству состояний преобразователя и, таким образом, другим способом доказать теорему 3.1, п. 1. Как уже было сказано, в удобной форме оценку на регулятор образа лучше той, что получена в теореме 3.1, п. 2, нам получить не удалось, хотя с помощью описанного только что подхода, видимо, такую оценку получить можно, как минимум для некоторых частных случаев обобщённо почти периодических последовательностей.

3.4. Стековые конечно-автоматные преобразования

Интересно понять, что произойдёт, если мы несколько расширим класс рассматриваемых преобразований.

По-видимому, простейшим обобщением можно считать стековый (по-другому, магазинный) конечно-автоматный преобразователь, по аналогии со стековыми (или магазинными) автоматами (pushdown automata), распознающими контекстно-свободные языки (например, см. [32]). *Стековый конечно-автоматный преобразователь* можно описать как конечно-автоматный преобразователь, к которому добавлен стек. (Стек, или магазин, — структура данных, в которой элементы хранятся в фиксированном порядке, и пользователь имеет доступ к последнему добавленному элементу, про который говорят, что он находится в голове стека.) На каждом шаге следующее состояние преобразователя и выходной символ определяются текущим состоянием, символом входной последовательности и символом в голове стека. Кроме того, на каждом шаге преобразователь должен решить, что делать со стеком: добавить символ, удалить символ или оставить стек без изменений. Преобразователь также может проверять стек на пустоту.

Как и следовало ожидать, стековые конечно-автоматные преобразователи не сохраняют свойство обобщённой почти периодичности.

Предложение 3.12. *Существуют стековый конечно-автоматный преобразователь и обобщённо почти периодическая последовательность, которая под действием этого преобразователя переходит в последовательность, не являющуюся обобщённо почти периодической.*

Лемма 3.13. *Существует почти периодическая последовательность над $\mathbb{B}$, у которой для любого n найдётся префикс, в котором нулей на n больше, чем единиц, и для любого n найдётся префикс, в котором единиц на n больше, чем нулей.*

Доказательство. Положим $x = 001 \otimes \bigotimes_{i=1}^{\infty} 0111$. Пусть $u_m = 001 \otimes \bigotimes_{i=1}^{m} 0111$. Докажем индукцией по m, что $|u_m|_0 - |u_m|_1 = (-1)^m 2^m$. Действительно, при $m = 0$ имеем $u_0 = 001$, $|u_0|_0 - |u_1|_1 = 2 - 1 = 1$.

Пусть теперь $|u_m|_0 - |u_m|_1 = (-1)^m 2^m$. По определению $u_{m+1} = u_m \bar{u}_m \bar{u}_m \bar{u}_m$, значит, $|u_{m+1}|_0 - |u_{m+1}|_1 = |u_m|_0 + 3|u_m|_1 - (|u_m|_1 + 3|u_m|_0) = 2(|u_m|_1 - |u_m|_0) = (-1)^{m+1} 2^{m+1}$. □

Доказательство предложения 3.12. По лемме 3.13 существует почти периодическая последовательность в алфавите $\mathbb{B}$, обладающая следующим свойством: для любого $n \in \mathbb{N}$ найдётся такой её префикс, в котором нулей на n больше, чем единиц, а также найдётся такой префикс, в котором единиц на n больше, чем нулей. Применим к этой последовательности следующий стековый конечно-автоматный преобразователь. У него два режима, a (начальный) и b. В режиме a он действует так: увидев в последовательности символ 0, он кладёт его в стек, а увидев 1, убирает 0 из стека. Когда стек опустошается (что соответствует начальному отрезку последовательности с равным количеством 0 и 1), преобразователь переключается в режим b. В режиме b преобразователь действует противоположным образом: при входе 1 кладёт 1 в стек, при входе 0 убирает 1 из стека, при опустошении стека переключается в режим a. В выходную последовательность преобразователь подаёт всегда номер своего режима. Видно тогда, что в выходной последовательности найдутся сколь угодно длинные отрезки подряд идущих a, и значит, она не является обобщённо почти периодической. □

Глава 4.

Вычислимые операторы на бесконечных последовательностях

Во многих вопросах, связанных с почти периодическими и с обобщённо почти периодическими последовательностями, естественным образом возникает алгоритмическая составляющая: можно ли ту или иную характеристику или то или иное свойство проверять алгоритмически, получая на вход последовательность. Основной пример, обсуждавшийся ранее — теорема 3.5, которая является эффективным аналогом теоремы 3.4. В настоящей главе в основном будут рассмотрены результаты, когда ответ на эти вопросы отрицательный. Мы доказываем, что некоторые свойства почти периодичности не имеют эффективных аналогов.

Формально, мы рассматриваем *вычислимые операторы на последовательностях*. Отметим, что тут возможны два подхода. В первом мы рассматриваем операторы, определённые на всевозможных последовательностях. Такие операторы в процессе вычисления могут обращаться к любым элементам последовательности (последовательность задаётся как оракул). Основное свойство вычислимых операторов — непрерывность относительно топологии, индуцированной метрикой d_C. Другими словами, перед тем, как за конечное время выдать ответ, оператор может прочитать лишь конечное количество элементов последовательности. Таким образом, для доказательства неразрешимости некоторого свойства таким оператором достаточно показать, что это свойство не является непрерывным.

При втором подходе рассматриваются операторы, определённые только на вычислимых последовательностях. На вход оператору подаётся конечная запись процедуры вычисления последовательности. Однако можно доказать, что при таком определении оператор также будет обладать свойством непрерывности (см. [36]). Поэтому достаточно иметь в виду только первый подход, а все результаты настоящего раздела верны и при втором подходе.

Почти никакие осмысленные свойства нельзя распознать, имея на входе только саму последовательность. Например, про $x \in \mathbb{B}^{\mathbb{N}}$ нельзя даже сказать, входит ли в неё символ 1: если алгоритм проверил некоторое конечное количество символов, и все они оказались 0, он не может гарантировать, что далее не встретится 1. Вопрос об эффективности станет более интересным, если мы на вход будем подавать какую-то дополнительную информацию. В случае с обобщённо почти периодическими последовательностями в качестве этой информации естественно взять регулятор почти периодичности. (Ясно, что точно так же можно рассматривать операторы, которым на вход подаются и функции, а не только последовательности.) Дополнительной мотивацией к рассмотрению задачи именно в такой формулировке является тот факт, что по теореме 1.6 любое свойство обобщённо почти периодической последовательности, выразимое в монадическом языке, можно вычислить, зная саму последовательность и её регулятор почти периодичности, то есть в регуляторе содержится достаточно много информации о последовательности.

Теперь мы видим, что при таком подходе сформулированная выше задача может быть решена эффективно: прочитав первые $f(1)$ символов последовательности, мы можем понять, входит ли в неё 1, а прочитав следующие $f(1)$ символов, сможем сказать даже, входит ли в неё 1 бесконечно много раз.

4.1. Неразрешимость некоторых свойств почти периодических последовательностей

Следующие несколько теорем — примеры результатов о неэффективности, то есть о несуществовании оператора, вычисляющего некоторое свойство или некоторую функцию на бесконечной последовательности. Все эти результаты довольно естественно ожидаемые. Особенно интересно сопоставить утверждения теоремы 4.1 и теоремы 3.5. Теорема 4.8 также связана с теоремой 3.5. Теоремы 4.1, 4.2, 4.8, п. 1 предложения 4.6

(для класса $\mathcal{AP}$, что в данном случае несущественно) были впервые объявлены в [41] и доказаны в [40]. Общая схема доказательства теоремы 4.3, сформулированная в качестве отдельного доказательства утверждения п. 2 предложения 4.6, изложена в [42], и в [43] приведено полное доказательство.

Под сходимостью $f_n \to f$, где $f_n, f\colon \mathbb{N} \to \mathbb{N}$, будем понимать условие $\forall i\ \exists n\ \forall m > n\ f_m(i) = f(i)$.

Теорема 4.1. *Не существует вычислимого оператора, получив на вход $x \in \mathcal{EAP}$ и $f \geqslant \mathrm{R}_x$, выдающего какое-либо $l \geqslant \mathrm{pr}(x)$.*

Доказательство. Напомним, что $\mathbf{t} = \lim_{n\to\infty} u_n$ — последовательность Туэ — Морса, где последовательность слов u_n определяется по индукции: $u_0 = 0$ и $u_{n+1} = u_n\bar{u}_n$. При этом $|u_n| = 2^n$. Последовательность $\mathbf{t}$ не содержит кубов (см. раздел 2.4). (На самом деле, как будет видно из доказательства, нам достаточно даже более слабого утверждения — отсутствия кубов специального вида.)

Для доказательства теоремы достаточно построить $x_n \in \mathcal{EAP}$, $x \in \mathcal{AP}$ с оценками на регуляторы $f_n \geqslant \mathrm{R}_{x_n}$ и $f \geqslant \mathrm{R}_x$, такие что $x_n \to x$, $f_n \to f$, но $\mathrm{pr}(x_n) \to \infty$.

Действительно, предположим, указанный в условии теоремы вычислимый оператор существует. Пусть, получив на вход $\langle x, f\rangle$, он выдаёт число $l \geqslant 0$ (оно может быть произвольным, так как $x \in \mathcal{AP}$). Во время вычисления l алгоритм вычисления значения оператора прочитал лишь конечное количество символов x и значений f, поэтому существует такое $N > l$, что этот алгоритм не знает $x(k)$ и $f(k)$ для $k > N$ (формально это означает, что на всех входах, различие с которыми только в позициях $k > N$, алгоритм работает так же, как и на $\langle x, f\rangle$). Поскольку $\mathrm{pr}(x_n) \to \infty$, найдётся n, для которого $\mathrm{pr}(x_n) > N$. Ясно тогда, что на входе $\langle x_n, f_n\rangle$ значение оператора такое же, как и на $\langle x, f\rangle$, и значит, выдаёт l, но $\mathrm{pr}(x_n) > N > l$.

Положим $x = \mathbf{t}$, $x_n = u_n u_n u_n x$. Тогда $\mathrm{pr}(x_n) \geqslant 2^n$. Действительно, если $\mathrm{pr}(x_n) < 2^n$, то $u_n u_n x = u_n u_n u_n \bar{u}_n \bar{u}_n u_n \ldots \in \mathcal{AP}$, и значит, $u_n u_n u_n$ входит в $x = \mathbf{t}$ — противоречие с бескубностью $\mathbf{t}$.

Осталось показать, что можно выбрать оценку на регуляторы f_n, f, так что $f_n \to f$. Для этого достаточно найти общий регулятор g для всех x_n — потом мы можем увеличить g и добиться того, чтобы g была общей оценкой на регулятор для всех x_n и для x. Положим $g = 4 \cdot \mathrm{R}_x$. Пусть слово v, $|v| = k$ входит в $x_n = u_n u_n u_n x$ бесконечно много раз. Возьмём отрезок $[i, j]$ последовательности x длины $4 \cdot \mathrm{R}_x(k)$ и покажем, что v в него входит. Если $j \geqslant 3 \cdot 2^n + \mathrm{R}_x(k)$, то v встречается на отрезке $x[3 \cdot 2^n, 3 \cdot 2^n + \mathrm{R}_x(k)]$ (по определению регулятора). Иначе $j < 3 \cdot 2^n + \mathrm{R}_x(k)$, откуда $i \leqslant 3 \cdot 2^n - 3\,\mathrm{R}_x(k)$. Но $i \geqslant 0$, значит, в этом случае $\mathrm{R}_x(k) \leqslant 2^n = |u_n|$. Таким образом, $x_n[i, i + \mathrm{R}_x(k)]$ целиком содержится в $u_n u_n$. Но $u_n u_n$ входит в x, и значит, $x_n[i, i + \mathrm{R}_x(k)]$ входит в x, поэтому v входит в $x[i, i + \mathrm{R}_x(k)]$, так как v входит в x бесконечно много раз, а значит, входит в любое подслово x длины $\mathrm{R}_x(k)$.

Однако g — ещё не искомая. Необходимо проследить за словами, которые встречаются в x_n конечное количество раз. Но ясно, что если какое-то v входит в x_n конечное количество раз, то $|v| = k > 2^n$ (иначе v входит в блок из двух последовательных слов u_n или $\bar{u}_n$, а значит, и в x), то есть такое может быть лишь для конечного множества различных n. Поэтому, рассмотрев все ситуации, когда слова фиксированной длины k входят в какие-то x_n конечное количество раз, нам, возможно, придётся увеличить значение $g(k)$, но лишь конечное количество раз. Значит, искомая оценка регуляторов найдётся. □

Как мы уже отмечали в разделе 2.3, имеем $\mathcal{EAP} \subsetneq \mathcal{GAP}$ (предложение 2.4). Можно показать, что эти классы невозможно разделять эффективно.

Теорема 4.2. *Не существует вычислимого оператора, по $x \in \mathcal{GAP}$ и $f \geqslant \mathrm{R}_x$ определяющего, верно ли, что $x \in \mathcal{EAP}$.*

Доказательство. Для доказательства достаточно построить такие $x_n \in \mathcal{EAP}$, $x \in \mathcal{GAP} \setminus \mathcal{EAP}$ и общую оценку на регулятор f для них, что $x_n \to x$.

