Уральский ордена Трудового Красного знамени
государственный университет имени А. М. Горького

На правах рукописи

Гамзова Юлия Васильевна

Комбинаторные свойства частичных слов

01.01.09 Дискретная математика и математическая кибернетика

Диссертация на соискание ученой степени
кандидата физико-математических наук

научные руководители
кандидат физико-математических наук,
доцент **Шур А. М.,**
кандидат физико-математических наук,
доцент **Суханов Е. В.**

Екатеринбург
2006

Оглавление

Введение

1°. Понятие о частичных словах

Символьные последовательности, или слова, представляют собой важный и популярный объект комбинаторных исследований. Задачи, связанные со словами, возникали в различных областях математики и других наук и привели к возникновению отдельного раздела дискретной математики, занимающейся изучением комбинаторных свойств слов. Историю комбинаторики слов принято отсчитывать от 1906 года, когда была опубликована работа [1]. С тех пор комбинаторика слов плодотворно развивалась и к настоящему времени обрела четкие контуры и собственную богатую проблематику. Имеющаяся литература весьма обширна. Упомянем здесь монографии [2–4], а также [5] и ряд других глав трехтомного «Справочника по формальным языкам».

Частичные слова представляют собой естественное обобщение понятия обычного слова. Частичное слово — это обычное слово, в котором часть букв по каким-то причинам отсутствует. Изучение комбинаторных свойств частичных слов «в явном виде» началось совсем недавно, в 1999 году, с работы [6].

Задачи, в которых возникают частичные слова, можно разделить на два типа. В задачах первого типа некоторая часть информации нам не важна (вне зависимости от того, доступна она или нет). Примером может служить задача поиска в тексте по шаблону, когда шаблон содержит символ (или символы) со значением «что угодно». В задачах второго типа некоторая часть информации для нас важна, но по каким-либо причинам недоступна. По имеющейся части информации и некоторым дополнительным условиям необходимо полностью восстановить информацию. Примером является задача восстановления фрагментов файлов, повреждённых при передаче или в результате порчи носителя. Очень интересными представляются также задачи молекулярной биологии (см. [7]), в частности, задача реконструкции нуклеотидных цепочек ДНК по частично расшифрованным фрагментам.

В данной работе изучаются комбинаторные свойства частичных слов, связанные с локальной и глобальной периодичностью.

2°. Определения и обозначения

1. Пусть A — конечный алфавит, A^+ и A^* — соответственно свободная полугруппа и свободный моноид над этим алфавитом. Элементы из A^+ и A^* называются словами, а множества слов — языками. Длина слова W обозначается через $|W|$. Для данной буквы $a \in A$ обозначим через $|U|_a$ количество букв a в слове U. На слово длины n можно смотреть как на функцию из множества $\{1, \ldots, n\}$ в алфавит. *Частичное* слово длины n над алфавитом A — это частичная функция
$$W : \{1, \ldots, n\} \to A.$$
Значение $W(i)$ мы будем называть *буквой в i-ой позиции* слова W. Слова (обычные и частичные) мы обозначаем буквами P, Q, R, S, T, U, V, W (с индексами), пустое слово обозначаем через λ.

2. Кроме конечных частичных слов, мы будем рассматривать также бесконечные частичные слова. Частичное *Z-слово* (*ω-слово*) над алфавитом A — это частичная функция $W : \mathbb{Z} \to A$ (соответственно, $W : \mathbb{N} \to A$).

3. Через $D(W)$ обозначим область определения частичной функции $W(i)$. Каждому частичному слову W над алфавитом A мы можем поставить в соответствие обычное слово $W_\diamond$ над расширенным алфавитом $A_\diamond = A \cup \{\diamond\}$ следующим образом:
$$W_\diamond(i) = \begin{cases} W(i), & i \in D(W) \\ \diamond, & i \notin D(W). \end{cases}$$
Отображение $W \to W_\diamond$ является биекцией. В дальнейшем под частичным словом мы будем понимать именно последовательность символов над алфавитом $A \cup \{\diamond\}$.

   Заметим, что роль символа $\diamond$ существенно отличается от роли остальных алфавитных символов. Его значение таково: «на этой позиции должен находиться алфавитный символ, неизвестно какой». В [6] символ $\diamond$ назывался *дырой.* Мы используем термин *джокер*, который кажется нам более уместным. Отличные от джокера символы алфавита, будем, как обычно, называть *буквами.*

4. Обычное слово W имеет *период* p, если $W(i) = W(i+p)$ для всех $i = 1, \ldots, |W| - p$. Понятие периода можно распространить на частичные слова двумя естественными способами, предложенными в [6]. Приведём соответствующие определения.

   Частичное слово W имеет *период* p, если для любых $i, j \in D(W)$

   $$i \equiv j (\mathrm{mod} p) \Rightarrow W(i) = W(j).$$

   Например, слово $W = ab\diamond a\diamond cabc$ имеет период 3.

   Частичное слово W имеет *локальный период* p, если для любого i такого, что $i, i+p \in D(W)$, выполняется $W(i) = W(i+p)$. Для обычных слов наличие локального периода p всегда влечет наличие периода p, но это свойство не выполняется для частичных слов. Например, частичное слово $W = abc\diamond bcd$ имеет локальный период 3, но не имеет периода 3. Заметим, что частичное слово имеет период p тогда и только тогда, когда в нём все джокеры можно заменить буквами так, что полученное обычное слово будет иметь период p.

5. Слово V называется *подсловом* слова W (или говорят, что W *содержит* V, обозначается $V \leq W$), если $W = PVQ$ для некоторых слов P, Q. При этом мы говорим, что V является

   - *префиксом* W, если $P = \lambda$;
   - *суффиксом* W, если $Q = \lambda$.

   Если V — подслово W, начинающееся в i-й и заканчивающееся в j-й позиции, то используется запись $V = W(i \ldots j)$.

   Любое представление слова в виде произведения подслов мы называем *разбиением* этого слова.

6. Будем обозначать через $\lfloor n \rfloor$ нижнее целое, через $\lceil n \rceil$ — верхнее целое числа n.

3°. Краткий обзор исследований по частичным словам

С момента возникновения комбинаторики частичных слов исследования по частичным словам велись довольно интенсивно и уже получено много интересных результатов. Комбинаторика частичных слов находится в начальной фазе развития, когда основной интерес представляет перенесение комбинаторных свойств обычных слов на более широкий класс частичных слов. Уже в [6] были получены для частичных слов аналоги некоторых комбинаторных свойств обычных слов. В частности, рассматривалось в частных случаях свойство взаимодействия периодов частичных слов. Исследование этого свойства продолжалось и в других работах. Свойство взаимодействия периодов тесно связано с нашими исследованиями; оно будет рассмотрено в следующем параграфе.

Рассматривались и другие свойства периодических частичных слов. В частности, в [8] проводились исследования аналога теоремы о критическом разбиении. Введем необходимые определения. Число $p \in \mathbb{N}$ называется *локальным периодом* (обычного) слова W *в позиции* k, $0 < k < |W|$, если слово $W(n_1 \dots n_2)$ имеет период p, где $n_1 = \max\{k - p + 1, 1\}$, $n_2 = \min\{k + p, |W|\}$. Смысл понятия локального периода таков: в p-окрестности позиции k слово W ведет себя как периодическое с периодом p.

Разбиение $W = UV$ называется *критическим*, если минимальный локальный период слова W в позиции $|U|$ совпадает с его минимальным периодом.

Теорема 0.1 (теорема о критическом разбиении, [9, 10]). *Всякое слово W, имеющее минимальный период p и содержащее не менее двух различных букв, допускает критическое разбиение $W = UV$ с условием $|U| < p$.*

Из доказательства этой теоремы следует эффективный алгоритм построения этого разбиения.

Введем для частичных слов понятие, аналогичное понятию локального периода для обычных слов. Число $p \in \mathbb{N}$ называется *точечным периодом* частичного слова W *в позиции* k, $0 < k < |W|$, если слово $W(n_1 \dots n_2)$ имеет период p, где $n_1 = \max\{k - p + 1, 1\}$, $n_2 = \min\{k + p, |W|\}$. Заметим, что для данного определения можно использовать любое из двух определений периода частичного слова, т.к. длина периодической части слишком мала для того, чтобы успели проявиться различия между локальной и глобальной периодичностью.

В [8] доказан «условный» аналог теоремы о критическом разбиении для частичных слов с одним джокером, т.е. выявлены необходимые и достаточные условия наличия в слове критического разбиения. В [11] этот результат обобщен для частичных слов с любым количеством джокеров.

Также для частичных слов проводились исследования множества периодов. Обозначим через $\mathcal{P}(U)$ множество периодов слова U. Для обычных слов известен следующий результат.

Теорема 0.2 ([12]). *Для любого слова U над алфавитом A существует слово V длины $|U|$ над алфавитом $\{0,1\}$ такое, что $\mathcal{P}(V) = \mathcal{P}(U)$.*

Доказательство этой теоремы, приведенное в [12], является довольно сложным. В [13] дается более простое конструктивное доказательство. Алгоритм из [13] был обобщен в [14] для частичных слов с одним джокером, в [15] — с двумя джокерами, в [16] — с произвольным количеством джокеров.

Также изучались свойства примитивных частичных слов. Будем говорить, что частичное слово U *содержится* в частичном слове V $(U \subset V)$, если $|U| = |V|$, $D(U) \subset D(V)$ и $U(i) = V(i)$ для всех $i \in D(U)$. Частичные слова U и V называются *совместимыми* $(U \uparrow V)$, если существует слово W, в котором содержатся и слово U, и слово V. Обычное слово U называется *примитивным*, если не существует слова V такого, что $U = V^n, n \geq 2$. Частичное слово U называется *примитивным*, если не существует слова V такого, что $U \subset V^n, n \geq 2$. В [17] приведены для примитивных частичных слов аналоги известных свойств примитивных обычных слов. В [18] приведен для частичных слов аналог известного критерия примитивности обычных слов, позволяющий создать линейный по времени алгоритм проверки примитивности частичного слова.

Большой интерес представляет изучение множеств частичных слов (*ч-языков*). В 2004 году Ф. Бланше-Садри сделала первый шаг в изучении свойств ч-языков, введя понятие ч-кодов в статье [19].

Напомним, что непустое множество слов $X \in A^+$ называется *кодом*, если для всех натуральных m, n и слов $U_1, \ldots, U_m, V_1, \ldots, V_n \in X$ выполняется

$$(U_1 U_2 \ldots U_m = V_1 V_2 \ldots V_n) \Rightarrow (m = n) \wedge (U_i = V_i), i = 1, \ldots, m$$

Непустое множество частичных слов $X \in A_\diamond^+$ называется *ч-кодом*, если для всех натуральных m, n и частичных слов $U_1, \ldots, U_m, V_1, \ldots, V_n \in X$ выполняется

$$(U_1 U_2 \ldots U_m \uparrow V_1 V_2 \ldots V_n) \Rightarrow (m = n) \wedge (U_i = V_i), i = 1, \ldots, m$$

В [19] приведены различные свойства ч-кодов. В [20] приводится алгоритм для проверки, является ли данный ч-язык ч-кодом.

П. Лёйпольд в статье [21] предложил несколько способов получения ч-языков, а также привел некоторые свойства этих языков.

Для получения *k-проколотого* языка в каждом слове «обычного» языка некоторые символы (не более k) заменяются джокерами. Проколотые языки можно генерировать (например, при помощи грамматик) двумя способами: либо сначала генерируется обычный язык, а затем выполняется прокалывание, либо джокер рассматривается как терминальный символ грамматики, порождающей сразу проколотый язык.

Обозначим через $\mathcal{C}^{(k-\diamond)}$ класс всех k-проколотых языков, полученных с помощью прокалывания языков из класса языков $\mathcal{C}$, а через $\mathcal{C}_{(k-\diamond)}$ — класс языков, порожденных грамматиками того же класса, что и класс языков $\mathcal{C}$, с дополнительным терминальным символом $\diamond$. Обозначим через REG (CF, CS, RE) класс всех языков, порожденных регулярными (соответственно, контекстно-свободными, контекстно-зависимыми, любыми) грамматиками.

Предложение 0.1 ([21]). *Для $\mathcal{C} \in \{REG, CF, CS, RE\}$ и любого натурального k класс языков $\mathcal{C}^{(k-\diamond)}$ является собственным подмножеством класса $\mathcal{C}_{(k-\diamond)}$.*

В [21] также приведен несколько другой подход к рассмотрению кодов в случае частичных слов. В отличие от ч-кода, элементами которого являются произвольные частичные слова, k-проколотый код является замыканием обычного кода относительно всевозможных k-прокалываний. Язык, порожденный проколотым кодом, содержит все частичные слова, которые могли быть получены из произведений исходных кодовых слов в результате потери некоторой части букв. В [21] приведены различные свойства k-проколотых кодов.

4°. Вопросы, рассматриваемые в диссертации

Перенесение свойств обычных слов на частичные слова приводит к постановкам разной значимости. Рассмотрение вместо обычных слов частичных слов может привести к расширению задачи или к ее сужению. Богатую проблематику для исследований по частичным словам предоставляет одно из ключевых свойств обычных периодических слов — *свойство взаимодействия периодов*. Свойство взаимодействия периодов (в «стандартной» формулировке) заключается в том, что достаточно длинное слово с периодами p и q имеет также и «производный» период $\textit{НОД}(p, q)$. Точная формулировка для обычных слов содержится в теореме Файна-Вильфа, которую мы приведём здесь в удобных для нас обозначениях.

Теорема 0.3 (Файн, Вильф [22]). *Пусть p и q – натуральные числа, тогда каждое слово длины не менее, чем $p+q-НОД(p,q)$ с периодами p и q имеет период $НОД(p,q)$. Указанная оценка является точной.*

Свойство взаимодействия периодов интенсивно изучалось и был получен ряд обобщений теоремы Файна-Вильфа. В частности, были получены результаты для трех и более периодов ([23–27]), для абелевых периодов ([28]), для неравенств ([29]), а также для полупериодических слов ([30]), для двумерных слов ([31]), для слов на деревьях ([32–34]).

Для частичных слов выполнение свойства взаимодействия периодов зависит не только от длины слова, но и от количества и расположения в нем джокеров, причем задача может ставиться как для периодов, так и для локальных периодов. Все это приводит к разнообразию нетривиальных постановок задач. Вначале рассмотрим соответствующие постановки для периодов.

1. Верно ли, что частичное слово достаточно большой длины, имеющее периоды p и q и фиксированное число джокеров k, обязательно имеет период $НОД(p,q)$?

2. Если ответ на первый вопрос — утвердительный, то какой длины должно быть частичное слово, имеющее периоды p,q и содержащее k джокеров, чтобы оно также гарантированно имело период $НОД(p,q)$ (тривиальный случай длины $k+1$ исключим из рассмотрения)?

Наименьшую такую длину будем называть *длиной взаимодействия периодов p и q при наличии k джокеров* и обозначать $L(k,p,q)$. Длину взаимодействия естественно рассматривать как функцию от k с параметрами p и q.

Как отмечалось в разных работах (см. [5,6]), утверждения о взаимодействии произвольных периодов p и q тривиально следуют из аналогичных утверждений для взаимно простых периодов ввиду следующего наблюдения.

Наблюдение. В случае, когда $НОД(p,q)=d>1$, можно заменить (частичное) слово U набором, состоящим из d (частичных) слов $U_1,\ldots,U_d$, где $U_i=U(i)U(d+i)U(2d+i)\ldots$. Каждое из этих слов будет иметь взаимно простые периоды p/d, q/d. При этом слово U будет иметь период d тогда и только тогда, когда все U_i имеют период 1.

Таким образом, достаточно исследовать свойство взаимодействия для взаимно простых периодов. В этом случае свойство взаимодействия периодов заключается в том, что в достаточно длинном слове с периодами p и q все буквы должны быть одинаковыми. В дальнейшем мы будем рассматривать только взаимно простые периоды, $2\le q<p$.

Свойство взаимодействия периодов можно обобщить следующим образом. Для данного частичного слова W рассмотрим множество частичных слов с теми же периодами и той же областью определения. Среди этих слов выберем слово U с максимальным количеством различных букв. Множество всех позиций слова U, содержащих одну и ту же букву, будем называть *доменом* W. Количество доменов слова W (т.е. количество различных букв в U) назовем *размерностью* слова W и обозначим через $r(W)$. Из определений следует, что количество букв в частичном слове не превышает его размерность, а две позиции могут содержать различные буквы только если эти позиции принадлежат разным доменам. Обобщенное свойство взаимодействия периодов состоит в выполнении некоторых условий для множества доменов частичного слова. Например, выполнение свойства взаимодействия (взаимно простых) периодов в стандартной формулировке эквивалентно тому, что размерность слова равна единице.

Из определений следует, что размерность и множество доменов частичных слов определяются только периодами и областью определения. Поэтому при изучении свойства взаимодействия периодов мы вместо частичных слов будем рассматривать расстановки. *Расстановка* R — это слово над алфавитом $\{0,1\}$, полученное из частичного слова W с помощью следующего гомоморфизма:

$$R(i) = \begin{cases} 1, & i \in D(W) \\ 0, & i \notin D(W). \end{cases}$$

Все частичные слова, которым соответствует данная расстановка R, имеют одинаковую размерность и множество доменов, поэтому мы будем говорить о множестве доменов и размерности расстановки, понимая под этим множество доменов и размерность соответствующих частичных слов.

Следующее несложное предложение показывает одно из направлений обобщения результата теоремы Файна-Вильфа для обычных слов.

Предложение 0.2 ([5]). *Если слово U имеет взаимно простые периоды p и q и $|U| = p+q-r$, где $1 \le r \le q$, то U содержит не более r различных букв.*

Наличие в частичном слове W (глобального) периода p накладывает ограничения на множество его доменов: размерность W не превосходит p, т.е. W содержит не более p различных букв. Таким образом, глобальная периодичность является достаточно сильным условием. Для периодических частичных слов имеет смысл пытаться получить аналог приведенного выше предложения, т.е. считать, что обобщенное свойство взаимодействия периодов состоит в выполнении некоторых ограничений для размерности частичного слова.

Таким образом, следует исследовать расстановки, длина которых меньше длины взаимодействия. В этом случае расстановки одинаковой длины с одинаковым количеством джокеров могут иметь различную размерность, и интересным представляется вопрос о доле расстановок определенной размерности.

Все предыдущее говорилось о глобальной периодичности, теперь речь пойдет о локальной периодичности. В [6] был доказан аналог теоремы Файна-Вильфа для локально периодических частичных слов с одним джокером.

Теорема 0.4 ([6]). *Пусть p и q – натуральные числа. Тогда любое частичное слово длины не менее, чем $p+q$ с одним джокером и локальными периодами p и q имеет период НОД(p,q). Указанная оценка является точной.*

Однако, если частичное слово содержит более одного джокера, то свойство взаимодействия локальных периодов может вообще не выполняться. Например, частичное слово U с периодами p и q, содержащее два джокера в позициях $q{+}1$ и $p{+}1$, может содержать в первой позиции любую букву. Приведенный пример показывает, что в частичном слове могут существовать некоторые локальные структуры, которые независимо от длины слова приводят к невыполнению свойства взаимодействия локальных периодов. Таким образом, для локально периодических слов более, чем с одним джокером, можно доказывать лишь «условные» теоремы. Такие результаты приведены в статьях [35,36]. Мы приведем здесь теорему в удобных для нас обозначениях. Будем называть *специальными* конечные подпоследовательности позиций джокеров, наличие которых приводит к невыполнению свойства взаимодействия локальных периодов во всем слове.

Теорема 0.5 ([36]). *Пусть частичное слово U без специальных подпоследовательностей имеет взаимно простые локальные периоды p и q и содержит k джокеров. Тогда если*

$$|U| \geq (\lfloor k/2 \rfloor + 1)(p+q) - (k+1) \bmod 2,$$

то U также имеет период 1.

Таким образом, для локально периодических частичных слов, не содержащих специальных подпоследовательностей, выполняется свойство взаимодействия периодов. Однако в бесконечном частичном слове вероятность наличия специальных подпоследовательностей равна единице (если джокеры распределены равномерно, т.е. джокер с равной ненулевой вероятностью может находиться в любой позиции). Исследование специальных подпоследовательностей пока не производилось и представляет большой интерес.

Локальная периодичность накладывает на слово гораздо более слабые ограничения, чем глобальная. В частности, наличие у слова локального периода p не ограничивает размерность слова. Поэтому для локально периодических частичных слов имеет смысл рассматривать в качестве обобщенного свойства взаимодействия локальных периодов выполнение некоторых условий для множества доменов. Выполнение обобщенного свойства взаимодействия локальных периодов тесно связано с наличием в частичном слове специальных подпоследовательностей. Изучение специальных подпоследовательностей позволяет ответить на многие вопросы о множестве доменов частичного слова.

Обратимся теперь к проблемам, решаемым в диссертации. В первой главе диссертации дается ответ на ряд вопросов, связанных со свойством взаимодействия (глобальных) периодов частичных слов. Перечислим эти вопросы.

1. Верно ли, что длина взаимодействия $L(k, p, q)$ существует для любых периодов p и q и количества джокеров k?

2. Какова «критическая» расстановка джокеров в частичном слове, длина которого на единицу меньше длины взаимодействия?

3. Как длина взаимодействия периодов зависит от количества джокеров в частичном слове?

   Для ответа на последний вопрос предполагается

   3.1. Оценить рост длины взаимодействия как функции от количества джокеров k в общем случае;

   3.2. Вычислить эту длину точно в тех частных случаях, в которых это возможно.

Во второй главе рассматриваются периодические частичные слова, длина которых меньше длины взаимодействия. Для таких слов изучаются статистические закономерности выполнения обобщенного свойства взаимодействия периодов. Мы ответим на следующие вопросы.

1. Чем определяется размерность расстановки?

2. Какова доля расстановок данной размерности r среди всех расстановок длины L с k джокерами и периодами p, q?

Третья глава посвящена изучению свойства взаимодействия локальных периодов частичных Z-слов. Изучаются специальные подпоследовательности.

1. Какой вид имеют специальные подпоследовательности?

2. Каковы необходимые и достаточные условия наличия в слове специальных подпоследовательностей?

3. В каком случае частичное Z-слово содержит хотя бы один бесконечный домен?

4. В каком случае частичное Z-слово содержит только бесконечные домены?

5°. Основные результаты диссертации

Первая глава диссертации посвящена изучению свойства взаимодействия периодов частичных слов (в «стандартной» формулировке).

В первом параграфе мы доказываем, что длина взаимодействия $L(k, p, q)$ существует для любых периодов p и q и количества джокеров k и получаем для длины взаимодействия следующую оценку.

Теорема 1.1. *Для любых взаимно простых $p > q \geq 2$ и любого $k \geq 0$*

$$L(k, p, q) \leq p + (k+1)q - 1.$$

Равенство достигается в двух случаях: когда $k = 0$ и когда $q = 2$ и p делит k.

Во втором параграфе мы находим точные оценки для длины взаимодействия в двух частных случаях: $q = 2$ и $k = 2$. Приведем здесь полученные результаты.