Положим $a_0 = 1$, $a_1 = 10011$, и далее по правилу: $a_{n+1} = a_n \bar{a}_n \bar{a}_n a_n a_n$. Обозначим слово $a_n a_n a_n a_n$ через c_n. Введём для удобства обозначение $l_n = 5^n - 1 = |c_0 c_1 \ldots c_{n-1}|$. Рассмотрим $x = c_0 c_1 c_2 c_3 \ldots$ и $y = \lim_{n\to\infty} a_n$. Как было доказано в предложении 2.4, $x \in \mathcal{GAP} \setminus \mathcal{EAP}$. Пусть $x_n = c_0 c_1 \ldots c_n y$. Поскольку $y \in \mathcal{AP}$ (так как порождено схемой $\langle \{a_n, \bar{a}_n\}, 5^n \rangle$), то $x_n \in \mathcal{EAP}$. Ясно, что $x_n \to x$, и осталось выбрать общую для всех x_n оценку f на регулятор. Мы получим конечное количество требований вида $f(k) \geqslant \alpha$, после чего в качестве $f(k)$ можем взять максимум по всем таким α.

Пусть $v = x_n[i,j]$, $|v| = k$ входит в $x_n = c_0c_1\ldots c_ny$ бесконечно много раз. Тогда v входит в y, а значит, и в некоторое a_m. Поэтому v входит в x бесконечно много раз, и достаточно взять $f(k) \geqslant \mathrm{R}_x(k) + \mathrm{R}_y(k)$.

Пусть $v = x_n[i,j]$, $|v| = k$ входит в x_n конечное количество раз. Ясно тогда, что $i < l_n$. Если $j > l_n$, то тогда $k > 5^n$, так как иначе v содержится в некотором a_m, а значит, входит в y бесконечно много раз. Но неравенство $k > 5^n$ выполнимо только для конечного количества значений n, что даёт лишь конечное количество условий на $f(k)$. Пусть теперь $j \leqslant l_n$. Но тогда v содержится в $c_0c_1\ldots c_n$, причём входит в x конечное количество раз (иначе бы v содержалось в каком-то a_m), поэтому в этом случае подойдёт $f(k) \geqslant \mathrm{R}_x(k)$. □

Следующий довольно общий результат был изначально мотивирован вопросом, заданным О. Купферман, о регулярности множества автоматных последовательностей — см. предложение 4.6, п. 2 и следствие 4.11, п. 2, доказанные в [42].

Теорема 4.3. *Пусть $\mathcal{M}$ — счётное множество последовательностей, такое что $\mathcal{P} \subseteq \mathcal{M}$. Тогда не существует вычислимого оператора, по $x \in \mathcal{GAP}$ и $f \geqslant \mathrm{R}_x$ определяющего, верно ли, что $x \in \mathcal{M}$.*

Доказательство. Достаточно построить последовательность x_n бесконечных слов и бесконечное слово x, так что x_n стремится к x, все слова x_n, x обобщённо почти периодичны и имеют общую оценку на регулятор почти периодичности $g \geqslant \mathrm{R}_{x_n}$, $g \geqslant \mathrm{R}_x$, причём все $x_n \in \mathcal{M}$, а $x \notin \mathcal{M}$. Действительно (повторим основную схему рассуждения ещё раз), предположим, вычислимый оператор из условия теоремы (распознающий принадлежность $\mathcal{M}$ обобщённо почти периодических последовательностей) существует. Подадим вычисляющему его алгоритму на вход x и g. Он выдаст отрицательный ответ, так как $x \notin \mathcal{M}$. Во время своей работы, перед тем как выдать ответ, алгоритм прочитал лишь какое-то конечное количество символов из x. Последовательность x_n стремится к x, значит, те же самые символы, которые алгоритм успел прочитать в x, стоят на тех же местах в каком-то x_m для некоторого m. Значит, тот же самый отрицательный ответ алгоритм должен выдать и на x_m и g — противоречие, так как $x_m \in \mathcal{M}$.

Построим теперь нужные x_n, x.

В разделе 2.5 мы описали универсальный метод построения почти периодических последовательностей. Напомним один из его вариантов.

Последовательность $\langle B_n, l_n\rangle$, где B_n — непустое множество непустых слов в фиксированном конечном алфавите A, l_n — натуральные числа, является A-$\mathcal{AP}$-схемой, если для неё выполнено для любого $n \in \mathbb{N}$:
(1) все слова из B_n имеют длину l_n;
(2) любое слово $u \in B_{n+1}$ представимо в виде $u = v_1v_2\dots v_k$, где $v_i \in B_n$, и для любого $w \in B_n$ существует i, такое что $v_i = w$. Последовательность x $\mathcal{AP}$-порождена A-$\mathcal{AP}$-схемой, если для любых i и n имеем $x[il_n, (i+1)l_n - 1] \in B_n$ (далее приставки $\mathcal{AP}$- и A-$\mathcal{AP}$- мы при необходимости опускаем).

Как мы уже говорили, каждая схема порождает какую-то последовательность, и каждая последовательность, порождённая схемой, является почти периодической. Более того, каждая почти периодическая последовательность порождается некоторой схемой.

Теперь мы усилим главное условие на B_n, а именно, будем рассматривать схемы, для которых верно:

$(*)$ для любого $n > 1$ каждое $u \in B_n$ имеет вид $u = v_1v_2\dots v_k$, где $v_i \in B_{n-1}$, причём для каждых $w_1, w_2 \in B_n$ найдётся $i < k$, такое что $v_iv_{i+1} = w_1w_2$.

Отметим, что так мы теряем свойство универсальности: очевидно, не каждая почти периодическая последовательность может быть порождена схемой, удовлетворяющей условию $(*)$.

Лемма 4.4. *Существует схема, для которой выполнено условие $(*)$, и которая порождает континуум различных последовательностей.*

Доказательство. Утверждение леммы почти очевидно, но тем не менее докажем его и построим необходимую $\mathbb{B}$-$\mathcal{AP}$-схему.

Определим рекурсивно схему $\langle l_n, B_n\rangle$: положим $l_0 = 1$ и $B_0 = \{0, 1\}$, и дальше для каждого $n \geqslant 1$ положим $l_n = (\#B_{n-1})^2 l_{n-1}$. Также положим B_n состоящим из всех слов длины l_n, которые удовлетворяют условию $(*)$.

Теперь объясним, почему эта схема порождает континуум различных последовательностей. Начнём строить какую-нибудь последовательность, порождаемую этой схемой. Для этого выберем последовательность слов w_i рекурсивно следующим образом: слово w_0 из множества B_0 выберем произвольно, а далее слово w_i для $i \geqslant 1$ будем выбирать произвольно из множества B_i, так чтобы оно являлось продолжением слова w_{i-1}. Заметим, что длина l_i выбрана достаточной для того, чтобы для любого $v \in B_{i-1}$ можно было найти более одного слова из B_i, являющегося продолжением v. Это значит, что каждый раз, когда мы выбираем w_i для какого-нибудь i, у нас есть хотя бы два различных варианта, независимо от всех предыдущих выборов. Значит, мы можем выбрать последовательность слов w_i континуумом различных способов. Осталось только заметить, что при любом таком выборе предельное бесконечное слово $\lim_{i\to\infty} w_i$ порождено схемой $\langle l_n, B_n\rangle$ по построению. □

Как следует из леммы 4.4, среди порождённых такими усиленными схемами найдутся последовательности, не принадлежащие $\mathcal{M}$, так как $\mathcal{M}$ счётно (это единственное место, где используется счётность $\mathcal{M}$). Возьмём одну из них — x, порождённую некоторой схемой $\langle B_n, l_n\rangle$.

Определим $p_n = x[0, l_n]$. Таким образом, $p_n \in B_n$ и $\lim_{n\to\infty} p_n = x$. Положим $x_n = p_n p_n p_n \ldots \in \mathcal{P}$. Ясно, что $x_n \to x$, кроме того, все $x_n \in \mathcal{M}$. Доказательство теоремы завершается доказательством следующей леммы. Она доказывает существование общего регулятора f для всех x_n. Увеличив его потом, если нужно, сделаем так, чтобы он подходил и для x.

Лемма 4.5. *Пусть x — последовательность, порождённая схемой $\langle B_n, l_n\rangle$, удовлетворяющей условию $(*)$. Определим $p_n = x[0, l_n]$ и положим $x_n = p_n p_n p_n \ldots \in \mathcal{P}$. Тогда существует общий регулятор почти периодичности f для всех последовательностей x_n.*

Доказательство. Сейчас мы получим конечное (для каждого k) количество требований вида $f(k) \geqslant \alpha$, после чего в качестве $f(k)$ можем взять максимум по всем таким α.

Пусть $v = x_n[i, j]$, $|v| = k$ (это сразу означает, что v входит в x_n бесконечно много раз, так как $x_n \in \mathcal{P}$). Неравенство $k \geqslant l_{n-1}$ может выполняться лишь для конечного количества различных n (при фиксированном k), что даёт лишь конечное количество условий на $f(k)$. Теперь можно считать, что $k < l_{n-1}$. Возьмём такое t, что $l_{t-1} < k \leqslant l_t$ (важно, что t не зависит от n и однозначно определяется по k). Тогда $t \leqslant n-1$. Существует m, такое что $ml_t \leqslant i$ и $j \leqslant (m+2)l_t - 1$, то есть v содержится в некотором ab, где $a, b \in B_t$, и значит, благодаря свойству $(*)$, v входит в любое $c \in B_{t+1}$. Но на каждом отрезке длины $2l_{t+1}$ последовательности x_n найдётся вхождение какого-нибудь $c \in B_{t+1}$ (целиком входящего в какое-то p_n), а значит, и вхождение v. Таким образом, достаточно $f \geqslant 2l_{t+1}$. □

□

Простым следствием из теоремы 4.3 является следующее предложение.

Предложение 4.6. *1) Не существует вычислимого оператора, по $x \in \mathcal{GAP}$ и $f \geqslant \mathrm{R}_x$ определяющего, верно ли, что $x \in \mathcal{P}$.*

2) Не существует вычислимого оператора, по $x \in \mathcal{GAP}$ и $f \geqslant \mathrm{R}_x$ определяющего, верно ли, что x автоматна.

3) Не существует вычислимого оператора, по $x \in \mathcal{GAP}$ и $f \geqslant \mathrm{R}_x$ определяющего, верно ли, что x морфическая.

Заметим также, что попутно мы доказали следующий факт (следует из лемм 4.4 и 4.5).

Предложение 4.7. *Существует бесконечное количество периодических последовательностей с общим регулятором почти периодичности.*

Эта серия примеров любопытна тем, что одна естественная мера повторяемости в этих последовательностях — оценка на регулятор почти периодичности — постоянна, тогда как другая такая мера — период последовательности — может быть сколь угодно большой. По-другому, это замечание можно сформулировать так: если естественным образом обобщить понятие регулятора почти периодичности на конечные слова и взять последовательность слов p_n — минимальных периодов x_n (в обозначениях леммы 4.5), то это будет пример сколь угодно длинных непериодических конечных слов (то есть имеющих минимальную длину периода, равную длине слова) с общей оценкой сверху на регулятор почти периодичности.

Следующее утверждение говорит о том, что даже приписав к почти периодической последовательности один символ, мы, вообще говоря, уже не можем проверять, является ли она по-прежнему почти периодической.

Теорема 4.8. *Не существует вычислимого оператора, по $x \in \mathcal{EAP}$, $f \geqslant \mathrm{R}_x$ и некоторому $l \geqslant \mathrm{pr}(x)$ находящего $\mathrm{pr}(x)$.*

Лемма 4.9. *Если $ax \in \mathcal{AP}$, где $a \in A^*$, и x периодическая с периодом l, то ax тоже периодическая с периодом l.*

Доказательство. Достаточно доказать лемму для однобуквенного слова a. Пусть $\alpha = 012\ldots(l-1)012\ldots(l-1)012\ldots(l-1)\ldots$ — периодическая последовательность, составленная из букв алфавита $A_l = \{0, 1, 2, \ldots, l-1\}$. Тогда по предложению 2.9 получаем, что $ax \times \alpha \in \mathcal{AP}$. В этой последовательности символ $\langle a, 0\rangle$ встречается бесконечно много раз, откуда ясно, что $a = x(l)$, что и требовалось. $\square$

Доказательство теоремы 4.8. Для доказательства достаточно построить такие $x_n \in \mathcal{EAP}$, $x \in \mathcal{AP}$ и общую для всех x_n оценку f на регулятор, что $x_n \to x$ и $\mathrm{pr}(x_n) = 1$ (из $x \in \mathcal{AP}$ следует $\mathrm{pr}(x) = 0$).

Заметим, что $1\mathbf{t} \in \mathcal{AP}$. Для этого достаточно проверить, что любой префикс этой последовательности входит в неё бесконечно много раз. Но действительно, для любого n слова $u_n u_n$ и $\bar{u}_n u_n$ входят в $\mathbf{t}$, а значит, и $1u_n$ тоже. Аналогично показывается, что $0\mathbf{t} \in \mathcal{AP}$.