Теорема 1.2. *Для любого нечетного p и любого $k \geq 0$*

$$L(k, p, 2) = (2\lfloor k/p \rfloor + 1)p + k \bmod p + 1.$$

Теорема 1.3. *Для любых взаимно простых* $p > q \geq 2$

$$L(2, p, q) = p + 2q - 1.$$

При более пристальном изучении длины взаимодействия $L(k, p, q)$ как функции от k с параметрами p и q можно заметить, что она имеет, за исключением некоторого начального отрезка, регулярное поведение (см. рис. 1.4). Если отбросить рассмотренный в Теореме 1.2 случай $q = 2$ и случай маленьких k, то можно значительно улучшить оценку для длины взаимодействия из Теоремы 1.1. В следующей теореме, являющейся центральным результатом главы 1, эта длина представлена в виде суммы линейной и периодической функций относительно k. Такое представление позволяет получить максимально точную «общую» (т.е. пригодную для любых p и q) оценку длины взаимодействия.

Теорема 1.4. *Пусть p и q — взаимно простые натуральные числа, $p > q \geq 3$, а натуральное число k таково, что $k \geq \lfloor 2p/3 \rfloor - 1$ при q=3 и $k \geq \lfloor 3p/q \rfloor + 3$ при $q \geq 4$. Пусть $L(k, p, q)$ — длина взаимодействия периодов p и q при наличии k джокеров. Тогда*

$$L(k, p, q) = \frac{pq}{p + q - 2} \cdot k + \Delta(k, p, q),$$

где функция $\Delta(k, p, q)$ является периодической по k с периодом $p + q - 2$ и удовлетворяет следующим условиям:

1. *для произвольных p и q*

$$2q \leq \max(\Delta(k, p, q)) < 4(q - 1);$$

2. *для любого периода q и любого $\varepsilon > 0$ существует p такое, что*

$$\max(\Delta(k, p, q)) > 4(q - 1) - \varepsilon$$

3. *для произвольных p и q*

$$\min(\Delta(k, p, q)) = \frac{pq}{p + q - 2}.$$

Пример графика зависимости длины взаимодействия от числа джокеров, вместе с графиками верхней и нижней оценок из Теоремы 1.4 и графиком верхней оценки из Теоремы 1.1, приведен на рис. 1.4.

Следствие 1.1. *В условиях Теоремы 1.4 длина взаимодействия удовлетворяет неравенству*

$$L(k, p, q) < \frac{pq}{p+q-2} \cdot k + 4(q-1).$$

Функция, стоящая в правой части неравенства, является наилучшей верхней оценкой для $L(k, p, q)$ в классе линейных по k функций с параметрами p и q.

Результаты Теорем 1.1 и 1.4 дополняют друг друга. Верхняя оценка для длины взаимодействия Теоремы 1.1 применима при любом количестве джокеров $k \geq 0$ и любых значениях периодов. Оценка Теоремы 1.4 применима с некоторыми ограничениями, но она существенно точнее, является двусторонней, и позволяет локализовать длину взаимодействия внутри полосы, ширина которой меньше, чем $4q$.

Заданные в Теореме 1.4 ограничения на k могут быть незначительно понижены (не более чем на единицу в общем случае, и несколько сильнее, если отдельно рассматривать разные случаи соотношения между p и q). Однако, такое понижение, не меняя сути результата, привело бы к значительному удлинению доказательства.

Полученные результаты позволяют провести интересное сравнение свойств локальной и глобальной периодичности с точки зрения свойства взаимодействия периодов. В отличие от глобальной периодичности, для локальных периодов в общем случае нельзя указать оценку для длины взаимодействия. Для того, чтобы можно было оценить длину взаимодействия, необходимо рассматривать только те расстановки, в которых размещение пропущенных символов удовлетворяет некоторому достаточно обременительному ограничению (см. обсуждение на с.11). Из Теоремы 0.5 следует, что в этом случае взаимодействие локальных периодов гарантировано, если доля пропущенных символов в слове не превосходит величину $2/(p+q)$, т.е. обратно пропорциональна большему периоду. В то же время, оценка из Теоремы 1.4 показывает, что допустимая (для гарантированного взаимодействия периодов) доля пропущенных символов в периодическом частичном слове может быть существенно большей; а именно, величина $k/L \sim (p+q-2)/pq$ обратно пропорциональна меньшему периоду.

Параграфы с третьего по пятый посвящены доказательству теоремы 1.4. В третьем параграфе формулируется обратная задача: задана длина слова L и его периоды p, q, необходимо найти минимальное количество джокеров, при котором слово может *не иметь* периода 1. В четвертом параграфе приводится решение обратной задачи. Мы определяем *специальную* расстановку джокеров в слове, доказываем, что при определенных условиях специальная расстановка является оптимальной, т.е именно при такой расстановке достигается искомое минимальное количество джокеров. Затем мы определяем количество джокеров в специальной расстановке. В пятом параграфе мы получаем решение исходной задачи.

Вторая глава посвящена статистическим закономерностям взаимодействия периодов.

Пусть взаимно простые периоды p и q, $p > q$ фиксированы. Будем обозначать через $M(r, L, k)$ множество всех расстановок длины L, содержащих k джокеров и имеющих размерность r. Если какой-то параметр заменен знаком $\forall$, то это означает, что рассматривается объединение таких множеств по всем возможным значениям параметра.

Обозначим через $P(r, L, k)$ долю расстановок длины L, содержащих k джокеров и имеющих размерность r, т.е. отношение мощности множества $M(r, L, k)$ к мощности множества $M(\forall, L, k)$. Во второй главе описывается эффективный алгоритм для определения $P(r, L, k)$. Время работы «наивного» алгоритма экспоненциально зависит от длины слова; время работы нашего алгоритма полиномиально зависит от L и p и экспоненциально — от параметра q.

Основная идея алгоритма излагается в шестом параграфе. Она заключается в разбиении всех расстановок из множества $M(r, L, k)$ на классы и эффективном перечислении классов и подсчете количества расстановок в каждом классе. Для того, чтобы разбить множество расстановок $M(r, L, k)$ на классы, сопоставим каждой расстановке *характеристическую* расстановку длины pq. В шестом параграфе доказан следующий результат.

Предложение 2.1. *Размерность расстановки совпадает с размерностью ее характеристической расстановки.*

Таким образом, для определения $P(r, L, k)$ необходимо найти все характеристические расстановки размерности r, а затем для каждой такой характеристической расстановки определить количество соответствующих ей расстановок из множества $M(r, L, k)$.

В седьмом параграфе рассматривается задача вычисления мощности класса расстановок $M(r, L, k)$, которым соответствует данная характеристическая расстановка H. В случае $L = lpq$ эта мощность зависит только от L, k и количества единиц в H. В случае $L = lpq + g, 0 < g < pq$ мощность зависит еще и от количества единиц в первых g позициях характеристической расстановки H. В седьмом параграфе приведены формулы для вычисления данной мощности и итоговая формула для величины $P(r, L, k)$.

Восьмой параграф посвящен описанию алгоритма для определения $P(r, L, k)$. Вначале вычисляются вспомогательные таблицы (время выполнения этого этапа полиномиально зависит от L и p и экспоненциально — от параметра q), а затем по формуле вычисляется $P(r, L, k)$ (время выполнения этого этапа полиномиально зависит от L, p и q).

В девятом параграфе приводятся и анализируются результаты, полученные с помощью программы, реализующей данный алгоритм. Назовем *плотностью* джокеров в конечном частичном слове отношение количества джокеров к длине слова. Плотность джокеров в бесконечном частичном слове определим как верхний предел плотности джокеров в его подсловах длины n при $n \to \infty$. В частности, в девятом параграфе исследовалась зависимость доли неунарных расстановок (т.е. расстановок, размерность которых больше единицы) длины L с k джокерами от плотности джокеров в слове. Примеры графиков данной зависимости приведены на рис.2.2, 2.3.

Третья глава посвящена изучению свойства взаимодействия локальных периодов частичных Z-слов.

В десятом параграфе описывается используемая техника. Обычному Z-слову с периодами p, q можно сопоставить *граф локальной периодичности* – граф, вершины которого соответствуют позициям букв данного слова, а ребра соединяют вершины u, v такие, что разность номеров соответствующих позиций частичного слова составляет p или q. Этот граф можно изобразить на бесконечном цилиндре двумя системами спиралей (см.рис.3.2). Граф локальной периодичности частичного Z-слова получается из графа обычного слова удалением вершин, соответствующих позициям джокеров, и инцидентных им ребер. Множество доменов частичного слова совпадает с множеством компонент связности данного графа.

В одиннадцатом параграфе вводятся необходимые определения и доказываются некоторые свойства специальных конечных подпоследовательностей. Нас будут интересовать минимальные по включению специальные подпоследовательности. Такие подпоследовательности мы будем называть *разрезами* (т.к. множество позиций разреза частичного Z-слова соответствует вершинному разрезу графа локальной периодичности обычного Z-слова). В одиннадцатом параграфе приводится классификация разрезов.

Двенадцатый параграф посвящен описанию алгоритма проверки наличия в частичном слове специальной подпоследовательности. Алгоритм ищет позицию, *замыкающую разрез* (это наибольшая в некотором линейном порядке $\prec$ из позиций разреза).

На множестве позиций частичного слова определим следующие отношения. Отношение ρ: $x\rho y \Leftrightarrow W(x) = W(y)$. Отношение ϕ («быть соседом»): $x\phi y \Leftrightarrow (x - y = ap + bq, a, b \in \{-1, 0, 1\})$. Рассмотрим также отношение $\phi_1 = \phi \cap \rho$. Для каждой позиции x определим *метку* следующим образом

$$Label(x) = \begin{cases} \min_{\prec}\{y : y\phi_1 x_1 \phi_1 x_2 \phi_1 \ldots \phi_1 x_n \phi_1 x, & \\ \qquad\qquad\qquad\qquad y, x_1, \ldots, x_n \prec x\}, & W(x) = 0 \\ \qquad\qquad\qquad\qquad\qquad\qquad \infty, & W(x) = 1. \end{cases}$$

Предложение 3.4. *Позиция x, замыкающая разрез, имеет двух соседей y, z с одинаковой конечной меткой. При этом $y \prec x, z \prec x$.*

На основании предложения 3.4 в двенадцатом параграфе приводится описание полиномиального алгоритма с конечной памятью для проверки наличия в частичном слове специальных подпоследовательностей, а также дается оценка сложности работы этого алгоритма. В этом алгоритме позиции частичного слова рассматриваются в порядке $\prec$. Для текущей позиции вычисляется метка и проверяется выполнение условий из предложения 3.4. Если эти условия выполняются, то текущая позиция объявляется замыкающей разрез.

Теорема 3.1. *Время работы алгоритма проверки наличия в частичном слове специальных подпоследовательностей линейно зависит от длины исследуемого слова. Используемый объем памяти линейно зависит от периодов p, q.*

В последнем параграфе третьей главы приводятся две модификации алгоритма из предыдущего параграфа. Первая из них позволяет определить количество доменов в конечном частичном слове. Вторая модификация алгоритма предназначена для поиска специальных подпоследовательностей особого вида — *преград*.

Таковы основные результаты диссертации. Они докладывались на международных конференциях «Mathematical Foundations of Computer Science» (Марианские Лазни, 2001), «WORDS'03» (Турку, 2003), российской конференции «Дискретный анализ и исследование операций» (Новосибирск, 2004), международной алгебраической конференции (Екатеринбург, 2005) и научных семинарах «Алгебраические системы» (УрГУ) и «Дискретная математика» (УрГУ). Результаты опубликованы в работах [38–45]. В совместных работах [38–41] руководителю принадлежат постановка задачи и общая методика исследований, доказательства всех основных утверждений принадлежат автору.

Работа выполнена под руководством доцента Арсения Михайловича Шура и доцента Евгения Витальевича Суханова, которым автор выражает глубокую благодарность за постоянное внимание и всестороннюю поддержку.