Заметим, что схема $\langle\{u_n, \bar{u}_n\}, 2^n\rangle$, удовлетворяющая свойству $(*)$ (см. с. 61), порождает последовательность $\mathbf{t}$. Отсюда и из леммы 4.5 следует, что можно выбрать такую последовательность $k_n \to \infty$, что все периодические последовательности вида $\mathbf{t}(0)\ldots\mathbf{t}(k_n)\mathbf{t}(0)\ldots\mathbf{t}(k_n)\mathbf{t}(0)\ldots$ имеют общий регулятор f. Выберем такую подпоследовательность m_n последовательности k_n, что все символы $\mathbf{t}(m_n)$ одинаковы. Предположим, что они равны 0.

Положим $x_n = 1x(0)\ldots x(m_n)x(0)\ldots x(m_n)x(0)\ldots$, $x = 1\mathbf{t}$. Ясно, что можно выбрать для них общую оценку сверху g на регулятор. Для этого достаточно, чтобы было выполнено $g(k) \geqslant f(k) + 1$ (исходя из рассмотрения слов, входящих бесконечное количество раз) и $g(k) \geqslant k$ (исходя из рассмотрения слов, входящих лишь конечное количество раз, что возможно только для префиксов, входящих ровно один раз).

Если бы $x_n \in \mathcal{AP}$, то по лемме 4.9 последовательность x_n периодична с периодом m_n, но мы специально подобрали $x_n(0) = 1 \neq x_n(m_n) = 0$. Таким образом, $\mathrm{pr}(x_n) = 1$.

Случай, когда все символы $x(m_n)$ равны 1, разбирается аналогично (тогда x_n и x начинаются с 0). $\square$

4.2. Нерегулярность некоторых множеств последовательностей

Теперь мы покажем, как, используя результаты этой главы, можно доказать про некоторые множества последовательностей, что они не распознаются конечным автоматом (имеется в виду автомат Бюхи или автомат Мюллера), то есть не являются регулярными.

Как утверждает теорема 3.3, по обобщённо почти периодической последовательности, какой-нибудь оценке сверху на её регулятор и автомату Бюхи можно алгоритмически определять, принимает ли данный автомат Бюхи данную последовательность. Говоря более точно, существует алгоритм, который, получив на вход оракул для последовательности $x \in \mathcal{GAP}$, оракул для некоторой функции $g \geqslant \mathrm{R}_x$ и автомат Бюхи F, выдаёт ответ, верно ли, что F принимает x.

Предположим, что некоторое множество $\mathcal{M}$ последовательностей является регулярным, то есть существует принимающий его автомат Бюхи. Тогда по теореме 3.3 существует алгоритм, который по обобщённо почти периодической последовательности и оценке сверху на её регулятор определяет, принадлежит ли эта последовательность множеству $\mathcal{M}$.

Таким образом, для того чтобы показать, что множество $\mathcal{M}$ нерегулярно, достаточно показать, что не существует алгоритма, который по оракулу для обобщённо почти периодической последовательности и для какой-то оценки сверху на её регулятор определяет принадлежность этой последовательности к множеству $\mathcal{M}$. (Обобщённо почти периодические последовательности в этом утверждении можно заменить на почти периодические.)

Так, из теоремы 4.2 и предложения 4.6 можно получить следующие результаты.

Следствие 4.10. *Множество $\mathcal{EAP}$ не регулярно.*

Следствие 4.11. *1) Множество $\mathcal{P}$ не регулярно.*

2) Множество автоматных последовательностей не регулярно.

3) Множество морфических последовательностей не регулярно.

Глава 5.

Свойства автоматных и морфических последовательностей

В этой главе мы изучаем, как соотносятся два класса последовательностей — почти периодические последовательности и морфические последовательности.

Основной вопрос главы — вопрос о разрешимости почти периодичности для морфических последовательностей. В полной общности вопрос остаётся открытым. В разделе 5.1 мы делаем некоторые предварительные замечания и доказываем некоторые вспомогательные и предварительные результаты. В разделе 5.2 мы приводим критерий почти периодичности для чисто морфических последовательностей (вопрос о таком критерии был поставлен в [4], раздел 10.12, проблема 5) и доказываем, что существует полиномиальный по времени алгоритм, определяющий по чисто морфической последовательности, является ли она почти периодической. В разделе 5.3 мы доказываем, что существует полиномиальный по времени алгоритм, определяющий по автоматной последовательности, является ли она почти периодической.

В разделе 5.4 мы рассматриваем вопрос о подсловной сложности последовательностей. Мы доказываем, что подсловная сложность последовательности, являющейся одновременно морфической и почти периодической, растёт не более чем линейно.

Результаты настоящей главы получены в [45, 48]. Все результаты получены автором, за одним исключением: изначально теоремы 5.3 и 5.6 были доказаны в [45] для чисто морфических последовательностей, порождённых нестирающими морфизмами. После этого в [48] первым автором было предложено, как можно обобщить эти результаты на случай произвольного морфизма. Для полноты изложения мы приводим эти теоремы сразу в полной общности.

5.1. Разрешимость почти периодичности для морфических последовательностей: предварительные результаты

Основной вопрос этой главы заключается в алгоритмической разрешимости следующей проблемы:

Вход: Два алфавита A и B, буква $s \in A$, морфизм $\phi\colon A^* \to A^*$, продолжаемый на s, и кодирование $h\colon A \to B$.

Вопрос: Является ли последовательность $h(\phi^\infty(s))$ почти периодической?

Вопрос о существовании такого алгоритма остаётся открытым, однако ниже мы отвечаем на него положительно в двух частных случаях. Перед этим мы дадим некоторые вспомогательные определения и докажем предварительные утверждения.

Для последовательности $\phi^\infty(s)$ легко можно найти, используя граф G_ϕ, какие символы алфавита A действительно входят в эту последовательность. Действительно, это в точности все те символы, к которым существует путь по рёбрам этого графа, начинающийся в s. Так что теперь без ограничения общности будем считать, что все символы алфавита A входят в $\phi^\infty(s)$.

Заметим также, что проверка правильности формата входа алгоритма требует проверки того, верно ли, что символ s является ϕ-возрастающим: это действительно можно сделать, как следует из леммы 5.5 ниже.

Сформулируем следующий критерий, который нам понадобится в дальнейшем.

Предложение 5.1. *Последовательность $\phi^\infty(s)$ почти периодическая тогда и только тогда, когда s входит в неё бесконечно много раз с ограниченными интервалами между соседними вхождениями.*

Доказательство. В одну сторону утверждение предложения прямо следует из определения.

Предположим теперь, что s входит в $\phi^\infty(s)$ бесконечно много раз с ограниченными интервалами. Тогда для любого m слово $\phi^m(s)$ также входит в $\phi^\infty(s)$ бесконечно много раз с ограниченными интервалами. Но любое слово u, входящее в $\phi^\infty(s)$, входит в некоторый префикс $\phi^m(s)$ и, следовательно, входит в последовательность бесконечно много раз с ограниченными интервалами между соседними вхождениями. $\square$

Морфизм неприводимый, если и только если его граф инцидентности строго связен. Для продолжаемых морфизмов это является также критерием примитивности. Поэтому следующее предложение является проверяемым за полиномиальное время критерием почти периодичности для возрастающих морфизмов.

Предложение 5.2. *Пусть A — алфавит, $s \in A$ и $\phi\colon A^* \to A^*$ — возрастающий морфизм, продолжаемый на s. Тогда последовательность $\phi^\infty(s)$ почти периодическая, если и только если морфизм ϕ примитивный.*

Доказательство. Предположим, ϕ — примитивный. Пусть n такое, что s входит в $\phi^n(a)$ для любого $a \in A$. Обозначим через l максимальную длину $\phi^n(a)$ для $a \in A$. Каждое подслово длины $2l$ последовательности $\phi^\infty(s)$ содержит хотя бы одно вхождение s. Следовательно, по предложению 5.1 последовательность $\phi^\infty(s)$ почти периодическая.

Обратно, пусть ϕ не примитивный. Существует буква $b \in A$, такая что для любого $n \geq 0$ слово $\phi^n(b)$ не содержит s. Для любого $n \geq 0$ слово $\phi^n(b)$ входит в $\phi^\infty(s)$. Поскольку $|\phi^n(b)| \to \infty$ при $n \to \infty$, последовательность $\phi^\infty(s)$ не почти периодическая. $\square$

Однако при обобщении даже на нестирающие морфизмы (то есть если разрешить образам букв иметь длину не только не менее 2, но и 1), уже невозможно по графу и даже матрице инцидентности морфизма определить, является ли соответствующая последовательность почти периодической. Это видно из следующего примера.

Пусть ϕ_1 — следующий морфизм: $0 \to 01$, $1 \to 120$, $2 \to 2$, и пусть ϕ_2 такой: $0 \to 01$, $1 \to 210$, $2 \to 2$. Эти морфизмы, продолжаемые на 0, порождают последовательности

$$\phi_1^\infty(0) = 011201202011202012011 20\ldots$$

и

$$\phi_2^\infty(0) = 01210221001222100101210\dots$$

Матрицы инцидентности этих морфизмов совпадают, но $\phi_1^\infty(0)$ почти периодическая, а $\phi_2^\infty(0)$ — нет. Действительно, в последовательность $\phi_2^\infty(0)$ входят сколь угодно длинные отрезки вида $222\dots22$, так что $\phi_2^\infty(0)$ не почти периодическая. Такой проблемы не возникает в $\phi_1^\infty(0)$. Поскольку 0 входит и в $\phi_1(0)$, и в $\phi_1(1)$, и слово 22 (как несложно показать) не входит в $\phi_1^\infty(0)$, то 0 входит в $\phi_1^\infty(0)$ с ограниченными интервалами. Отсюда $\phi_1^\infty(0)$ почти периодична по предложению 5.1. В теореме 5.3 мы формулируем общий критерий почти периодичности для чисто морфических последовательностей.

5.2. Разрешимость почти периодичности для чисто морфических последовательностей

В этом разделе мы приводим критерий почти периодичности для чисто морфических последовательностей в теореме 5.3, а также более эффективный вариант критерия в теореме 5.6, который можно проверить за полиномиальное время.

Итак, пусть $\phi\colon A^* \to A^*$ — некоторый морфизм, продолжаемый на некотором символе $s \in A$. Мы хотим получить критерий почти периодичности последовательности $\phi^\infty(s)$ и доказать, что его можно проверить за полиномиальное время. Говоря о полиномиальном времени работы здесь, мы имеем в виду время, полиномиальное от n и k, где n — размер алфавита A и $k = \max\{|\phi(a)| : a \in A\}$, поскольку ясно, что морфическую последовательность можно описать словом длины, полиномиальной от n и k.

Теорема 5.3. *Пусть A — алфавит, $s \in A$, и $\phi\colon A^* \to A^*$ — морфизм, продолжаемый на s. Чисто морфическая последовательность $\phi^\infty(s)$ почти периодична тогда и только тогда, когда выполнены следующие два условия:*

1) *для любой ϕ-возрастающей буквы a, входящей в $\phi^\infty(s)$, существует число $n \in \mathbb{N}$, такое что s входит в $\phi^n(a)$, и*

2) *лишь конечное множество различных ϕ-ограниченных слов входит в $\phi^\infty(s)$.*

Доказательство. $\Rightarrow$. Предположим, что $\phi^\infty(s)$ почти периодическая. Тогда существует натуральное число l, такое что s входит в любое подслово длины l последовательности $\phi^\infty(s)$.

1) Пусть a — ϕ-возрастающая буква, входящая в $\phi^\infty(s)$. Для любого $n \in \mathbb{N}$ слово $\phi^n(a)$ является подсловом $\phi^\infty(s)$. Если n достаточно велико, слово $\phi^n(a)$ имеет длину хотя бы l. Следовательно, s входит в $\phi^n(a)$ для всех достаточно больших n.

2) Поскольку s — ϕ-возрастающая, s не может входить ни в какое ϕ-ограниченное подслово последовательности $\phi^\infty(s)$. Следовательно, все ϕ-ограниченные подслова $\phi^\infty(s)$ имеют длину не более l.

$\Leftarrow$. Предположим, что выполняются условия 1) и 2). Из условия 1) следует, что существует положительное n, такое что для любой ϕ-возрастающей буквы a, входящей в $\phi^\infty(s)$, s входит в $\phi^n(a)$. Из условия 2), существует положительное M, такое что любое ϕ-ограниченное подслово последовательности $\phi^\infty(s)$ имеет длину не больше M. Обозначим через K максимальную длину слов $\phi^n(a)$ для $a \in A$.

Пусть w — подслово длины $KM + K$ последовательности $\phi^\infty(s)$. Существует подслово v длины M в $\phi^\infty(s)$, такое что $\phi^n(v)$ является подсловом w. Поскольку v длиннее, чем любое ϕ-ограниченное подслово $\phi^\infty(s)$, какая-то ϕ-возрастающая буква a входит в v. Следовательно, s входит в $\phi^n(a)$, $\phi^n(a)$ — подслово $\phi^n(v)$, и $\phi^n(v)$ — подслово w. Отсюда s входит в w.