Глава I

Свойство взаимодействия периодов

В данной главе изучается свойство взаимодействия периодов частичных слов. Основными результатами данной главы являются теорема о существовании длины взаимодействия $L(k, p, q)$ для любых периодов p, q и количества джокеров k (§1) и теорема, дающая точную оценку длины взаимодействия при некоторых ограничениях (§3).

§1 Существование длины взаимодействия

В данном параграфе мы доказываем, что длина взаимодействия $L(k, p, q)$ существует для любых периодов p и q и количества джокеров k и получаем для длины взаимодействия линейную верхнюю оценку.

§1.1 Используемая техника

При исследовании частичных слов удобно рассматривать бинарные отношения и их графы. Пусть частичное слово U длины L, имеющее периоды p, q, содержит k джокеров в позициях $i_1, \ldots, i_k$. Будем соотносить с этим словом бинарные отношения R_p и R_q на множестве $D(U)$. Отношение R_p (отношение R_q) состоит из всех пар (s, t), где $s \equiv t \pmod p$ (соответственно, $s \equiv t \pmod q$) и между s и t нет такого $n \in D(U)$, что $s \equiv n \pmod p$ (соответственно, $s \equiv n \pmod q$). Отношения R_p, R_q являются симметричными, поэтому пару $(D(U), R_p \cup R_q)$ можно рассматривать как неориентированный граф. Этот граф мы будем называть *графом периодичности*. Отношение $(R_p \cup R_q)^*$ (рефлексивно-транзитивное замыкание отношения $R_p \cup R_q$) является отношением эквивалентности. По определению периода в U на эквивалентных позициях обязательно стоят одинаковые буквы. Отношению эквивалентности R^* соответствует некоторое разбиение множества $D(U)$. Количество классов этого разбиения равно количеству компонент связности соответствующего графа и равно размерности слова U.

Пример 1.1. Пусть $n = 9, k = 3, i_1 = 2, i_2 = 4, i_3 = 6, q = 3, p = 5$. Тогда $U = a_1 \diamond a_3 \diamond a_5 \diamond a_7 a_8 a_9$. Граф обычного слова с такими периодами и граф U выглядят следующим образом:

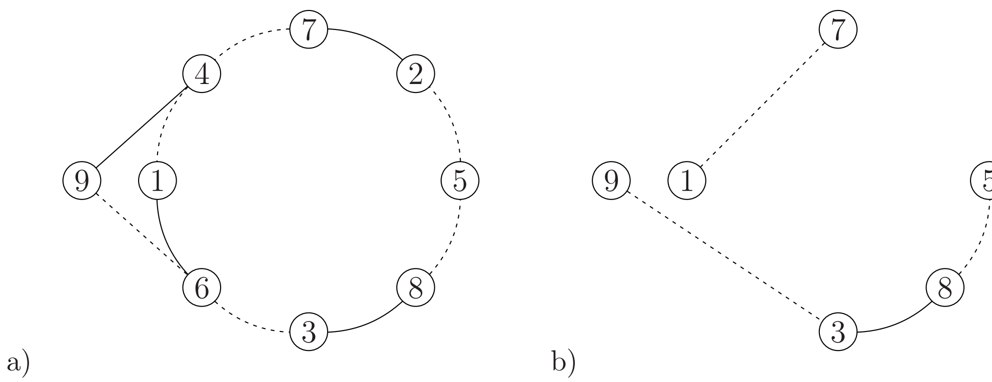


Рис. 1.1: Графы обычного и частичного слова. Сплошные линии обозначают ребра, соответствующие отношению R_p, штриховые — отношению R_q.

Таким образом, максимальное возможное количество различных букв в U равно 2 (компоненты связности соответствующего графа $\{1, 7\}$ и $\{3, 5, 8, 9\}$).

§1.2 Теорема о существовании длины взаимодействия

Напомним, что мы рассматриваем взаимно простые периоды $p, q, 2 \le q < p$. Следующая теорема утверждает, что длина взаимодействия $L(k, p, q)$ существует для любых периодов p, q и любого количества джокеров k.

Теорема 1.1. *Для любых взаимно простых $p > q \ge 2$ и любого $k \ge 0$*

$$L(k, p, q) \le p + (k + 1)q - 1.$$

Равенство достигается в двух случаях: когда $k = 0$ и когда $q = 2$ и p делит k.

Замечание 1.1. *Из Теоремы 1.1 следует, что для любого k, не являющегося степенью двойки, можно подобрать такие взаимно простые натуральные p и q, чтобы существовало частичное слово длины $p+(k+1)q-2$, имеющее периоды p и q, но не имеющее периода 1. В этом случае можно положить p равным некоторому нечетному делителю k, а q равным 2.*

Замечание 1.2. *Для обычных слов ($k = 0$) Теорема 1.1 дает ту же оценку, что и теорема Файна-Вильфа для взаимно простых периодов.*

Для доказательства первого утверждения Теоремы 1.1 необходимы две леммы. Первая из них является аналогом Теоремы 0.4 для глобальных периодов.

Лемма 1.1. *Рассмотрим частичное слово U длины $p+q$, имеющее периоды p и q. Пусть джокеры в U расположены следующим образом — ровно один джокер в позициях $1, \dots, q, p+1, \dots, p+q$ и любое количество джокеров в позициях $q+1, \dots, p$. Тогда U имеет период 1.*

Доказательство. Рассмотрим слово без джокеров длины $p+q-1$, имеющее периоды p, q. По теореме Файна-Вильфа это слово имеет период 1, поэтому соответствующий граф является связным. Рассмотрим,какие вершины могут быть смежными с некоторой вершиной t. Поскольку длина слова меньше, чем $p+q+1$, возможны следующие случаи: 1)$t+q, t+p$; 2)$t+q$; 3)$t+q, t-q$; 4)$t-q$; 5)$t-q, t-p$. Случаи 2) и 4) описывают вершины q и p соответственно; любая другая вершина имеет две смежных. Так как граф связный, то он является цепью.

Добавим в граф вершину $(p+q)$. Она смежна вершинам p и q. Таким образом, полученный граф является простым циклом.

Исследуем процесс размещения в слове джокеров. Когда мы ставим джокер в позицию t, $q+1 \leq t \leq p$, граф изменяется следующим образом: удаляются вершина t и ребра $(t-q, t)$ и $(t, t+q)$, а вершины $t-q$ и $t+q$ соединяются ребром. Таким образом, полученный граф тоже является простым циклом, причем он будет оставаться таковым при размещении в слове любого количества джокеров в позициях $q+1, \dots, p$. Размещение джокера в префиксе или суффиксе длины q означает удаление соответствующей вершины из графа. Таким образом, полученный граф (граф U) является цепью, т.е. связен. Следовательно, U имеет период 1. □

Наблюдение 1.1. *Граница $p+q$ является точной. Это связано с тем, что граф слова длины $p+q-1$ является цепью, поэтому джокер в любой из позиций $1, \dots, q-1, p+1, \dots, p+q-1$ разбивает эту цепь на две части.*

В соответствии с Леммой 1.1 мы будем называть джокер *существенным* для слова U длины $p+q$, имеющего периоды p, q, если он расположен в префиксе или суффиксе длины q.

Лемма 1.2. *Рассмотрим частичное слово U, имеющее периоды p и q. Если в этом слове есть подслово длины $p+q$, в котором содержится только один существенный джокер, то U имеет период 1.*

Доказательство. Предположим, что это подслово расположено в U на участке $t+1, \ldots, t+p+q$. По Лемме 1.1 буквы $U(t+1), \ldots, U(t+p+q)$ равны некоторой букве (назовем ее a). Тогда достаточно показать, что любая буква из U равна a. Рассмотрим позицию s, не входящую в выбранный сегмент и пометим все позиции, равные s по модулю q. Тогда помеченными окажутся одна из позиций $t+1, \ldots, t+q$ и одна из позиций $t+p+1, \ldots, t+p+q$. По условию только одна из них может быть занята джокером; значит, хотя бы в одной из этих позиций стоит буква a. Все буквы в помеченных позициях должны быть равны. Таким образом, $U(s) = a$. □

Доказательство Теоремы 1.1. Докажем, что любое частичное слово с периодами p и q, содержащее k джокеров и не имеющее периода 1, короче, чем приведенная граница. По Лемме 1.2 такое слово должно содержать по крайней мере 2 существенных джокера в каждом подслове длины $p+q$. Каждое подслово длины $p+q$ содержит $2q$ позиций, в которых могут стоять существенные джокеры. Поэтому любой джокер может быть существенным только для $2q$ подслов. Более того, первый в слове джокер не может стоять на последнем месте в подслове, содержащем два джокера, поэтому он может являться существенным только для $2q-1$ подслов. То же самое по симметрии выполняется и для последнего джокера в слове.

Поскольку каждое подслово длины $p+q$ должно содержать по меньшей мере два существенных джокера, количество подслов длины $p+q$ в слове не должно превышать $(2qk-2)/2$, т.е. $qk-1$. С другой стороны, слово длины L содержит $L-p-q+1$ таких подслов. Таким образом,

$$L - p - q + 1 \leq qk - 1,$$

откуда следует, что $L \leq p + (k+1)q - 2$. Следовательно, частичное слово длины не менее $p + (k+1)q - 1$ имеет период 1. Первое утверждение Теоремы 1.1 доказано.

Теперь докажем второе утверждение. При $k = 1$ оценку можно улучшить до $p + q$ по Лемме 1.1. Поэтому будем считать, что $k \geq 2$. Прежде всего докажем, что оценка может быть точной только для $q = 2$. Предположим противное: $q \geq 3$ и существует частичное слово W длины $p + (k + 1)q - 2$, имеющее периоды p и q и содержащее k джокеров, но не имеющее периода 1.

Рассмотрим в W подслова длины $p + q$. Поскольку длина W равна $p + (k + 1)q - 2$, слово W содержит $qk - 1$ подслов длины $p + q$. Т.к. W не имеет периода 1, по лемме 1.2 каждое подслово W длины $p + q$ содержит не менее 2 существенных джокеров. Это возможно только в случае, когда каждый джокер является существенным для максимально возможного количества подслов: первый и последний джокеры являются существенными для $2q-1$ подслов, остальные джокеры — для $2q$ подслов (см. доказательство первого утверждения). Поэтому два первых джокера должны находиться в позициях $p + q - 1$ и $p + q$. Кроме того, каждое подслово должно содержать ровно два джокера.

Можно представить, что мы рассматриваем слово W через окно, размер которого $p + q$ символов. Окно разделено на три части, размер которых q, $p - q$ и q символов соответственно (см. Рис. 1.2а,б). В начале окно установлено в префиксе W. Далее окно посимвольно передвигается вправо. На каждом шаге в боковых частях окна должно находиться ровно два джокера.

Два первых джокера изначально находятся в двух крайних справа позициях окна и при передвижении окна вправо сдвигаются влево. Третий джокер должен появиться в окне в тот момент, когда первый джокер попадает в центральную часть окна (Рис. 1.2а). Таким образом, третий и четвертый джокеры находятся в слове W в позициях $p + 2q - 1$,$p + 2q$ (эти позиции существуют, так как $|W| \geq p + 3q - 2$ и между вторым и третьим джокером есть хотя бы одна буква, поскольку $q \geq 3$).

Пока первый джокер проходит центральную часть окна, несколько раз может повториться подобная ситуация: когда текущая пара джокеров попадает в центральную часть, другая пара появляется справа на расстоянии q от текущей пары.

Когда первая пара джокеров попадает в левую часть окна, другая пара должна переходить в центральную часть (Рис. 1.2б). Пусть V — слово, которое в этот момент находится в окне. С одной стороны, в V имеются джокеры, расположенные в позициях q, $q + 1$ и p, $p + 1$. С другой стороны, расстояние между первыми джокерами этих пар кратно q. Тогда q делит $p - q$, что противоречит предположению НОД$(p, q) = 1$.

Теперь предположим, что $q = 2$. Первые два джокера находятся в a)

b)

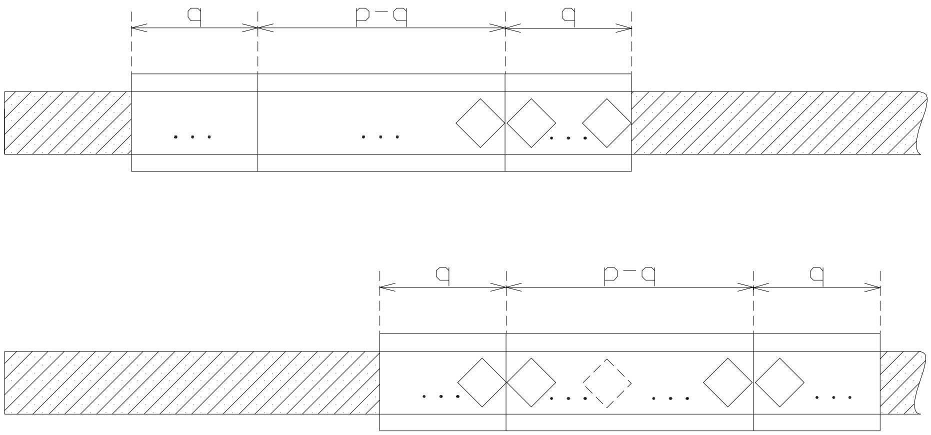


Рис. 1.2: Движение окна по слову.

позициях $p+1$ и $p+2$. В рассмотренном ранее случае пары джокеров в W были разделены хотя бы одной буквой. Теперь в правой части окна только два символа, поэтому третий джокер возникает сразу после второго. Более того, можно заметить, что W содержит p последовательных джокеров в позициях $p+1,\dots,2p$. За ними следуют p букв, затем снова p джокеров и т.д. Таким образом, если $k = sp$ для некоторого s, то длина W равна $(2s+1)p$, что совпадает с предложенной оценкой $p+(k+1)q-2$. С другой стороны, если $k \neq sp$, то символ, следующий за k-ым джокером в W, тоже должен быть джокером. Таким образом, последним символом W должен быть джокер. Этот джокер является существенным только для одного подслова длины $p+q$, в то время как он должен являться существенным для $2q-1$ подслов.

Мы доказали, что необходимые условия могут выполняться только при $q = 2$ и $k = sp$ для некоторого s. Теперь докажем, что слово W длины $(2s+1)p$, имеющее периоды 2 и p и содержащее sp джокеров, расположенных описанным выше способом, содержит по меньшей мере две различных буквы. Как уже было доказано, джокеры в W должны быть расположены следующим образом:

$$W = a_1a_2\dots a_p\underbrace{\diamond\diamond\dots\diamond}_{p}a_{2p+1}\dots a_{(2s-1)p}\underbrace{\diamond\diamond\dots\diamond}_{p}a_{2sp+1}a_{2sp+2}\dots a_{(2s+1)p}.$$

В соответствующем графе ребро соединяет вершины, равные по модулю 2 или по модулю p. В слове W на любых двух позициях, находящихся на расстоянии p друг от друга, расположены одна буква и один джокер.

Таким образом, одинаковые буквы находятся на расстоянии, кратном $2p$. Это означает, что граф имеет две компоненты связности, состоящие из вершин с четными и нечетными номерами соответственно. Следовательно, в W могут присутствовать две различные буквы. Таким образом, Теорема 1.1 доказана. $\square$

§2 Точные оценки в частных случаях

В данном параграфе мы находим точные оценки для длины взаимодействия в двух важных частных случаях: $q = 2$ и $k = 2$.

Теорема 1.2. *Для любого нечетного p и любого $k \geq 0$*

$$L(k, p, 2) = (2\lfloor k/p \rfloor + 1)p + k \bmod p + 1.$$

Доказательство. Сначала рассмотрим случай $k = sp$ для некоторого $s \in \mathbb{N}$. Согласно теореме 1.1 в этом случае $L(k, p, 2) = p + 2k + 1$, что совпадает с доказываемой оценкой. Таким образом, существует слово W длины $(L(k, p, 2) - 1) = (p + 2k)$ с k джокерами и периодами 2 и p, не имеющее периода 1. Из доказательства теоремы 1.1 следует, что джокеры в W должны быть расположены следующим образом:

$$W = a_1 a_2 \ldots a_p \underbrace{\diamond\diamond \ldots \diamond}_{p} a_{2p+1} \ldots a_{(2s-1)p} \underbrace{\diamond\diamond \ldots \diamond}_{p} a_{2sp+1} a_{2sp+2} \ldots a_{(2s+1)p}.$$

При этом каждый джокер является существенным для максимально возможного количества подслов длины $p + q$: первый и последний джокеры являются существенными для $2q - 1$ подслов, остальные джокеры — для $2q$ подслов.

Теперь рассмотрим случай $k = sp + t, 0 < t < p$. Как следует из доказательства теоремы 1.1, в этом случае невозможно расставить джокеры так, что каждый джокер будет существенным для максимально возможного количества подслов длины $p+q$. Рассмотрим слово W длины $L(k, p, 2)-1$ с k джокерами и периодами 2 и p, не имеющее периода 1. В этом слове джокеры должны быть расставлены так, чтобы как можно больше джокеров являлись существенными для максимально возможного количества подслов длины $p + q$. Для этого sp джокеров должны быть расставлены в подслове длины $p+2sp$ так же, как в предыдущем случае. Оставшиеся t джокеров должны стоять в позициях, соседних с этим подсловом, так как иначе в слове возникнет подслово длины $p+q$, в котором менее 2 существенных джокеров. Тогда

$|W| = (2s+1)p + t = (2\lfloor k/p \rfloor + 1)p + k \bmod p$, а расстановка джокеров в слове W, например, такая:

$$W = a_1 a_2 \ldots a_p \underbrace{\diamond\diamond\ldots\diamond}_{p} a_{2p+1} \ldots a_{(2s-1)p} \underbrace{\diamond\diamond\ldots\diamond}_{p} a_{2sp+1} a_{2sp+2} \ldots a_{(2s+1)p} \underbrace{\diamond\diamond\ldots\diamond}_{t}.$$

□

Теорема 1.3. *Для любых взаимно простых $p > q \geq 2$*

$$L(2, p, q) = p + 2q - 1.$$

Доказательство. Вначале докажем, что $L(2, p, q) \leq p + 2q - 1$. Рассмотрим частичное слово U длины $p+2q-1$, имеющее периоды p и q и содержащее 2 джокера. Докажем, что U имеет период 1. Для этого рассмотрим возможные варианты расстановки джокеров в слове U.

1. Предположим, что в U имеется джокер в одной из позиций $q + 1, \ldots, p, p + q + 1, \ldots, p + 2q - 1$. Тогда в префиксе слова U длины $p + q$ находится не больше одного существенного джокера. Аналогично, если джокер расположен в одной из позиций $1, \ldots, q - 1, 2q, \ldots, p + q - 1$, то в суффиксе находится не больше одного существенного джокера. По Лемме 1.2 в этих случаях U имеет период 1.

2. Предположим, что в U имеется джокер в позиции t, причем $p+1 \leq t < 2q$ (этот случай невозможен при $p + 1 \geq 2q$). Тогда для подслова длины $p+q$, начинающегося с позиции $t-p+1$, этот джокер не является существенным, так как расположен в p-ой позиции. По Лемме 1.2 U в этом случае имеет период 1.

3. Осталась одна возможность: джокеры расположены в q-ой и $(p+q)$-ой позициях. Необходимо доказать, что в этом случае граф U является связным. Как уже указывалось, граф обычного слова длины $p + q$, имеющего периоды p и q, является простым циклом (см. доказательство Леммы 1.1). Вершины q и $p + q$ являются смежными. Поэтому при удалении этих вершин из графа (т.е. размещении джокеров в позициях q и $p + q$) граф останется связным. Таким образом, префикс U длины $p + q$ является связным. Симметрично, связным будет и граф суффикса. Эти подслова имеют $p+1$ общих символов, т.е. по крайней мере одну общую букву. Отсюда следует, что граф U является связным.

Теперь докажем, что $L(2, p, q) \geq p+2q-1$. Расмотрим слово W длины $p+2q-2$, имеющее периоды p и q и содержащее два джокера в позициях $q-1$,q (см. пример на Рис.1.3). Граф префикса длины $p + q$ получен из простого цикла удалением двух несмежных вершин, поэтому он состоит из двух цепей, не связанных друг с другом. Вершина $p+q+1$ смежна с вершинами $p+1$ и $q+1$. Эти вершины смежны вершине 1, поэтому они и так находятся в одной компоненте. Следовательно, полученный граф по-прежнему имеет две компоненты связности. Аналогично добавление вершин $p+q+2, \ldots, p+2q-2$ не уменьшает количество компонент связности. Только добавление вершины $p+2q-1$ делает граф связным, так как она соединяет вершины $p+q-1$ и $2q-1$, которые не были связаны, поскольку $W(q-1) = \diamond$. □

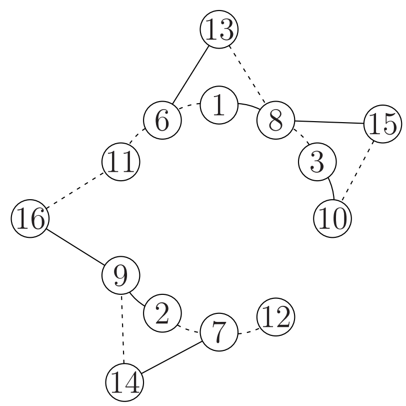


Рис. 1.3: Граф частичного слова, имеющего периоды $q = 5, p = 7$, длины $p+2q-1 = 16$. Удаление вершины 16 разбивает граф на две компоненты.

§3 Формулировка прямой и обратной задачи. Бланки и расстановки

Если отбросить рассмотренный в Теореме 1.2 случай $q = 2$ и случай маленьких k, то можно значительно улучшить оценку для длины взаимодействия из Теоремы 1.1. В следующей теореме, являющейся центральным результатом главы 1, эта длина представлена в виде суммы линейной и периодической функций относительно k. Такое представление позволяет получить максимально точную «общую» (т.е. пригодную для любых p и q) оценку длины взаимодействия.

Теорема 1.4. *Пусть p и q — взаимно простые натуральные числа, $p > q \geq 3$, а натуральное число k таково, что $k \geq \lfloor 2p/3 \rfloor - 1$ при q=3 и $k \geq \lfloor 3p/q \rfloor + 3$ при $q \geq 4$. Пусть $L(k, p, q)$ — длина взаимодействия периодов p и q при наличии k джокеров. Тогда*

$$L(k,p,q) = \frac{pq}{p+q-2} \cdot k + \Delta(k,p,q),$$

где функция $\Delta(k,p,q)$ является периодической по k с периодом $p+q-2$ и удовлетворяет следующим условиям:

1. *для произвольных p и q*

$$2q \le \max(\Delta(k,p,q)) < 4(q-1);$$

2. *для любого периода q и любого $\varepsilon > 0$ существует p такое, что*

$$\max(\Delta(k,p,q)) > 4(q-1) - \varepsilon$$

3. *для произвольных p и q*

$$\min(\Delta(k,p,q)) = \frac{pq}{p+q-2}.$$

Пример 1.2. График зависимости длины взаимодействия от числа джокеров при $p = 13$, $q = 3$, вместе с графиками верхней и нижней оценок из теоремы 1.4, приведены на рис. 1.4. Отметим, что, согласно теореме 1.4, эти оценки верны при $k \ge 7$. Штриховая линия соответствует верхней оценке из теоремы 1.1. Эта оценка верна при любом k.