Таким образом, мы показали, что s входит в любое подслово длины $KM+K$ последовательности $\phi^\infty(s)$. Поэтому $\phi^\infty(s)$ почти периодическая по предложению 5.1. $\square$

Теперь объясним, как получить полиномиальный по времени критерий. Сначала в предложении 5.4 мы приводим различные переформулировки условия 2) из теоремы 5.3. Потом мы переформулируем критерий почти периодичности, так чтобы его легко было проверять за полиномиальное время.

Предложение 5.4 ([13]). *Пусть A — алфавит, $s \in A$, и $\phi\colon A^* \to A^*$ — морфизм, продолжаемый на s. Следующие три условия эквивалентны.*

1) *В последовательности $\phi^\infty(s)$ есть бесконечно много различных ϕ-ограниченных подслов.*

2) Существует число n, буква a, входящая в $\phi^\infty(s)$, и два слова u, $v \in A^*$, удовлетворяющие следующим условиям:

- *u — не ϕ-стираемое,*
- *u — ϕ-ограниченное и*
- *$\phi^n(a) = uav$ или $\phi^n(a) = vau$.*

3) Существует непустое ϕ-ограниченное слово w, такое что w^n — подслово $\phi^\infty(s)$ для каждого $n \in \mathbb{N}$.

Теперь покажем, что по морфизму $\phi\colon A^* \to A^*$ можно эффективно находить множества I_ϕ, B_ϕ, E_ϕ.

Лемма 5.5. *Множества I_ϕ, B_ϕ и E_ϕ вычислимы по алфавиту A и морфизму $\phi\colon A^* \to A^*$ за время $\mathrm{poly}(n, k)$, где $n = |A|$ и $k = \max_{b \in A} |\phi(b)|$.*

Доказательство. Рассмотрим отношение эквивалентности "$\equiv$" на вершинах графа G_ϕ: $a \equiv b$, если и только если из a можно дойти в b по рёбрам графа и наоборот, из b можно дойти в a. Ясно, что если $a \equiv b$, то $|\phi^m(a)| = \Theta(|\phi^m(b)|)$ при $m \to \infty$. Построим новый граф H_ϕ: множеством его вершин будет множество классов эквивалентности отношения "$\equiv$". Ребро идёт из вершины C в вершину D в H_ϕ, если $\exists a \in C$ $\exists b \in D$, так что $\phi(a)$ содержит b. Для каждой вершины C графа H_ϕ определим число $\kappa_C = \max\{\text{количество вхождений символов из } C \text{ в } \phi(a) : a \in C\}$. Определим $S_i = \{D \text{ в } H_\phi : \max\{\kappa_C : \text{из } D \text{ можно дойти по рёбрам в } C\} = i\}$. Ясно, что H_ϕ, все κ_C и S_i можно посчитать за полиномиальное время.

Несложно видеть, что $\forall i \geq 2$ $\forall C \in S_i$ $\forall a \in C$ $|\phi^m(a)|$ растёт экспоненциально при $m \to \infty$, и поэтому $a \in I_\phi$. Кроме того, множество S_0 состоит из одноэлементных подмножеств E_ϕ.

Для каждой строго связной компоненты C графа G_ϕ имеем $\kappa_C = 1$, если и только если подграф G_ϕ, индуцированный C, — ориентированный цикл.

Теперь рассмотрим подграф графа H_ϕ, индуцированный S_1. Ясно, что $\forall C \in S_1$ $\forall a \in C$ $a \notin E_\phi$. Пусть $S_1 = U \cup V$, где $U = \{C \in S_1 : \kappa_C = 0\}$, $V = \{C \in S_1 : \kappa_C = 1\}$. Далее, пусть $V = X \cup Y$, где $X = \{C \in V : \text{из } C \text{ можно по рёбрам дойти до некоторой другой вершины } D \in V\}$, $Y = V \setminus X$. Несложно видеть, что $\bigcup_{C \in X} C \subseteq I_\phi$, $\bigcup_{C \in Y} C \subseteq B_\phi$. Далее, заметим, что любое множество $C \in U$ одноэлементно. Если в $D \in X$ можно по рёбрам прийти из $\{a\} \in U$, то $a \in I_\phi$, иначе $a \in B_\phi$.

Ясно, что всё вышеописанное можно проверить за полиномиальное время. □

Пусть $\phi\colon A^* \to A^*$ — морфизм. Как мы помним, слово ϕ-стираемо, если и только если оно состоит из ϕ-стираемых символов. Таким образом, по слову легко определить, является ли оно ϕ-стираемым.

Построим размеченный *префиксный граф* L_ϕ. Пусть I_ϕ — его множество вершин. Из каждой вершины b выходит ровно одно ребро. Чтобы его построить, найдём представление $\phi(b) = ucv$, где $c \in I_\phi$, u — максимальный префикс слова $\phi(b)$, содержащий только символы из множества B_ϕ. Из определения I_ϕ и B_ϕ следует, что u не совпадает с $\phi(b)$, поэтому такое представление корректно. Теперь построим в графе L_ϕ ребро, идущее из b в c, и отметим на этом ребре слово u.

Аналогично строится *суффиксный граф* R_ϕ. В этом случае мы находим представление $\phi(b) = vcu$, где $u \in B_\phi^*$, $c \in I_\phi$, и отмечаем слово u на ребре из b в c.

Теперь мы сформулируем конструктивную версию критерия из теоремы 5.3.

Теорема 5.6. *Пусть A — алфавит, $s \in A$ и $\phi\colon A^* \to A^*$ — морфизм, продолжаемый на s. Последовательность $\phi^\infty(s)$ почти периодическая, если и только если выполнены следующие два условия:*

1) *граф G_ϕ, ограниченный на множество I_ϕ, строго связен, и*

2) *в каждом из графов L_ϕ и R_ϕ каждое ребро каждого цикла отмечено ϕ-стираемым словом.*

Напомним, что мы предполагаем, что все символы из алфавита A входят в $\phi^\infty(s)$.

Доказательство. Несложно видеть, что условие 1) этой теоремы эквивалентно условию 1) теоремы 5.3. Предложение 5.4 служит объяснением того, почему то же верно и для условий 2). □

Рассмотрим примеры с ϕ_1 и ϕ_2, приведённые выше. Для $\phi \in \{\phi_1, \phi_2\}$ имеем: $I_\phi = \{0, 1\}$ и $B_\phi = \{2\}$. На каждом ребре графа R_ϕ в обоих случаях написано пустое слово. Это не так для L_ϕ: на ребре, идущем из 1 в 1, написано пустое слово для ϕ_1 и слово 2 для ϕ_2. Слово 2 не является стираемым, поскольку его образ равен 2, так что $\phi_1^\infty(0)$ почти периодична, но $\phi_2^\infty(0)$ не почти периодична в соответствии с критерием.

Теорема 5.7. *Существует алгоритм, работающий время* $\mathrm{poly}(n,k)$*, который по последовательности* $\phi^\infty(s)$ *определяет, является ли она почти периодической.*

Доказательство. Несложно видеть, благодаря лемме 5.5, что условия из теоремы 5.6 можно проверить за полиномиальное время. □

Также представляется полезным сформулировать в явном виде критерий для случая алфавита из двух символов.

Следствие 5.8. *Пусть* $\phi\colon \{0,1\}^* \to \{0,1\}^*$ *— морфизм, продолжаемый на 0. Тогда последовательность* $\phi^\infty(0)$ *почти периодическая тогда и только тогда, когда выполнено одно из следующих условий:*

1) $\phi(0)$ *содержит только символы 0, но не 1;*

2) $\phi(1)$ *содержит 0;*

3) $\phi(1) = \Lambda$*;*

4) $\phi(1) = 1$ *и* $\phi(0) = 0u0$ *для некоторого слова u.*

Пример последовательности, порождённой морфизмом ϕ, таким что $\phi(0) = 0010$, $\phi(1) = 1$, служит наглядной нетривиальной иллюстрацией к случаю 4 следствия 5.8: ϕ не примитивный, $\phi^\infty(0)$ апериодическая, но почти периодическая.

5.3. Разрешимость почти периодичности для автоматных последовательностей

Заметим, что классы автоматных и почти периодических последовательностей находятся в общем положении. Ясно, например, что существуют почти периодические последовательности, не являющиеся автоматными — почти периодических последовательностей континуум, в то время как автоматных, очевидно, счётное количество. И наоборот, существуют автоматные последовательности, не являющиеся почти периодическими. Действительно, последовательность 0110100010000001000…, у которой на каждом 2^n-м месте стоит 1, а на остальных 0, 2-автоматна (легко построить автомат, отделяющий числа вида $100\ldots0$ в двоичной системе счисления от всех остальных), но не является почти периодической.

Поэтому естественно задать вопрос о разрешимости свойства почти периодичности для автоматных последовательностей. Оказывается, на этот вопрос можно ответить положительно. Доказательству этого результата посвящён настоящий раздел. Мы показываем, что существует алгоритм, определяющий по автоматной последовательности, является ли она почти периодической. Как именно точно описать последовательность на входе алгоритма, не очень важно, достаточно понимать, что можно описать последовательность словом длины, полиномиальной от n и k, где n — размер алфавита порождающего k-равномерного морфизма.

Итак, пусть $\phi\colon A^* \to A^*$ — k-равномерный морфизм, продолжаемый на некотором символе $s \in A$, $h\colon A \to B$ — кодирование. Мы хотим получить критерий почти периодичности последовательности $h(\phi^\infty(s))$. После этого мы покажем, что наш критерий можно проверить за полиномиальное время.

Для каждого $l \in \mathbb{N}$ определим отношение эквивалентности на A: $b \sim_l c$ если и только если $h(\phi^l(b)) = h(\phi^l(c))$. Это отношение можно естественно продолжить на A^*: $u \sim_l v$, если и только если $h(\phi^l(u)) = h(\phi^l(v))$. На самом деле, условие $u \sim_l v$ значит, что $|u| = |v|$ и $u(i) \sim_l v(i)$ для всех i, таких что $1 \le i \le |u|$.

Пусть $R = R(h, \phi)$ — количество всевозможных отношений $\sim_l$.

Лемма 5.9. $R \le 2^{n^2}$.

Доказательство. Количество всевозможных бинарных отношений на множестве из n элементов не превосходит 2^{n^2}. □

Сформулируем некоторые свойства последовательности отношений $(\sim_l)_{l\in\mathbb{N}}$.

Лемма 5.10. *Если $\sim_r$ совпадает с $\sim_s$, то $\sim_{r+p}$ совпадает с $\sim_{s+p}$ для любого p.*

Доказательство. Действительно, предположим, что $\sim_r$ совпадает с $\sim_s$. Тогда $b \sim_{r+1} c$ равносильно $\phi(b) \sim_r \phi(c)$ равносильно $\phi(b) \sim_s \phi(c)$ равносильно $b \sim_{s+1} c$. Поэтому если $\sim_r$ совпадает с $\sim_s$, то $\sim_{r+1}$ совпадает с $\sim_{s+1}$, откуда следует утверждение леммы. □

В соответствии с этой леммой, последовательность $(\sim_l)_{l\in\mathbb{N}}$ — заключительно периодическая, причём сумма периода и предпериода этой последовательности не превосходит R. Таким образом, мы получаем следующую лемму.

Лемма 5.11. *Для некоторых $p, q \in \mathbb{N}$, таких что $p + q = R$, для всех i и всех $t \geq p$ отношение $\sim_t$ совпадает с отношением $\sim_{t+iq}$.*

Заметим, что $R \leq B_n$, где B_n — n-е число Белла, то есть количество всевозможных отношений эквивалентности на множестве из n элементов. Известно, что $B_n = 2^{O(n \log n)}$, но такая оценка нам не понадобится (нам достаточно леммы 5.9).

В следующих нескольких предложениях, постепенно переформулируя его, мы получаем критерий почти периодичности автоматной последовательности, окончательно сформулированный в теореме 5.15. После чего мы доказываем в теореме 5.16, что этот критерий можно проверить за полиномиальное время.

Следующее предложение прямо следует из определения почти периодичности, поскольку все слова $h(\phi^m(s))$ являются префиксами последовательности $h(\phi^\infty(s))$.

Предложение 5.12. *Последовательность $h(\phi^\infty(s))$ почти периодическая тогда и только тогда, когда для каждого m найдётся такое l, что слово $h(\phi^m(s))$ входит в $h(\phi^\infty(s))$ на каждом отрезке длины l.*

Теперь сформулируем немного более сложный вариант.

Предложение 5.13. *Последовательность $h(\phi^\infty(s))$ почти периодическая тогда и только тогда, когда для любого m существует такое l, что символы, $\sim_m$-эквивалентные символу s, входят в $\phi^\infty(s)$ на любом отрезке длины l.*

Доказательство. $\Leftarrow$. Если расстояние между соседними вхождениями в последовательность $\phi^\infty(s)$ символов, $\sim_m$-эквивалентных s, не больше t, то расстояние между соседними вхождениями слова $h(\phi^m(s))$ в $h(\phi^\infty(s))$ не больше tk^m.

$\Rightarrow$. Предположим, $h(\phi^\infty(s))$ почти периодическая. Пусть $y = 012\ldots(k^m-2)(k^m-1)01\ldots(k^m-1)0\ldots$ — периодическая последовательность с периодом k^m. Тогда по предложению 2.9 последовательность

$h(\phi^\infty(s)) \times y$ почти периодическая. Это означает, что расстояния между соседними k^m-выровненными вхождениями слова $h(\phi^m(s))$ в $h(\phi^\infty(s))$ ограничены. Остаётся только заметить, что если $h(\phi^\infty(s))[ik^m, (i+1)k^m - 1] = h(\phi^m(s))$, то $\phi^\infty(s)(i) \sim_m s$. $\square$

Пусть Y_m — следующее утверждение: символы, $\sim_m$-эквивалентные s, входят в $\phi^\infty(s)$ бесконечно много раз с ограниченными интервалами между соседними вхождениями.

Предположим, что для некоторого T утверждение Y_T верно. Тогда $h(\phi^T(s))$ входит в $h(\phi^\infty(s))$ с ограниченными интервалами. Следовательно, для всех $m \leq T$ слово $h(\phi^m(s))$ также входит в $h(\phi^\infty(s))$ ограниченными интервалами, поскольку $h(\phi^m(s))$ — префикс слова $h(\phi^T(s))$. Таким образом, нам не нужно проверять, что условие Y_m выполнено для любого m, а только для любого $m \geq T$ для некоторого T.

Более того, из леммы 5.11 следует, что достаточно проверить только одно такое утверждение, как в следующем предложении.

Предложение 5.14. *Для любого $r \geq R$ выполнено: последовательность $h(\phi^\infty(s))$ почти периодическая тогда и только тогда, когда символы, $\sim_r$-эквивалентные символу s, входят в $\phi^\infty(s)$ бесконечно много раз с ограниченными интервалами.*

И наконец, окончательная версия критерия.

Теорема 5.15. *Для любого $r \geq R$: последовательность $h(\phi^\infty(s))$ почти периодическая тогда и только тогда, когда существует m, такое что для всех $b \in A$ некоторый символ, $\sim_r$-эквивалентный символу s, входит в слово $\phi^m(b)$.*

Действительно, если символы из некоторого фиксированного множества входят в последовательность с ограниченными интервалами, то для достаточно большого m они входят в каждый k^m-выровненный отрезок длины k^m последовательности.

Теперь мы докажем, что условие теоремы 5.15 можно проверить за полиномиальное время.

Теорема 5.16. *Существует полиномиальный по времени алгоритм, разрешающий по данной автоматной последовательности, является ли она $h(\phi^\infty(s))$ почти периодической.*

Доказательство. Нам необходимо показать две вещи. Во-первых, как выбрать некоторое $r \geq R$ и за полиномиальное время найти множество всех символов, $\sim_r$-эквивалентных символу s (сложность в том, что R может быть экспоненциально большим). Во-вторых, как проверить, что для некоторого m символы из этого множества входят в $\phi^m(b)$ для каждого $b \in A$.

Начнём со второго. Предположим, мы нашли множество H всех символов, $\sim_r$-эквивалентных s. Для $m \in \mathbb{N}$ обозначим через $P_m^{(b)}$ множество всех символов, входящих в $\phi^m(b)$. Наша цель — проверить, существует ли такое m, что для каждого b выполнено $P_m^{(b)} \cap H \neq \varnothing$. Прежде всего, заметим, что если $\forall b \ P_m^{(b)} \cap H \neq \varnothing$, то $\forall b \ P_l^{(b)} \cap H \neq \varnothing$ для всех $l \geq m$. Во-вторых, заметим, что последовательность n-кортежей множеств $(\langle P_m^{(b)} \rangle_{b \in A})_{m=0}^{\infty}$ заключительно периодическая (напомним, что n — размер алфавита A). Действительно, последовательность множеств $(P_m^{(b)})_{m=0}^{\infty}$ для каждого b, очевидно, заключительно периодическая, причём и период, и предпериод не превосходят 2^n. Таким образом, период последовательности $(\langle P_m^{(b)} \rangle_{b \in A})_{m=0}^{\infty}$ не превосходит наименьшего общего кратного периодов $(P_m^{(b)})_{m=0}^{\infty}$ для $b \in A$, и предпериод не превосходит максимального предпериода $(P_m^{(b)})_{m=0}^{\infty}$ для $b \in A$. Значит, период не превосходит $(2^n)^n = 2^{n^2}$ и предпериод не превосходит 2^n. Отсюда, вспоминая первое замечание, получаем, что достаточно выбрать одно фиксированное $m \geq 2^{n^2} + 2^n$, после чего вычислить $\langle P_m^{(b)} \rangle_{b \in A}$ и проверить пересечения с H. В-третьих, заметим, что существует полиномиальная по времени процедура, которая по графу инцидентности некоторого морфизма ψ вычисляет граф инцидентности морфизма ψ^2. Применив эту процедуру $n^2 + 1$ раз, мы получаем граф, с помощью которого легко найти $\langle P_{2^{n^2+1}}^{(b)} \rangle_{b \in A}$. Заметим, что $2^{n^2+1} > 2^{n^2} + 2^n$.

Аналогичные рассуждения используются в разрешении следующей задачи. Здесь мы описываем полиномиальный по времени алгоритм, который находит множество всех символов, $\sim_r$-эквивалентных s, для некоторого $r \geq R$.

Мы рекурсивно строим серию графов T_i. Их общее множество вершин — множество всех неупорядоченных пар $\{b, c\}$, таких что $b, c \in A$ и $b \neq c$. Таким образом, количество вершин в каждом графе $\frac{n(n-1)}{2}$. Множество вершин, соединённых с $\{b, c\}$ в графе T_i, обозначим $V_i(b, c)$.

Определим граф T_0. Пусть $V_0(b, c)$ — множество всех таких пар

$\{\phi(b)(j), \phi(c)(j)\}$ для $j = 1, \ldots, k$, что $\phi(b)(j) \neq \phi(c)(j)$. Другими словами, $b \sim_{l+1} c$, если и только если $x \sim_l y$ для всех $\{x, y\} \in V_0(b, c)$.

Таким образом, $b \sim_2 c$, если и только если для всех $\{x, y\} \in V_0(b, c)$ и для всех $\{z, t\} \in V_0(x, y)$ имеем $z \sim_0 t$. В графе T_1 положим $V_1(b, c)$ равным множеству всех $\{x, y\}$, таких что найдётся путь длины два из $\{b, c\}$ в $\{x, y\}$ в T_0. Граф T_1 обладает следующим свойством: $b \sim_2 c$, если и только если $x \sim_0 y$ для всех $\{x, y\} \in V_1(b, c)$. И даже более общо: $b \sim_{l+2} c$, если и только если $x \sim_l y$ для всех $\{x, y\} \in V_1(b, c)$.

Теперь мы можем повторить операцию, проделанную с T_0 для получения T_1. А именно, в T_2 положим $V_2(b, c)$ равным множеству всех $\{x, y\}$, таких что найдётся путь длины два из $\{b, c\}$ в $\{x, y\}$ в T_1. Отсюда получаем: $b \sim_{l+4} c$, если и только если $x \sim_l y$ для всех $\{x, y\} \in V_2(b, c)$.

Аналогично строятся все T_i. Более явно, $V_i(b, c)$ равно множеству всех пар $\left\{\phi^{2^i}(b)(j),\ \phi^{2^i}(c)(j)\right\}$ для $1 \leq j \leq k^{2^i}$ и $\phi^{2^i}(b)(j) \neq \phi^{2^i}(c)(j)$. Отметим также, что $\phi^{2^i}(b) = \phi^{2^i}(c)$, если и только если $\{b, c\}$ имеет исходящую степень 0 в графе T_i.

Из леммы 5.9 следует, что $\log_2 R \leq n^2$. Поэтому после повторения нашей процедуры n^2 раз мы получим граф T_{n^2}, такой что $b \sim_{2^{n^2}} c$, если и только если $x \sim_0 y$ для всех $\{x, y\} \in V_{n^2}(b, c)$. Напомним, что $x \sim_0 y$ означает $h(x) = h(y)$, так что теперь мы можем вычислить множество символов, $\sim_{2^{n^2}}$-эквивалентных s. $\square$

5.4. Линейность подсловной сложности почти периодических морфических последовательностей

Напомним, что подсловной сложностью последовательности x называется такая функция $\mathrm{p}\colon \mathbb{N} \to \mathbb{N}$, что $\mathrm{p}(n)$ равно количеству слов длины n, входящих в последовательность x. Для последовательностей в алфавите из m символов подсловная сложность может варьироваться от 1 до m^n. Как доказано в [22] (см. предложение 2.5), подсловная сложность последовательности ограничена тогда и только тогда, когда последовательность является заключительно периодической.

В серии работ, завершающейся работой [26], доказано, что подсловная сложность чисто морфической последовательности может удовлетворять одной из следующих пяти асимптотик: $O(1)$, $\Theta(n)$, $\Theta(n \log\log n)$,

$\Theta(n \log n)$, $\Theta(n^2)$. Про подсловную сложность морфических последовательностей произвольного вида долгое время было ничего неизвестно. В [27] показано, что существуют примеры морфических последовательностей с подсловной сложностью вида $\Theta(n^{1+\frac{1}{k}})$ для каждого $k \in \mathbb{N}$. Наконец, в [11] был получен следующий результат.

Теорема 5.17 ([11]). *Подсловная сложность морфической последовательности имеет один из следующих видов: $O(n \log n)$, $\Theta(n^{1+\frac{1}{k}})$ для некоторого $k \in \mathbb{N}$. Каждый из указанных классов непуст.*

Таким образом, для полного описания возможных асимптотик подсловной сложности морфической последовательности осталось разобраться подробнее со случаем $O(n \log n)$.

В случае автоматных последовательностей ситуация существенно проще.

Теорема 5.18 ([9]). *Подсловная сложность автоматной последовательности не более чем линейна.*

Несложно показать (например, см. [14]), что подсловная сложность почти периодической последовательности в алфавите из m символов не может быть больше $m^{\alpha n}$ для некоторого фиксированного $\alpha < 1$. Однако можно показать, что для любого $\beta < 1$ существует почти периодическая последовательность в алфавите из m символов с подсловной сложностью не меньше $m^{\beta n}$ (например, см. [46]).

Оказывается, подсловная сложность последовательности, являющейся одновременно морфической и почти периодической, устроена гораздо проще. Следующий результат был получен в работе [48] (см. также [43]).

Теорема 5.19. *Если x — почти периодическая морфическая последовательность, то $p_x(n) = O(n)$.*

Доказательство теоремы 5.19 содержится в нескольких следующих леммах. Ключевой является лемма 5.24. Другие важные вспомогательные леммы — это леммы 5.20 и 5.22. Леммы 5.21 и 5.23 технические.

Лемма 5.20. *Подсловная сложность чисто морфической последовательности, порождённой примитивным морфизмом, не более чем линейна.*

Лемма 5.20 в явном виде доказана в [29] или в [4] (теорема 10.4.12), но также следует из результатов работы [26].

Лемма 5.21. *Пусть A, B — конечные алфавиты, $\phi\colon A^* \to B^*$ — нестирающий морфизм, и пусть максимальная длина $\phi(a)$ по всем $a \in A$ равна M. Тогда $p_{\phi(x)}(n) \leqslant M p_x(n)$ для каждой последовательности $x \in A^{\mathbb{N}}$ и $n \in \mathbb{N}$.*

Лемму 5.21 можно найти в [4] (теорема 10.2.4).

Лемма 5.22 ([26]). *Пусть A — конечный алфавит, $s \in A$, и $\phi\colon A^* \to A^*$ — морфизм, продолжаемый на s. Предположим, что множество всевозможных ϕ-ограниченных подслов последовательности $\phi^\infty(s)$ конечно. Тогда $\phi^\infty(s)$ может быть представлена как образ под действием нестирающего морфизма чисто морфической последовательности, порождённой возрастающим морфизмом.*

Доказательство леммы 5.22 воспроизводится также в [48].

Лемма 5.23. *Пусть B — конечный алфавит, $\phi\colon B^* \to B^*$ — возрастающий морфизм. Тогда существует натуральное n и такая буква $t \in B$, что морфизм ϕ^n продолжаем на t.*

Доказательство. Пусть $b \in B$. Поскольку алфавит B конечный, существуют такие i, j, что $i < j$ и $\phi^i(b)$ и $\phi^j(b)$ начинаются с одинаковой буквы — обозначим её t. Следовательно $\phi^{j-i}(t)$ начинается с буквы t. Поскольку морфизм ϕ возрастающий, морфизм ϕ^{j-i} также возрастающий. Отсюда ϕ^{j-i} продолжаем на t. □

Лемма 5.24. *Для каждой чисто морфической последовательности x, порождённой возрастающим морфизмом, существует чисто морфическая последовательность y, порождённая примитивным морфизмом, такая что $\mathrm{Fac}(y) \subseteq \mathrm{Fac}(x)$.*

Доказательство. Доказательство проходит индукцией по размеру алфавита последовательности x.

Пусть $x = \phi^\infty(s)$, причём ϕ возрастающий. Пусть B — строго связная компонента графа инцидентности G_ϕ, такая что из B не выходит рёбер, ведущих вне B. Тогда ϕ, ограниченный на B, является возрастающим неприводимым морфизмом $B^* \to B^*$ (здесь мы отождествляем подграф B

и множество его вершин). По лемме 5.23 существуют $t \in B$ и n, такие что ϕ^n продолжаем на t. Если ϕ^n примитивный, то лемма доказана: можно взять последовательность $(\phi^n)^\infty(t)$ в качестве y. Действительно, t входит в x, и поэтому $(\phi^n)^m(t)$ входит в x для любого m.