Далее в первой главе мы всюду считаем, что $p > q \ge 3$. Для доказательства теоремы 1.4 рассмотрим обратную задачу:

задана длина слова L и его периоды p, q, необходимо найти минимальное количество джокеров, при котором слово может не иметь периода 1.

Заметим, что для каждого периода p произвольного слова U и для любого i, $1 \le i \le p$, все буквы, находящиеся в позициях $i, i+p, i+2p, \ldots$ равны независимо от расположения джокеров в этих позициях. Указанное множество позиций мы будем называть p-классом; оно содержится в i-ом классе вычетов по модулю p.

Мы рассматриваем случай двух различных периодов p, q у слова U, и, соответственно, будем рассматривать p- и q-классы. Заметим, что если две позиции из разных q-классов, занятые буквами, принадлежат одному p-классу (т.е. «связаны» при помощи периода p), то в этих q-классах содержится одна и та же буква. Если же любая пара позиций из данных q-классов, принадлежащая одному p-классу, содержит хотя бы один джокер, то эти q-классы могут содержать различные буквы.

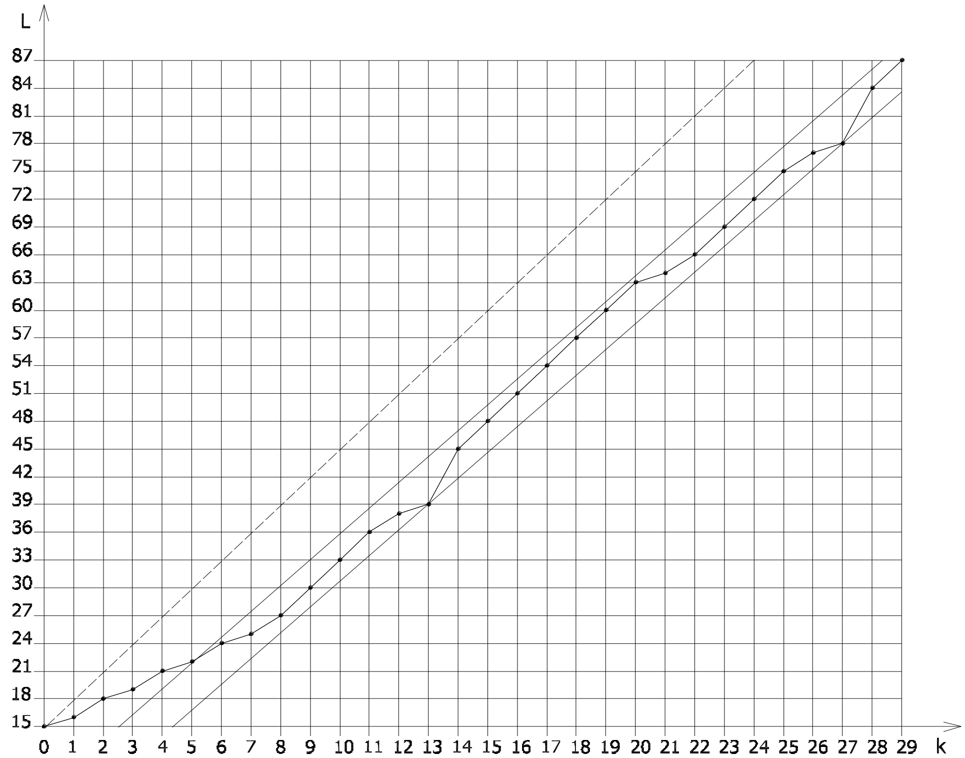


Рис. 1.4: *график зависимости длины взаимодействия от числа джокеров.*

Таким образом, чтобы построить слово U длины L с периодами p и q, содержащее m различных букв ($m \leq q$), нужно разбить q-классы (их q штук) на m групп и разместить джокеры так, чтобы период p связывал между собой только такие позиции, не занятые джокерами, которые принадлежат q-классам из одной группы.

Чтобы решить поставленную обратную задачу, надо найти наиболее экономный способ размещения джокеров для случая $m = 2$. Процедура построения слова U в этом случае будет выглядеть следующим образом. Разделим q-классы на 2 группы: T и S, содержащие соответственно t и s классов ($0 < t \leq s,\ t + s = q$). Договоримся отождествлять номера q-классов и сами q-классы. В частности, нам будет удобнее относиться к элементам множеств S и T как к номерам. Для каждого p-класса выберем одну из групп T или S и разместим джокеры во всех позициях этого p-класса, принадлежащих к q-классам из выбранной группы. Если мы хотя бы по разу выбрали каждую из групп, то в q-классах как из T, так и из S останутся позиции, не занятые джокерами. В них мы разместим буквы (одну для T, другую для S) и, тем самым, получим слово U, не имеющее периода 1. Несложно заметить, что число джокеров в U, которое мы стремимся минимизировать, зависит от t и от способа выбора групп для q-классов.

Опуская детали, вышеизложенную процедуру можно представить так: имеется некоторый «шаблон» слова U, в котором мы по определённым правилам размещаем алфавитные символы (вначале — джокеры, а затем — буквы). Интуитивное понятие «шаблона» мы сейчас оформим в виде специальной числовой таблицы — *бланка*, играющей ключевую роль в дальнейших рассмотрениях.

Запишем позиции слова U в таблицу, как показано на рис. 1.5a).

$$\begin{array}{cccc} 1 & p+1 & 2p+1 & \dots \\ 2 & p+2 & 2p+2 & \dots \\ \vdots & \vdots & \vdots & \vdots \\ p & 2p & 3p & \dots \end{array} \qquad \begin{array}{cccc} 1 & 1 \oplus (p \bmod q) & 1 \oplus (2p \bmod q) & \dots \\ 2 & 2 \oplus (p \bmod q) & 2 \oplus (2p \bmod q) & \dots \\ \vdots & \vdots & \vdots & \vdots \\ p \bmod q & 2p \bmod q & 3p \bmod q & \dots \end{array}$$

a) b)

Рис. 1.5: *построение бланка.*

Занумеруем q-классы числами от 0 до $q-1$ (так, что i-ый класс содержит позиции с номерами вида $i + nq$) и заменим в приведённой таблице каждую позицию на номер содержащего её q-класса. Полученную таблицу будем называть *бланком*. Она имеет вид, изображённый на рис. 1.5б) ($\oplus$ обозначает сложение по модулю q).

Для построения слова U мы заменяем некоторые номера в строках (строки соответствуют p-классам) джокерами так, чтобы в любой строке оставались номера классов только из группы S или только из группы T.

Пример 1.3. Пусть имеется слово длины 15, имеющее периоды 5,7. Минимальное количество джокеров, при котором такое слово может не быть унарным, равно 2 (Это следует из теоремы 1.3). Бланк для этого слова выглядит следующим образом (рис. 1.6а):

Легко убедиться, что наименьшее количество джокеров достигается при $T = \{2, 4\}$ и $S = \{0, 1, 3\}$. Тогда нам необходимо поставить по одному джокеру в 4-ую и 5-ую строки, например, в позиции 4 и 5 (рис. 1.6б). Теперь в каждой строке находятся номера q-классов, принадлежащих одной группе. Заменив номера из S на a, номера из T на b и записав буквы в обычном порядке, получим бинарное слово $aba\diamond\diamond ababaababa$.

Замечание 1.3. *В связи с рассматриваемой задачей нас интересует только стадия размещения джокеров в бланке (и последующий подсчёт джокеров). Стадию размещения букв мы будем опускать.*

1	3	0
2	4	
3	0	
4	1	
0	2	
1	3	
2	4	

a)

1	3	0
2	4	
3	0	
$\diamond$	1	
$\diamond$	2	
1	3	
2	4	

b)

Рис. 1.6: *пример бланка и расстановки джокеров в нем.*

Замечание 1.4. *Параметры L, p, q ($L > p > q > 1$, НОД$(p, q) = 1$) однозначно задают бланк. Поэтому бланк будем обозначать через $B(L, p, q)$.*

Расстановкой джокеров в слове (бланке), или просто *расстановкой* будем называть множество позиций в слове или в бланке, на которых стоят джокеры. Пусть заданы периоды p, q и длина L. Расстановку джокеров в бланке $B(L, p, q)$ будем называть *допустимой*, если слово длины L с периодами p, q и указанной расстановкой джокеров может не иметь периода 1 (т.е. содержать две различные буквы). Допустимую расстановку с минимальным количеством джокеров будем называть *оптимальной*, а число джокеров в ней обозначать через k^*. Рассматриваемая нами обратная задача состоит, таким образом, в вычислении k^*.

Для данного t (напомним, что $t = |T|$) рассмотрим все допустимые расстановки джокеров и выберем расстановку с минимальным количеством джокеров. Эту расстановку будем называть *оптимальной при* t, а количество джокеров в ней обозначать через $k^*(t)$.

Будем называть *специальной* расстановку следующего вида: $|T| = 1$, во всех строках, кроме одной, джокерами заменяются все вхождения элемента из T и только они, а в оставшейся строке — все вхождения элементов из S и только они. Специальная расстановка, очевидно, является допустимой.

Лемма 1.3. *При $q \geq 3$ и $L \geq 2p-2$ существует специальная расстановка в бланке $B(L, p, q)$, оптимальная при $t = 1$.*

Доказательство. В бланке $B(L, p, q)$ не более двух одноэлементных строк. Если такие строки есть, поместим в множество T q-класс, в них не встречающийся (такой найдётся, поскольку q-классов не менее трёх); если таких строк нет, поместим в T любой q-класс. Теперь в каждой строке элементов из T содержится не больше, чем элементов из S. Для фиксированного одноэлементного T рассмотрим допустимую расстановку, в которой джокеры заменяют элементы S более, чем в одной строке. Заменив в одной из строк джокерами элементы T вместо элементов S, вновь получим допустимую расстановку; количество джокеров при этом может лишь уменьшиться. Таким образом, при каждом фиксированном одноэлементном T некоторая специальная расстановка содержит минимальное количество джокеров. Следовательно, найдётся специальная расстановка, которая является оптимальной при $t = 1$. □

Следствие 1.1. *Для любого бланка $B(L, p, 3)$ такого, что $L \geq 2p - 2$, существует оптимальная специальная расстановка.*

Доказательство. Условие $q = 3$ необходимо влечёт $t = 1$. Таким образом, специальная расстановка, оптимальная при $t = 1$ (такая существует по лемме 1.3), является оптимальной. □

Следующее предложение, позволяющее решить обратную задачу, является ключевым в данной главе.

Предложение 1.1. *Для любого бланка $B(L, p, q)$ такого, что $q \geq 4$, $L \geq 3p + q$, существует оптимальная специальная расстановка.*

Замечание 1.5. *В примере 1.3 выполняется $k^*(1) = 3$, и, таким образом, никакая специальная расстановка не оптимальна (но длина бланка в этом примере не удовлетворяет приводимым в предложении 1.1 ограничениям).*

§4 Решение обратной задачи

Данный параграф посвящён доказательству предложения 1.1, а также решению обратной задачи на основе подсчёта джокеров в оптимальной специальной расстановке.

Для доказательства предложения 1.1 разделим бланк на части следующим образом (см. рис. 1.7).

Часть бланка, соответствующую префиксу слова длины $pq \cdot \left\lfloor \frac{L}{pq} \right\rfloor$, будем называть *телом*, а оставшуюся часть — *хвостом*. Тело состоит из $\left\lfloor \frac{L}{pq} \right\rfloor$ блоков — прямоугольников высоты p и ширины q — это части **1**.