Предположим, что морфизм ϕ^n не примитивный. Тогда B является собственным подграфом G_ϕ, поскольку иначе ϕ^n одновременно продолжаемый и неприводимый, и значит, примитивный. Обозначим морфизм ϕ^n, ограниченный на B, через ψ.

Морфизм ψ — возрастающий, порождает чисто морфическую последовательность $x' = \psi^\infty(t)$. Алфавит последовательности x' строго меньше алфавита последовательности x. Ясно также, что $\mathrm{Fac}(x') \subseteq \mathrm{Fac}(x)$. По предположению индукции найдётся чисто морфическая y, порождённая примитивным морфизмом, такая что $\mathrm{Fac}(y) \subseteq \mathrm{Fac}(x')$. Отсюда получаем $\mathrm{Fac}(y) \subseteq \mathrm{Fac}(x)$. □

Теперь мы готовы доказать теорему 5.19.

Доказательство теоремы 5.19. Пусть $x = h(\phi^\infty(s))$ — почти периодическая морфическая последовательность, порождённая морфизмом $\phi\colon A^* \to A^*$, $h\colon A \to B$ — кодирование.

Возможны два случая.

1) В последовательность $\phi^\infty(s)$ входит бесконечно много ϕ-ограниченных подслов. Тогда по предложению 5.4 существует непустое $w \in A^*$, такое что w^n входит в $\phi^\infty(s)$ для каждого n. Следовательно, $(h(w))^n$ входит в x для каждого n, и поэтому x периодическая, и её подсловная сложность $O(1)$.

2) В последовательность $\phi^\infty(s)$ входит лишь конечное множество ϕ-ограниченных подслов. Тогда по лемме 5.22 $\phi^\infty(s)$ может быть представлена как $\phi^\infty(s) = g(\psi^\infty(t))$ для некоторых морфизмов $\psi\colon C^* \to C^*$ и $g\colon C^* \to A^*$, так что ψ возрастающий и g нестирающий. По лемме 5.24 существует чисто морфическая y, порождённая примитивным морфизмом, такая что $F(y) \subseteq F(\psi^\infty(t))$. Следовательно, $F(h(g(y))) \subseteq F(x)$, но x почти периодическая, поэтому по предложению 2.1 имеем $F(h(g(y))) = F(x)$. Отсюда для некоторой константы M

$$p_x(n) = p_{h(g(y))}(n) \leq M p_y(n) = O(n),$$

где промежуточное неравенство следует из леммы 5.21, а последнее равенство — из леммы 5.20. □

Глава 6.

Мера апериодичности бесконечных последовательностей

6.1. Определения и простейшие оценки

Периодические последовательности имеют самую простую структуру среди всех последовательностей над конечным алфавитом, поэтому естественно было бы научиться измерять, насколько данная последовательность далека от любой периодической. Для этого мы вводим понятие меры апериодичности последовательности. Неформально, мера апериодичности последовательности — это максимальное такое число α, что последовательность с любым своим нетривиальным сдвигом имеет хотя бы долю α различий. Насколько известно автору, отдельного исследования этого понятия и его свойств ранее не проводилось, началом этому можно считать работу [47], в которой введено понятие меры апериодичности бесконечной символической последовательности и доказаны некоторые его свойства. Отметим, что вскоре после отправки первой версии работы [47] для представления на международном симпозиуме Computer Science in Russia (CSR 2009) появилась работа [15], посвящённая обобщению понятия меры апериодичности, некоторые из результатов этой работы усиливают некоторые из результатов [47]. В изложении результатов этой главы мы следуем работе [47]. Все результаты этой главы, кроме теоремы 6.10, получены автором.

Формальное определение этого понятия основано на дискретном аналоге расстояния Безиковича d_B, использованном в [22] для определения почти периодических по Безиковичу последовательностей. Тот же подход был использован в [12] для определения α-апериодических последовательностей. По существу, в [12] было замечено, в нашей терминологии, что если $\mathrm{AM}(x) > \alpha$ для последовательности x, то x находится на расстоянии хотя бы $\alpha/2$ от любой заключительно периодической последовательности, в смысле расстояния d_B.

Напомним, что расстояние Безиковича определяется как $d_B(x, y) = \liminf_{n\to\infty} \frac{1}{n}\#\{i : 0 \leqslant i \leqslant n-1, x(i) \neq y(i)\}$. Определим *меру апериодичности* $\mathrm{AM}(x) = \inf\{d_B(x, L^n x) : n \geqslant 1\}$, где L обозначает операцию левого сдвига. Другими словами, $\mathrm{AM}(x)$ — это максимальное такое число от 0 до 1, что можно утверждать, что у последовательности x с любым её собственным сдвигом хотя бы доля $\mathrm{AM}(x)$ символов различны. Определим $s_n = \limsup_{m\to\infty} \frac{1}{m}\#\{0 \leqslant i \leqslant m-1 : x(i) = x(i+n)\}$. Тогда $\mathrm{AM}(x) = 1 - \sup\{s_n : n \geqslant 1\}$.

Одним из мотивов для изучения понятия меры апериодичности стала следующая гипотеза Б. Дюрана, А. Ромащенко и А. Шеня: *Для любого $\alpha < 1$ существует автоматная последовательность x, такая что* $\mathrm{AM}(x) \geqslant \alpha$. Эта гипотеза доказана в [47], см. теорему 6.10. Теорема 6.10 позволяет упростить конструкцию сильно апериодического замощения из [12].

Кроме того, понятие меры апериодичности достаточно естественно и просто, чтобы представлять самостоятельный интерес.

Ясно, что если последовательность x заключительно периодическая с периодом p, то $d_B(x, L^p x) = 0$ и, следовательно, $\mathrm{AM}(x) = 0$. Обратное не всегда верно — несложно привести пример апериодической последовательности, мера апериодичности которой равна 0. Тем не менее естественно считать, что чем меньше число $\mathrm{AM}(x)$, тем ближе x к периодической последовательности. В [47] доказано, что $\mathrm{AM}(x) = 0$, если x — последовательность Штурма (см. теорему 6.3 ниже).

Если в вышеупомянутой гипотезе не требовать автоматности последовательности, то утверждение становится почти очевидным.

Теорема 6.1. *Пусть x — случайная последовательность в алфавите из k символов. Тогда* $\mathrm{AM}(x) = 1 - \frac{1}{k}$.

Доказательство. В случайной последовательности в алфавите из k символов частота встречаемости каждого символа равна $\frac{1}{k}$. Ясно также, что если рассмотреть разность по модулю k случайной последовательности со своим сдвигом на произвольное фиксированное $t \geqslant 1$, то эта разность также будет случайна. Из этого следует, что доля совпадений при любом сдвиге будет также $\frac{1}{k}$. $\square$

Отметим, что, формально говоря, мы понимаем утверждение теоремы 6.1 стандартным способом с точки зрения теории меры. При таком подходе утверждение "случайный объект из некоторого множества M обладает свойством P" интерпретируется как "множество объектов из M, не обладающих свойством P, имеет меру ноль". Здесь если Σ_k^∞ — множество бесконечных последовательностей, мы рассматриваем равномерную борелевскую меру μ_k на Σ_k^∞, так что $\mu_k(u) = \frac{1}{k^{|u|}}$ — здесь множество последовательностей, начинающихся со слова u, мы обозначаем просто u. Вышеприведённое доказательство несложно переносится на этот язык стандартными рассуждениями.

Несложно показать, что для фиксированного алфавита можно дать и верхнюю оценку на меру апериодичности.

Теорема 6.2. *Если последовательность x состоит из не более чем k символов, то* $\mathrm{AM}(x) \leqslant 1 - \frac{1}{2k}$.

Доказательство. Без ограничения общности будем считать, что алфавит последовательности — это $\Sigma_k = \{0, \ldots, k-1\}$.

Пусть $u = x[l, l+N]$ — некоторый отрезок последовательности x. Для $0 \leqslant j \leqslant k-1$ обозначим через r_j количество вхождений символа j в u. Имеем $\sum_{j=0}^{k-1} r_j = N+1$. Для $j \in A$ найдётся $\frac{r_j(r_j-1)}{2}$ пар (p, q), таких что $p, q \in [l, l+N]$, $p < q$ и $x(p) = x(q) = j$. Следовательно,

$$\begin{aligned}
&\#\{(p,q) : p, q \in [l, l+N],\ p < q,\ x(p) = x(q)\} = \\
&= \sum_{j=0}^{k-1} \frac{1}{2} r_j(r_j - 1) = \\
&= \sum_{j=0}^{k-1} \frac{1}{2} r_j^2 - \frac{N+1}{2} \geqslant \\
&\geqslant \frac{1}{2k}(N+1)^2 - \frac{1}{2}(N+1).
\end{aligned}$$

Здесь мы использовали неравенство Коши $\sum_{j=0}^{k-1} r_j^2 \geqslant \frac{1}{k}\left(\sum_{j=0}^{k-1} r_j\right)^2$.

Теперь приблизим сумму $\sum_{n=1}^{N} s_n$, исходя из определения s_n:

$$\sum_{n=1}^{N} \frac{1}{m} \#\{i \in [m-1] : x(i) = x(i+n)\} =$$

$$= \frac{1}{m} \#\{(i,n) : i \in [m-1], n \in [1,N],\ x(i) = x(i+n)\} \geqslant$$

$$\geqslant \sum_{t=1}^{\lfloor m/N \rfloor - 1} \frac{1}{m} \#\{(i,n) : i \in [Nt, N(t+1)-1], n \in [1,N],\ x(i) = x(i+n)\} \geqslant$$

$$\geqslant \sum_{t=1}^{\lfloor m/N \rfloor - 1} \frac{1}{m} \#\{(p,q) : p, q \in [Nt, N(t+1)],\ p < q,\ x(p) = x(q)\} \geqslant$$

$$\geqslant \frac{\lfloor m/N \rfloor}{m} \left(\frac{(N+1)^2}{2k} - \frac{N+1}{2} \right) \to \frac{(N+1)^2}{2kN} - \frac{N+1}{2N}$$

при $m \to \infty$. Таким образом, $\sum_{n=1}^{N} s_n \geqslant \frac{(N+1)^2}{2kN} - \frac{N+1}{2N}$, и следовательно, $s_n \geqslant \frac{(N+1)^2}{2kN^2} - \frac{N+1}{2N^2}$ для некоторого $1 \leqslant n \leqslant N$. При $N \to \infty$ можно найти значение s_n, сколь угодно близкое к $\frac{1}{2k}$, следовательно, $\mathrm{AM}(x) \leqslant 1 - \frac{1}{2k}$. $\square$

В [15] получена оценка $\mathrm{AM}(x) \leqslant 1 - \frac{1}{k}$ на меру апериодичности произвольной последовательности x в алфавите из k символов.

6.2. Мера апериодичности конкретных последовательностей

Следующий результат служит подтверждением нашему неформальному пониманию о последовательностях Штурма как о близких к периодическим.

Теорема 6.3. *Если x — последовательность Штурма, то* $\mathrm{AM}(x) = 0$.

Доказательство. Пусть $x = c_{\alpha,\rho}$ — последовательность Штурма (заданная как механическая последовательность; в рамках этого раздела мы используем обозначение c для механических последовательностей). По определению, $0 < \alpha < 1$, число α иррационально, и $0 \leqslant \rho < 1$.

Докажем, что s_n может быть сколь угодно близким к 1. Тогда по определению меры апериодичности получим $\mathrm{AM}(x) = 0$.

Пусть $\varepsilon > 0$. Поскольку α иррационально, можно найти такое n, что $\mathrm{fr}(n\alpha) < \varepsilon$ (здесь $\mathrm{fr}(y) = y - \lfloor y \rfloor$ — дробная часть действительного числа y). Напомним, что $x(j) = 0$, если $0 \leqslant \mathrm{fr}(\alpha j + \rho) < 1 - \alpha$, и $x(j) = 1$, если $1 - \alpha \leqslant \mathrm{fr}(\alpha j + \rho) < 1$ для любого j. Заметим также, что $\mathrm{fr}(\alpha(i+n) + \rho) = \mathrm{fr}(\mathrm{fr}(\alpha i + \rho) + \mathrm{fr}(\alpha n))$. Следовательно,

$$\begin{aligned}
&\{i \in [m-1] : x(i) \neq x(i+n)\} \\
&\subseteq \{i \in [m-1] : 1 - \alpha - \varepsilon \leqslant \mathrm{fr}(\alpha i + \rho) < 1 - \alpha\} \\
&\qquad \cup \{i \in [m-1] : 1 - \varepsilon \leqslant \mathrm{fr}(\alpha i + \rho) < 1\}.
\end{aligned}$$

Отсюда $\#\{i \in [m-1] : x(i) \neq x(i+n)\} \leqslant \#\{i \in [m-1] : 1 - \alpha - \varepsilon \leqslant \mathrm{fr}(\alpha i + \rho) < 1 - \alpha\} + \#\{i \in [m-1] : 1 - \varepsilon \leqslant \mathrm{fr}(\alpha i + \rho) < 1\}$. Но эта величина $< 2\varepsilon m$ асимптотически при $m \to \infty$. Действительно, для любого иррационального β, любого действительного γ и любых действительных a, b, таких что $0 \leqslant a < b \leqslant 1$, имеем $\lim_{m\to\infty} \frac{1}{m}\#\{i \in [m-1] : a \leqslant \mathrm{fr}(\beta i + \gamma) \leqslant b\} = b - a$. Другими словами, последовательность $(\mathrm{fr}(\beta i + \gamma))_{i=0}^{\infty}$ равномерно распределена на отрезке $[0, 1]$ (теорема Кронекера — Вейля, например, см. [35]).

Таким образом, $s_n > 1 - 2\varepsilon$. Поскольку ε можно выбрать сколь угодно малым, имеем $\mathrm{AM}(x) = 0$. □

Интересно теперь вычислить меру апериодичности хорошо известной последовательности Туэ — Морса (см. раздел 2.4).

Теорема 6.4. $\mathrm{AM}(\mathbf{t}) = \frac{1}{3}$.

Доказательство. Пусть $s_n^m = \frac{1}{m}\#\{i \in [m-1] : \mathbf{t}(i) = \mathbf{t}(i+n)\}$.

Прежде всего, заметим, что $s_0^m = 1$ для любого m. Докажем теперь следующую лемму.

Лемма 6.5. *Для любых m и n выполнено:*

$$\begin{aligned}
s_{2n}^{2m} &= s_n^m, \\
s_{2n}^{2m+1} &= \frac{m+1}{2m+1} s_n^{m+1} + \frac{m}{2m+1} s_n^m, \\
s_{2n+1}^{2m} &= 1 - \frac{1}{2}(s_n^m + s_{n+1}^m), \\
s_{2n+1}^{2m+1} &= \frac{m+1}{2m+1}(1 - s_n^{m+1}) + \frac{m}{2m+1}(1 - s_{n+1}^m).
\end{aligned}$$

Доказательство. Идея доказательства этих соотношений состоит в рассмотрении символов последовательности отдельно с чётными и нечётными индексами. Действительно,

$$\begin{aligned}
s_{2n+1}^{2m+1} &= \frac{1}{2m+1}\#\{i \in [2m] : \mathbf{t}(i) = \mathbf{t}(i+2n+1)\} = \\
&= \frac{1}{2m+1}\#\{i \in [m] : \mathbf{t}(2i) = \mathbf{t}(2i+2n+1)\} + \\
&\quad \frac{1}{2m+1}\#\{i \in [m-1] : \mathbf{t}(2i+1) = \mathbf{t}(2i+1+2n+1)\} = \\
&= \frac{1}{2m+1}\#\{i \in [m] : \mathbf{t}(i) \neq \mathbf{t}(i+n)\} + \\
&\quad \frac{1}{2m+1}\#\{i \in [m-1] : \mathbf{t}(i) \neq \mathbf{t}(i+n+1)\} = \\
&= \frac{m+1}{2m+1}(1 - s_n^{m+1}) + \frac{m}{2m+1}(1 - s_{n+1}^m).
\end{aligned}$$

Здесь мы использовали соотношения $\mathbf{t}(2i) = \mathbf{t}(i)$ и $\mathbf{t}(2i+1) = 1 - \mathbf{t}(i)$, которым удовлетворяет последовательность Туэ — Морса. Остальные соотношения леммы доказываются аналогично. □

Из леммы 6.5 следуют соотношения

$$s_1^{2m} = \frac{1}{2} - \frac{1}{2}s_1^m, \quad s_1^{2m+1} = \frac{m}{2m+1}(1 - s_1^m).$$

Заметим, что по определению $s_1^1 = 0$. Докажем, что $\lim_{m\to\infty} s_1^m = \frac{1}{3}$. Пусть $s_1^m = \frac{1}{3} + a_m$. Тогда $a_1 = -\frac{1}{3}$, $a_{2m} = -\frac{1}{2}a_m$ и $a_{2m+1} = -\frac{m}{2m+1}a_m - \frac{1}{6m+3}$ для каждого m. Положим $b_m = 3ma_m$. Тогда $b_1 = -1$, $b_{2m} = 6ma_{2m} = -3ma_m = -b_m$ и $b_{2m+1} = 3(2m+1)a_{2m+1} = -3ma_m - 1 = -b_m - 1$, откуда легко видеть, что $|b_m| = O(\log m)$. Следовательно, $\lim_{m\to\infty} a_m = 0$ и существует предел $\lim_{m\to\infty} s_1^m = s_1 = \frac{1}{3}$.

Теперь из леммы 6.5 при переходе к пределу по m следует, что предел $s_n = \lim_{m\to\infty} s_n^m = \lim_{m\to\infty} \frac{1}{m}\#\{i \in [m-1] : \mathbf{t}(i) = \mathbf{t}(i+n)\}$ существует для любого $n \geqslant 2$. Более того, получаем следующие соотношения:

$$s_{2n} = s_n, \quad s_{2n+1} = 1 - \frac{1}{2}(s_n + s_{n+1})$$

для каждого n. При этом $s_0 = 1$ и $s_1 = \frac{1}{3}$.

Отсюда легко видеть по индукции, что $\frac{1}{3} \leqslant s_n \leqslant \frac{2}{3}$ для любого $n \geqslant 1$. Поскольку $s_3 = \frac{2}{3}$, получаем $\mathrm{AM}(\mathbf{t}) = 1 - \frac{2}{3} = \frac{1}{3}$. □

Заметим, что по существу в [22] было доказано $\mathrm{AM}(\mathbf{t}) \geqslant 1/4$.

Изначально предполагалось, что получить автоматную последовательность с мерой апериодичности, сколь угодно близкой к 1, можно следующим способом.

Следующий класс последовательностей — обобщение последовательности Туэ — Морса — был назван последовательностями Пруэ в [1] (см. [28]) и впоследствии активно изучался (например, см. [3, 31]). Пусть $\phi\colon \{0,\ldots,k-1\}^* \to \{0,\ldots,k-1\}^*$, так что:

$$\begin{aligned}
&\phi(0) = 0\,1\,2\,3\ldots(k-2)\,(k-1)\\
&\phi(1) = 1\,2\,3\ldots(k-2)\,(k-1)\,0\\
&\phi(2) = 2\,3\ldots(k-2)\,(k-1)\,0\,1\\
&\ldots\\
&\phi(k-1) = (k-1)\,0\,1\,2\,3\ldots(k-2).
\end{aligned}$$

По-другому, $(\phi(i))(j) = i + j$ (до конца этого раздела все суммы и разности берутся по модулю k) для $0 \leqslant i, j \leqslant k-1$. Положим $\mathbf{t}_k = \phi^\infty(0)$. Заметим, что $\mathbf{t}_k$ удовлетворяет соотношению $\mathbf{t}_k(ik+p) = \mathbf{t}_k(i) + p$ для любого i и для любого p, такого что $0 \leqslant p \leqslant k-1$.

Изначально предполагалось, что $\mathbf{t}_k$ имеет высокую меру апериодичности. Однако оказалось, что это не так, как показывает следующий результат.

Теорема 6.6. $\mathrm{AM}(\mathbf{t}_k) \leqslant \frac{2}{k+1} - \frac{2}{k^{k-1}(k+1)}$.

Доказательство. Пусть $s_n^m(d) = \frac{1}{m}\#\{i : i \in [m-1], \mathbf{t}_k(i+n) - \mathbf{t}_k(i) = d\}$. Заметим, что $s_0^m(0) = 1$ и $s_0^m(d) = 0$ для любого m и для любого d, такого что $1 \leqslant d \leqslant k-1$.

Лемма 6.7. *Для любых m, n и для любых p, t, таких что $0 \leqslant p, t \leqslant k-1$, выполнено:*

$$s_{kn+p}^{km+t}(d) = \frac{1}{km+t}\left(\sum_{j=0}^{k-p-1} m_j s_n^{m_j}(d-p) + \sum_{j=k-p}^{k-1} m_j s_{n+1}^{m_j}(d-p)\right),$$

где $m_j = m+1$ для $j < t$ и $m_j = m$ для $j \geqslant t$.

Или по-другому:

$$\begin{aligned}
s_{kn+p}^{km+t}(d) = \frac{1}{km+t}\big(&(k-p)(m+1)s_n^{m+1}(d-p)+\\
&+(t-k+p)(m+1)s_{n+1}^{m+1}(d-p) + (k-t)m s_{n+1}^{m}(d-p)\big)
\end{aligned}$$

при $t \geqslant k-p$ *и*

$$s_{kn+p}^{km+t}(d) = \frac{1}{km+t}\left(t(m+1)s_n^{m+1}(d-p)+\right.$$
$$\left.+(k-p-t)ms_n^m(d-p)+pms_{n+1}^m(d-p)\right)$$

при $t < k-p$.

Доказательство. Указанные соотношения можно увидеть из следующего равенства, воспользовавшись соотношением $\mathbf{t}_k(ik+p) = \mathbf{t}_k(i)+p$ для любого i и для любого p, такого что $0 \leqslant p \leqslant k-1$:

$$\begin{aligned}
s_{kn+p}^{km+t} &= \frac{1}{km+t}\#\{i \in [km+t-1] : \mathbf{t}_k(i+kn+p) - \mathbf{t}_k(i) = d\} = \\
&= \frac{1}{km+t}(\#\{i \in [m] : \mathbf{t}_k(ki+kn+p) - \mathbf{t}_k(ki) = d\}+ \\
&\quad + \#\{i \in [m] : \mathbf{t}_k(ki+1+kn+p) - \mathbf{t}_k(ki+1) = d\}+ \\
&\quad + \dots + \#\{i \in [m] : \mathbf{t}_k(ki+k-p-1+kn+p)- \\
&\qquad - \mathbf{t}_k(ki+k-p-1) = d\}+ \\
&\quad + \#\{i \in [m-1] : \mathbf{t}_k(ki+k-p+kn+p) - \mathbf{t}_k(ki+k-p) = d\}+ \\
&\quad + \dots + \#\{i \in [m-1] : \mathbf{t}_k(ki+k-1+kn+p)- \\
&\qquad - \mathbf{t}_k(ki+k-1) = d\}).
\end{aligned}$$

Здесь, чтобы избежать дальнейшего усложнения, мы предположили, что $k-p=t$. В общем случае переход от $i \in [m]$ к $i \in [m-1]$ происходит в t-м слагаемом. □

Лемма 6.8. *Существуют пределы* $s_1(0) = \lim_{m\to\infty} s_1^m(0) = \frac{k-1}{k^k-1}$ *и* $s_1(d) = \lim_{m\to\infty} s_1^m(d) = k^{k-d}\frac{k-1}{k^k-1}$ *для* $1 \leqslant d \leqslant k-1$.

Доказательство. Из леммы 6.7 получаем следующие соотношения:

$$s_1^{km+t}(d) = \frac{1}{km+t}ms_1^m(d-1)$$

для $d \neq 1$ и

$$s_1^{km+t}(1) = 1 - \frac{m}{km+t}(1-s_1^m(0)).$$

Положим $s_1^m(d) - k^{k-d}\frac{k-1}{k^k-1} = a_m(d)$ для $1 \leqslant d \leqslant k-1$ и $s_1^m(0) - \frac{k-1}{k^k-1} = a_m(0)$. Наша цель — показать, что $a_m(d) \to 0$ при $m \to \infty$. Положим также $m \cdot a_m(d) = b_m(d)$.

Из соотношений выше получаем для $2 \leqslant d \leqslant k-1$: $a_{km+t}(d)+k^{k-d}\frac{k-1}{k^k-1} = \frac{1}{km+t}m(k^{k-d+1}\frac{k-1}{k^k-1} + a_m(d-1))$, откуда следует $b_{km+t}(d) = b_m(d-1) - tk^{k-d}\frac{k-1}{k^k-1}$ для $2 \leqslant d \leqslant k-1$. Далее, из соотношений выше получаем также: $a_{km+t}(0) + \frac{k-1}{k^k-1} = \frac{1}{km+t}m(k\frac{k-1}{k^k-1} + a_m(k-1))$, откуда $b_{km+t}(0) = b_m(k-1) - t\frac{k-1}{k^k-1}$. И наконец, из этих соотношений получаем: $a_{km+t}(1) + k^{k-1}\frac{k-1}{k^k-1} = 1 - \frac{m}{km+t}(1 - a_m(0) - \frac{k-1}{k^k-1})$, откуда $b_{km+t}(1) = b_m(0) - tk^{k-1}\frac{k-1}{k^k-1} + t$. Из этих соотношений на $b_m(d)$ несложно видеть, что $|b_m(d)| = O(\log m)$ для любого d. Отсюда следует, что $a_m(d) \to 0$ при $m \to \infty$ для любого d. $\square$

Теперь из леммы 6.7 при переходе к пределу по m для каждого $n \geqslant 2$ следует существование предела $s_n(d) = \lim_{m\to\infty} s_n^m(d)$. Более того, получаем следующие соотношения:

$$s_{kn+p}(d) = \frac{k-p}{k}s_n(d-p) + \frac{p}{k}s_{n+1}(d-p) \tag{6.1}$$

для любых n, d, p, таких что $0 \leqslant d, p \leqslant k-1$. Положим также $s_0(0) = 1$ и $s_0(d) = 0$ для $1 \leqslant d \leqslant k-1$.