В хвосте выделим части **2** — это прямоугольники высоты q, ширина которых равна $\left\lfloor \frac{L \bmod pq}{p} \right\rfloor$ или $\left\lceil \frac{L \bmod pq}{p} \right\rceil$. Таких частей $\left\lfloor \frac{p}{q} \right\rfloor$ штук. Кроме того, в бланке присутствуют одна часть **3** и одна часть **4**. Часть **3** —

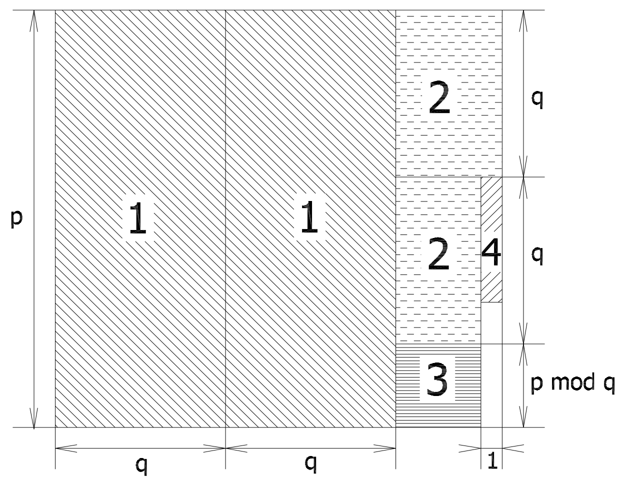


Рис. 1.7: *части бланка.*

прямоугольник ширины $\left\lfloor \frac{L \bmod pq}{p} \right\rfloor$ и высоты $(p \bmod q)$. Часть **4** — прямоугольник ширины 1 и высоты $(L \bmod q)$, соответствующий суффиксу слова длины $(L \bmod q)$.

Определим *расстановку для части* **i** как пересечение части **i** и расстановки в данном бланке. Определения допустимой и оптимальной (при t) расстановок для части **i** аналогичны соответствующим определениям для расстановок. Количество джокеров в оптимальной при t для части **i** расстановке будем обозначать через ${k_i}^*(t)$. Очевидно, что сумма значений ${k_i}^*(t)$ по всем частям бланка не превосходит $k^*(t)$.

Для доказательства предложения 1.1 рассмотрим отдельно количество джокеров в расстановках, оптимальных для каждой части. Лемма 1.4 посвящена исследованию части **1**, леммы 1.5-1.7 — части **2**, лемма 1.8 — части **3**. Рассмотрение расстановок для части **4** не представляет интереса: любая строка в этой части содержит только один элемент, а значит, оптимальная расстановка при любом t вообще не содержит джокеров.

Лемма 1.4. *Для любого* $2 \leq t \leq \lfloor q/2 \rfloor$ *справедливо*

$$ {k_1}^*(t) - {k_1}^*(1) = (p-2)(t-1). $$

Доказательство. Рассмотрим часть **1**. Каждая строка блока является перестановкой множества $\{0, \ldots, q-1\}$. В самом деле, последовательность элементов первой строки имеет вид

$$ \pi = (1,\ 1 \oplus (p \bmod q),\ 1 \oplus (2p \bmod q), \ldots, 1 \oplus ((q-1)p \bmod q)); $$

все элементы π различны, так как *НОД*$(p, q) = 1$, а значит, π является перестановкой множества $\{0, \ldots, q-1\}$. Остальные же строки являются циклическими перестановками π. Итак, в каждой строке блока номер каждого q-класса встречается ровно один раз. Таким образом, каждая строка содержит t элементов из T и s элементов из S.

Хотя бы в одной строке мы должны разместить t джокеров, хотя бы в одной — s джокеров, и в каждой из оставшихся — либо t, либо s джокеров. Поскольку $t \leq s$, то оптимальная расстановка для части **1** следующая: по t джокеров в $p-1$ строке, и $s = q - t$ джокеров — в оставшейся строке. Тогда общее количество джокеров в части **1** при оптимальной для неё расстановке равно

$$k_1{}^*(t) = (p-1)t + (q-t).$$

При $t \geq 2$ в части **1** необходимо поставить на $k_1{}^*(t) - k_1{}^*(1) = (p-2)(t-1)$ джокеров больше, чем при $t = 1$.

□

Перейдем к рассмотрению части **2**.

Отличительная особенность частей, принадлежащих хвосту бланка, состоит в том, что они пересекают не все строки бланка (т.е. не все p-классы). Это значит, что в расстановке, оптимальной для такой части, джокеры могут во всех строках заменять только элементы из T: строки, в которых джокерами заменяются элементы из S, обязательно присутствуют в бланке, но могут не пересекать данную часть. Наличие таких строк будет учтено ниже, в доказательстве предложения 1.1.

Обозначим ширину части **2** через w $(1 \leq w \leq q)$. Для удобства рассмотрения дополним эту часть бланка до квадрата $q \times q$ (см. рис. 1.8). Тогда в первой строке полученной таблицы будет находиться рассмотренная выше перестановка π, а в остальных строках — все ее различные циклические перестановки (ср. доказательство леммы 1.4). Заштрихованные клетки на рисунке содержат элементы из T. Ясно, что для произвольного бланка перестановкой строк можно добиться того, чтобы эти элементы размещались «лесенкой», как в примере на рис. 1.8. Ниже мы будем считать, что в рассматриваемой части вида **2** такая перестановка строк произведена.

Множеству T поставим в соответствие слово Z_T длины q над алфавитом $\{0, 1\}$, которое получается из первой строки «дополненной» части **2** заменой элементов T на 1 и элементов S на 0. Обозначим через $Z_T(w, j)$ префикс длины w циклической перестановки слова Z_T, начинающейся с j-го символа. Слова $Z_T(w, j)$ соответствуют всем q строкам рассматриваемой части **2**. *Характеристическим числом j-ой строки* назовём величину

$$\chi(j) = \min\{|Z_T(w, j)|_0, |Z_T(w, j)|_1\},$$

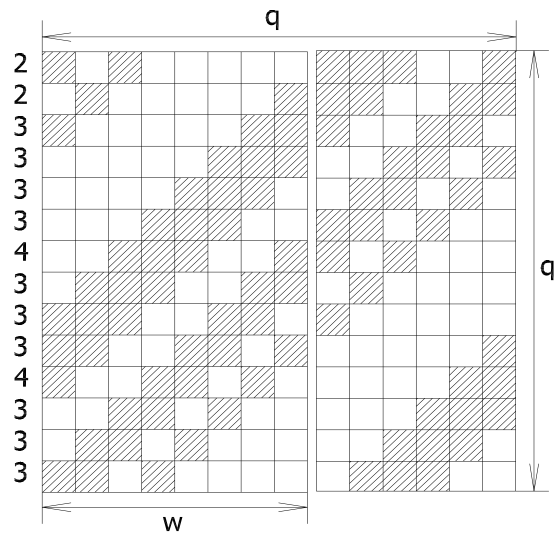


Рис. 1.8: *пример «дополненной» части* **2** *(номера q-классов не проставлены). Элементы T соответствуют заштрихованным клеткам, элементы S — пустым. Слева указаны характеристические числа строк.*

где $|Z_T(w,j)|_0, |Z_T(w,j)|_1$ — это число нулей (соответственно, единиц) в слове $Z_T(w,j)$. Для j-ой строки части **2** $\chi(j)$ равняется минимальному количеству джокеров, необходимому в этой строке для того, чтобы расстановка была допустимой. *Характеристическим числом множества* T (обозначается $\chi(T)$) будем называть величину

$$\chi(T) = \sum_{j=1}^{q} \chi(j).$$

Таким образом, $\chi(T)$ равно минимальному числу джокеров в части **2** в допустимой расстановке при данном множестве T.

Лемма 1.5. *1) Если $t \leq \lfloor w/2 \rfloor$, то $\chi(T) = tw$ для любого множества T.*

2) Если $t > \lfloor w/2 \rfloor$, то $\chi(T) = \lfloor \frac{w^2}{2} \rfloor$ для любого множества T такого, что Z_T есть циклическая перестановка слова $1 \ldots 10 \ldots 0$.

Множество T, такое что Z_T есть циклическая перестановка слова $1 \ldots 10 \ldots 0$, будем обозначать через $\overline{T}$.

Доказательство. Первое утверждение леммы практически очевидно. В самом деле, при $t \leq \lfloor w/2 \rfloor$ в каждом слове $Z_T(w,j)$ единиц не более $\lfloor w/2 \rfloor$, т.е. не больше, чем нулей. Следовательно, $\chi(j) = |Z_T(w,j)|_1$ для всякого j. Каждая единица из Z_T встречается в w словах $Z_T(w,j)$, а значит, всего этих словах содержится tw единиц, откуда $\chi(T) = tw$.

Перейдём к доказательству второго утверждения. «Дополненная» часть **2** в случае множества $\overline{T}$ перестановкой строк приводится к следующему виду (см. рис. 1.9).

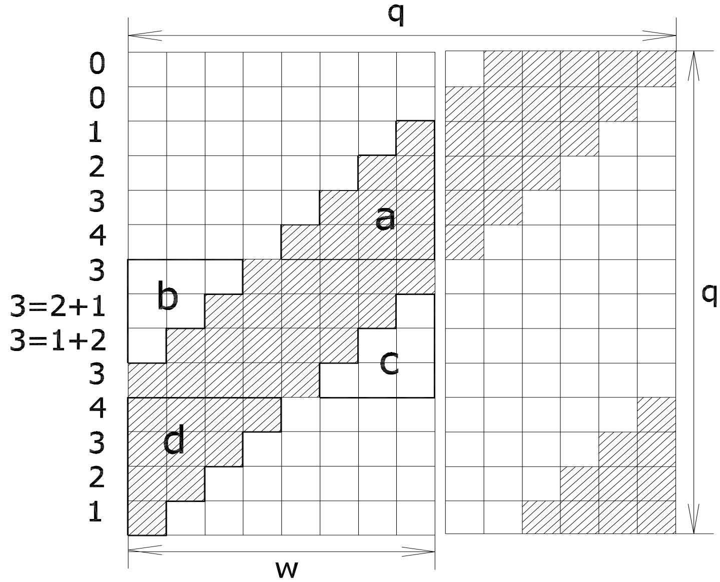


Рис. 1.9: *элементы* $\overline{T}$ *соответствуют заштрихованным клеткам. Слева указаны характеристические числа строк.*

Строки, в которых содержатся элементы как из S, так и из T, то есть такие строки, в которых надо ставить джокеры, пересекаются с выделенными на рис. 1.9 частями **a**, **b**, **c**, **d**. Для строк, пересекающих часть **a** или часть **d**, $\chi(j) = |Z_T(w,j)|_1$, т.е. джокерами нужно заменять элементы множества $\overline{T}$ (это клетки частей **a** и **d**). Для строк, пересекающих часть **b** и/или часть **c**, $\chi(j) = |Z_T(w,j)|_0$, и джокерами нужно заменять элементы множества S (это клетки частей **b** и **c**). Таким образом, $\chi(\overline{T})$ есть суммарное число клеток в частях **a**, **b**, **c** и **d**. Первое слагаемое приводимой формулы соответствует части **a**, второе — **b**, третье — **c**, четвертое — **d**:

$$\chi(\overline{T}) = \sum_{j=1}^{\lfloor \frac{w}{2} \rfloor} j + \sum_{j=1}^{\lfloor \frac{w-1}{2} \rfloor} j + \sum_{j=1}^{\lfloor \frac{w-1}{2} \rfloor} j + \sum_{j=1}^{\lfloor \frac{w}{2} \rfloor} j \ .$$

Вычисляя значение суммы при нечётном w, получаем

$$\chi(\overline{T}) = 4 \cdot \frac{(1 + \frac{w-1}{2})(\frac{w-1}{2})}{2} = \frac{w^2-1}{2} = \left\lfloor \frac{w^2}{2} \right\rfloor ,$$

а при чётном w —

$$\chi(\overline{T}) = 2 \cdot \frac{(1+\frac{w-2}{2})(\frac{w-2}{2})}{2} + 2 \cdot \frac{(1+\frac{w}{2})(\frac{w}{2})}{2} = \frac{w^2}{2} = \left\lfloor \frac{w^2}{2} \right\rfloor .$$

□

Лемма 1.6. *При любом фиксированном $t > \lfloor w/2 \rfloor$ выполняется*

$$\min_{|T|=t} \chi(T) = \chi(\overline{T}),$$

где $\overline{T}$ таково, что $Z_{\overline{T}}$ есть циклическая перестановка слова $1 \dots 10 \dots 0$.