Доказательство теоремы завершается следующей леммой, из которой по определению следует, что $\mathrm{AM}(\mathbf{t}_k) \leqslant \frac{2}{k+1} - \frac{2}{k^{k-1}(k+1)}$.

Лемма 6.9.

$$s_{k^k-1}(0) = \frac{1}{k^k}\left(1 + \frac{k-1}{k+1}\frac{k^{2k}-1}{k^k-1}\right) = 1 - \frac{2}{k+1} + \frac{2}{k^{k-1}(k+1)}$$

Доказательство. Докажем для $0 \leqslant i \leqslant k$ индукцией по i, что $s_{k^i-1}(k-i) = \frac{1}{k^i}\left(1 + \frac{k-1}{k+1}\frac{k^{2i}-1}{k^k-1}\right)$. Действительно, $s_0(0) = s_0(k) = 1$. Кроме того, из соотношений (6.1) имеем:

$$\begin{aligned}
&s_{k^{i+1}-1}(k-(i+1)) = \\
&\quad = \frac{1}{k}s_{k^i-1}(k-i) + \frac{k-1}{k}s_{k^i}(k-i) = \\
&\quad = \frac{1}{k}s_{k^i-1}(k-i) + \frac{k-1}{k}k^i\frac{k-1}{k^k-1} = \\
&\quad = \frac{1}{k}\cdot\frac{1}{k^i}\left(1 + \frac{k-1}{k+1}\frac{k^{2i}-1}{k^k-1}\right) + \frac{k-1}{k^{i+1}(k+1)}k^{2i}\frac{k^2-1}{k^k-1} = \\
&\quad = \frac{1}{k^{i+1}}\left(1 + \frac{k-1}{k+1}\frac{k^{2i}-1}{k^k-1}\right) + \frac{1}{k^{i+1}}\frac{k-1}{k+1}\frac{k^{2i+2}-k^{2i}}{k^k-1} = \\
&\quad\quad = \frac{1}{k^{i+1}}\left(1 + \frac{k-1}{k+1}\frac{k^{2i+2}-1}{k^k-1}\right).
\end{aligned}$$

$\square$

$\square$

К сожалению, не удалось доказать равенство $\mathrm{AM}(\mathbf{t}_k) = \frac{2}{k+1} - \frac{2}{k^{k-1}(k+1)}$, хотя можно предположить, что оно верно. Для его доказательства необходимо найти максимум последовательности $s_n(0)$, рекурсивно определённой в доказательстве выше. Правдоподобно было бы предположить, что этот максимум достигается как раз в $n = k^k - 1$. Это подтверждается компьютерными экспериментами.

Теперь приведём конструкцию автоматной последовательности с мерой апериодичности, сколь угодно близкой к 1.

Пусть $k \geqslant 3$ и $\phi\colon \{0, \ldots, k-1\}^* \to \{0, \ldots, k-1\}^*$ — морфизм, такой что $(\phi(i))(j) = i+1+2+\cdots+(j-1)+j$ для $0 \leqslant i, j \leqslant k-1$, где $+$ всегда берётся по модулю k, а $u(i)$ обозначает i-й символ конечного слова u. Положим $x_k = \phi^\infty(0)$. Например, если $k = 5$, то ϕ получается таким:

$$\begin{aligned}
\phi(0) &= 01310\\
\phi(1) &= 12421\\
\phi(2) &= 23032\\
\phi(3) &= 34143\\
\phi(4) &= 40204,
\end{aligned}$$

и $x_5 = 0131012421341431242101310124 21\ldots$

Можно доказать (см. [47], результат получен вторым автором), что если k простое, то $\mathrm{AM}(x_k) = 1 - \frac{2}{k}$. Таким образом, получаем следующий результат.

Теорема 6.10 ([47]). *Для любого $\alpha < 1$ существует автоматная последовательность x, такая что $\mathrm{AM}(x) \geqslant \alpha$.*

Литература

[1] Adler A., Li S.-Y. R. Magic cubes and Prouhet sequences. *American Mathematical Monthly*, 84:618–627, 1977.

[2] Allouche J.-P., Shallit J. The ubiquitous Prouhet–Thue–Morse sequence // *Sequences and their applications*, Proceedings of SETA'98, pp. 1–16. Springer-Verlag, 1999.

[3] Allouche J.-P., Shallit J. Sums of digits, overlaps, and palindromes. *Discrete Mathematics and Theoretical Computer Science*, 4:1–10, 2000.

[4] Allouche J.-P., Shallit J. *Automatic Sequences: Theory, Applications, Generalizations* // Cambridge University Press, 2003.

[5] Allouche J.-P. Sur la complexité des suites infinies // *Bulletin of the Belgian Mathematical Society—Simon Stevin*, 1(2):133–143, 1994.

[6] Berstel J. Axel Thue's work on repetitions in words // *Seŕies Formelles et Combinatoire Algébrique*, pp. 65–80, 1992.

[7] Büchi J. R. Weak second-order arithmetic and finite automata // *Z. Math. Logik Grundlagen Math.*, 6:66–92, 1960.

[8] Büchi J. R. On a decision method in restricted second-order arithmetic // *Proceedings of International Congress for Logic, Methodology and Philosophy of Science*, pp. 1–11. Stanford University Press, 1962.

[9] Cobham A. Uniform tag sequences // *Mathematical Systems Theory*, 6:164–192, 1972.

[10] Dekking F. M. Iteration of maps by an automaton // *Discrete Mathematics*, 126(1–3):81–86, 1994.

[11] Devyatov R. On subword complexity of morphic sequences // *Proceedings of the 3rd International Computer Science Symposium in Russia (CSR 2008), Moscow, Russia*, *Lecture Notes in Computer Science*, vol. 5010, pp. 146–157. Springer-Verlag, 2008.

[12] Durand B., Romashchenko A., Shen A. Fixed point and aperiodic tilings // *Proceedings of the 12th International Conference on Developments in Language Theory (DLT 2008), Kyoto, Japan*, *Lecture Notes in Computer Science*, vol. 5257, pp. 276–288. Springer-Verlag, 2008. See also preprint arXiv:0802.2432.

[13] Ehrenfeucht A., Rozenberg G. Repetition of subwords in DOL languages // *Information and Control*, 59(1–3):13–35, 1983.

[14] Ferenczi S. Complexity of sequences and dynamical systems // *Discrete Mathematics*, 206(1):145–154, 1999.

[15] Grant E., Shallit J., Stoll T. Bounds for the discrete correlation of infinite sequences on k symbols and generalized Rudin–Shapiro sequences // Preprint arXiv:0812.3186, 2008.

[16] Harju T., Karhumäki J. Morphisms // In G. Rozenberg, A. Salomaa, editors, *Handbook of formal languages*, vol. 1, pp. 439–510. Springer-Verlag, 1997.

[17] Jacobs K. *Maschinenerzeugte 0-1-Folgen* // Selecta Mathematica II. Springer-Verlag, 1970. Русский перевод: Якобс К. Машинно-порождённые 0–1 последовательности // Сборник Эббинхауз Г.-Д., Якобс К., Ман Ф.-К. *Машины Тьюринга и рекурсивные функции*, сс. 216–247. Мир, 1972.

[18] Keane M. Generalized morse sequences // *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 10:335–353, 1968.

[19] Lothaire M. *Combinatorics on Words* // Cambridge University Press, 2nd edition, 1997.

[20] Lothaire M. *Algebraic Combinatorics on Words* // Cambridge University Press, 2002.

[21] McNaughton R. Testing and generating infinite sequences by a finite automaton // *Information and Control*, 9:521–530, 1966.

[22] Morse M., Hedlund G. A. Symbolic dynamics // *American Journal of Mathematics*, 60(4):815–866, 1938.

[23] Morse M., Hedlund G. A. Symbolic dynamics ii: Sturmian trajectories // *American Journal of Mathematics*, 62(1):1–42, 1940.

[24] Morse M. Recurrent geodesics on a surface of negative curvature // *Transactions of the American Mathematical Society*, 22:84–100, 1921.

[25] Muchnik An., Semenov A., Ushakov M. Almost periodic sequences // *Theoretical Computer Science*, 304(1–3):1–33, 2003.

[26] Pansiot J.-J. Complexité des facteurs des mots infinis engendrés par morphismes itérés // *Proceedings of the 11th International Colloquium on Automata, Languages, and Programming (ICALP'84), Antwerp, Belgium*, *Lecture Notes in Computer Science*, vol. 172, pp. 380–389. Springer-Verlag, 1984.

[27] Pansiot J.-J. Subword complexities and iteration // *Bulletin of the European Association for Theoretical Computer Science*, 26:55–62, 1985.

[28] Prouhet M. E. Mémoire sur quelques relations entre les puissances des nombres. *Comptes Rendus des Séances de l'Académie des Sciences*, 33:225, 1851. Available at `http://gallica.bnf.fr/ark:/12148/bpt6k29901.image.f227.langFR`.

[29] Queffélec M. *Substitution Dynamical Systems—Spectral Analysis* // *Lecture Notes in Mathematics*, vol. 1284. Springer-Verlag, 1987.

[30] Rozenberg G., Salomaa A., editors. *Lindenmayer Systems: Impacts on Theoretical Computer Science, Computer Graphics, and Developmental Biology* // Springer-Verlag, 1992.

[31] Séébold P. On some generalizations of the Thue–Morse morphism. *Theoretical Computer Science*, 292:283–298, 2003.

[32] Sipser M. *Introduction to the Theory of Computation* // PWS Publishing Company: Boston, 2nd edition, 2005.

[33] Thue A. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen // *Selected mathematical papers of Axel Thue*, pp. 413–478. Oslo, Norway: Universitetsforlaget, 1977.

[34] Thue A. Über unendliche Zeichenreihen // *Selected mathematical papers of Axel Thue*, pp. 139–158. Oslo, Norway: Universitetsforlaget, 1977.

[35] Каток А. Б., Хасселблат Б. *Введение в современную теорию динамических систем* // Факториал, 1999. Пер. с англ.: Katok A., Hasselblatt B. *Introduction to the Modern Theory of Dynamical Systems* // Cambridge University Press, 1995.

[36] Роджерс Х. *Теория рекурсивных функций и эффективная вычислимость* // Мир, 1972. Пер. с англ.: Rogers H., *Theory of Recursive Functions and Effective Computability* // MIT Press, 1967.

[37] Семёнов А. Л. О некоторых расширениях арифметики сложения натуральных чисел // *Известия АН СССР, серия математическая*, 43(5):1175–1195, 1979.

[38] Семёнов А. Л. Логические теории одноместных функций на натуральном ряде // *Известия АН СССР, серия математическая*, 47(3):623–658, 1983.

Работы автора по теме диссертации

[39] Притыкин Ю. Л. Конечно-автоматные преобразования строго почти периодических последовательностей // *Математические заметки*, 80(5):751–756, 2006.

[40] Притыкин Ю. Л. Почти периодичность, конечно-автоматные преобразования и вопросы эффективности // *Известия ВУЗ. Математика*, в печати. См. препринт arXiv:cs/0607009.

[41] Притыкин Ю. Л. Конечно-автоматные преобразования почти периодических последовательностей и алгоритмическая неразрешимость // *Труды XXVIII Конференции молодых учёных*, сс. 177–181, мех.-мат. ф-т МГУ им. Ломоносова, 2006.

[42] Притыкин Ю. Л. О нерегулярности некоторых множеств бесконечных слов // *Труды 30-й конференции молодых учёных Информационные Технологии и Системы (ИТиС 2007)*, сс. 134–137, Институт проблем передачи информации РАН, Москва, 2007.

[43] Мучник Ан. А., Притыкин Ю. Л., Семёнов А. Л. Последовательности, близкие к периодическим // Препринт arXiv:0903.5316, 2009.

[44] Pritykin Yu., Raskin M. Almost periodicity and finite automata // *Electronic Proceedings of Workshop on Infinite Words, Automata, and Dynamics (satellite to the 2nd International Computer Science Symposium in Russia (CSR 2007)), Ekaterinburg, Russia*, 2007.

[45] Pritykin Yu. On almost periodicity criteria for morphic sequences in some particular cases // *Proceedings of the 11th International Conference on Developments in Language Theory (DLT 2007), Turku, Finland*, *Lecture Notes in Computer Science*, vol. 4588, pp. 361–370. Springer-Verlag, 2007.

[46] Pritykin Yu. Information in infinite words // *Proceedings of the 6th International Conference on Words (Words 2007), Marseille, France*, pp. 254–261. Institute de Mathématiques de Luminy, Marseille, France, 2007.

[47] Pritykin Yu., Ulyashkina J. Aperiodicity measure for infinite sequences // *Proceedings of the 4th International Computer Science Symposium in Russia (CSR 2009), Novosibirsk, Russia*, Lecture Notes in Computer Science, vol. 5675, pp. 274–285. Springer-Verlag, 2009.

[48] Nicolas F., Pritykin Yu. On uniformly recurrent morphic sequences // *International Journal of Foundations of Computer Science*, to appear.