Доказательство. Вначале распространим определение слова $Z_T(w, j)$ на случай произвольного целого j. А именно, определим $Z_T(w, j)$ как префикс длины w циклической перестановки слова Z_T, начинающейся с позиции $(j \bmod q)$ (а если $j \bmod q = 0$, то с q-й позиции). Данное определение можно наглядно представить следующим образом: слово Z_T — «циклическое», то есть за его последней буквой снова следует первая, и т.д. Тогда j-ая позиция как раз и совпадает с позицией $(j \bmod q)$. Величину $\chi(j)$ при таком определении можно рассматривать как функцию $\mathbb{Z} \to \mathbb{N}$, периодическую с периодом q. Рассмотрим также вспомогательную функцию $\phi(j) = |Z_T(w, j)|_1$ (это тоже периодическая с периодом q функция $\mathbb{Z} \to \mathbb{N}$). Заметим, что для произвольного $m \in \mathbb{Z}$

$$\chi(j) = \left\{ \begin{array}{rl} \phi(j), & \phi(j) \le \frac{w}{2} \\ w - \phi(j), & \phi(j) > \frac{w}{2} \end{array} \right. , \tag{1.1}$$

$$\sum_{j=m}^{m+q-1} \chi(j) = \chi(T), \sum_{j=m}^{m+q-1} \phi(j) = tw.$$

Нам будет удобно доопределить функции ϕ и χ на всей числовой прямой. Функцию $\phi(x)$ определим как кусочно-линейную, построенную по точкам $\phi(j)$, а функцию $\chi(x)$ — заменив j на x в равенстве (1.1). Рассмотрим свойства и график функций $\phi(x), \chi(x)$ (пример приведён на рис. 1.10).

1) Поскольку число единиц (как и число нулей) в соседних словах $Z_T(w, j)$ отличается не более, чем на единицу, имеем для любого $j \in \mathbb{Z}$

$$(\phi(j+1) - \phi(j)), (\chi(j+1) - \chi(j)) \in \{1, 0, -1\},$$

откуда следует, что любой линейный участок графика $\phi(x)$ или $\chi(x)$ имеет угловой коэффициент -1, 0 или 1.

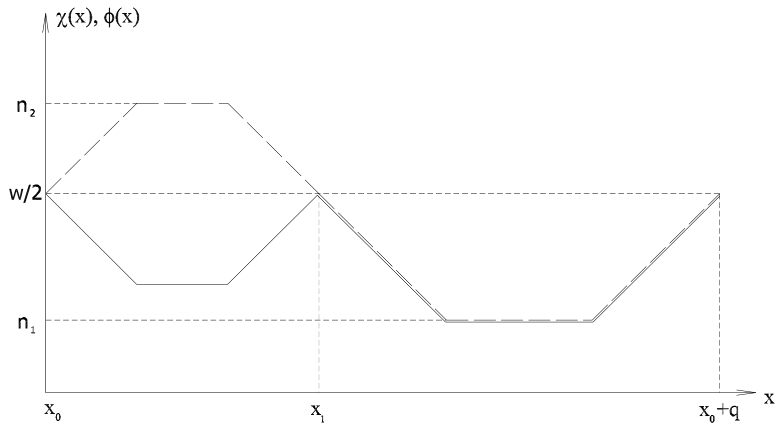


Рис. 1.10: *график $\overline{\chi}(x)$ (сплошная линия) и $\overline{\phi}(x)$ (штриховая линия) на периоде.*

2) Из (1.1) следует, что график $\chi(x)$ получается из графика $\phi(x)$ отражением относительно прямой $y = \frac{w}{2}$ той его части, которая выше этой прямой.

3) Для любых $m \in \mathbb{Z},\ x_0 \in \mathbb{R}$ выполняется

$$\int\limits_{x_0}^{x_0+q} \phi(x)dx = \int\limits_{m}^{m+q} \phi(x)dx = \sum_{j=m}^{m+q-1} \phi(j) = tw\ ,$$

$$\int\limits_{x_0}^{x_0+q} \chi(x)dx = \int\limits_{m}^{m+q} \chi(x)dx = \sum_{j=m}^{m+q-1} \chi(j) = \chi(T).$$

Таким образом, при фиксированном t константа $\int\limits_{x_0}^{x_0+q} \phi(x)dx$ не зависит от выбора множества T; нам необходимо доказать, что константа $\int\limits_{x_0}^{x_0+q} \chi(x)dx$, принимающая, вообще говоря, различные значения при различных T, минимальное значение принимает на множестве $\overline{T}$, указанном в условии леммы.

Возьмем функции $\overline{\phi}(x)$, $\overline{\chi}(x)$, соответствующие множеству $\overline{T}$. Тогда график функции $\overline{\phi}(x)$ на некотором периоде будет иметь вид, приведенный на рис. 1.10 (штриховая линия). Константный промежуток со значением $\overline{\phi}(x) = n_2$ соответствует максимальному количеству единиц в слове $Z_{\overline{T}}(w, j)$ (или максимальному количеству элементов T в строке части **2**), а константный промежуток со значением $\overline{\phi}(x) = n_1$ соответствует минимальному количеству единиц в слове $Z_{\overline{T}}(w, j)$ (или минимальному количеству элементов T в строке части **2**). Согласно свойству 2), график $\overline{\chi}(x)$ будет иметь вид, изображённый на рис. 1.10 сплошной линией.

Оценим константы n_1 и n_2. Пользуясь тем, что $Z_{\overline{T}}$ — циклическая перестановка слова $1\ldots 10\ldots 0$, выберем j_0 такое, что $Z_{\overline{T}}(w, j_0)$ является префиксом этого слова. Тогда $Z_{\overline{T}}(w, j_0)$ содержит максимальное количество (то есть n_2) единиц. Если $w \leq t$, то слово $Z_{\overline{T}}(w, j_0)$ целиком состоит из единиц, откуда $n_2 = w$. Если же $t < w$, то данное слово содержит все t единиц, имеющихся в $Z_{\overline{T}}$, откуда $n_2 = t$. В итоге, $n_2 = \min\{w, t\}$. Рассуждая аналогично, получаем $n_1 = \max\{0, w-s\}$. Заметим, что для любого множества T ($|T| = t$) всякое слово $Z_T(w, j)$ содержит не более, чем t и не более, чем w единиц. Следовательно, $\phi(x) \leq n_2$ для любого рассматриваемого T и для любого $x \in \mathbb{R}$. Аналогично приходим к выводу, что $\phi(x) \geq n_1$.

Кроме того, из условия $t \leq s$ следует, что $n_1 \leq w - n_2$, т.е. левый (на рис. 1.10) локальный минимум функции $\overline{\chi}(x)$ находится выше правого. В самом деле, если $n_2 = w$, то $t \geq w$, откуда $s \geq w$ и $n_1 = 0$. Если же $n_2 = t < w$, то $w - t \geq w - s$ и $w - t > 0$, откуда $w - t \geq n_1$.

Обозначим $\overline{X}_{\geq} = \{x|\, \overline{\phi}(x) \geq \frac{w}{2}\}, \overline{X}_{<} = \{x|\, \overline{\phi}(x) < \frac{w}{2}\}$. Тогда

$$\int\limits_{x_0}^{x_0+q} \overline{\chi}(x)dx = \int\limits_{\overline{X}_{\geq}} \overline{\chi}(x)dx + \int\limits_{\overline{X}_{<}} \overline{\chi}(x)dx = \int\limits_{x_0}^{x_1} \overline{\chi}(x)dx + \int\limits_{x_1}^{x_0+q} \overline{\chi}(x)dx.$$

Теперь возьмем произвольное слово Z_T с условием $|T| = t$, рассмотрим соответствующие ему функции $\phi(x)$ и $\chi(x)$ и докажем, что

$$\int\limits_{x_0}^{x_0+q} \overline{\chi}(x)dx \leq \int\limits_{x_0}^{x_0+q} \chi(x)dx.$$

Обозначим $X_{\geq} = \{x|\, \phi(x) \geq \frac{w}{2}\}, X_{<} = \{x|\, \phi(x) < \frac{w}{2}\}$. Условимся также обозначать через $\rho(M)$ длину (или *меру*) множества M, где M — объединение конечного числа непересекающихся промежутков из отрезка $[x_0, x_0 + q]$.

Как показано выше, $n_1 \leq \phi(x) \leq n_2$. Если $\phi(x) \leq \frac{w}{2}$ всегда, то

$$\int\limits_{x_0}^{x_0+q} \chi(x)dx = \int\limits_{x_0}^{x_0+q} \phi(x)dx = tw > \left\lfloor \frac{w^2}{2} \right\rfloor = \int\limits_{x_0}^{x_0+q} \overline{\chi}(x)dx.$$

Пусть существует точка, где $\phi(x) > \frac{w}{2}$. Если для любого y выполняется $\rho(\{x|\, \chi(x) \geq y\}) \geq \rho(\{x|\, \overline{\chi}(x) \geq y\})$, то тривиально $\int\limits_{x_0}^{x_0+q} \chi(x)dx \geq$

$\int\limits_{x_0}^{x_0+q}\overline{\chi}(x)dx$. Предположим, что $\int\limits_{x_0}^{x_0+q}\chi(x)dx < \int\limits_{x_0}^{x_0+q}\overline{\chi}(x)dx$. Тогда существует y_0 такое, что $\rho(\{x|\,\chi(x) < y_0\}) > \rho(\{x|\,\overline{\chi}(x) < y_0\})$.

Оценим значение y_0. Из непрерывности $\phi(x)$ следует, что она по крайней мере дважды на периоде принимает значение $\frac{w}{2}$; в тех же точках $\chi(x) = \frac{w}{2}$, т.е. $\chi(x)$ на периоде хотя бы дважды достигает максимального значения. Заметим, что функция $\overline{\chi}(x)$ достигает значения $\frac{w}{2}$ ровно дважды на периоде (с точках x_0 и x_1) и спускается до локальных минимумов с наибольшей возможной скоростью — с постоянным угловым коэффициентом, равным -1. Следовательно, для любого $y \geq w-n_2$ (как было доказано, это больший локальный минимум функции $\overline{\chi}(x)$) имеем $\rho(\{x|\,\chi(x) < y\}) \leq \rho(\{x|\,\overline{\chi}(x) < y\})$. Таким образом, $y_0 < w - n_2$. Но в этом случае значения, не превосходящие y_0, функция $\chi(x)$ принимает только в области $X_<$, а функция $\overline{\chi}(x)$ — в области $\overline{X}_<$. В силу свойств функции $\overline{\chi}(x)$ в области $\overline{X}_< = (x_1; x_0 + q)$ (см. рис. 1.10) из неравенства $\rho(\{x|\,\chi(x) < y_0\} \cap X_<) > \rho(\{x|\,\overline{\chi}(x) < y_0\} \cap \overline{X}_<)$ следует неравенство $\rho(\{x|\,\chi(x) < y\} \cap X_<) > \rho(\{x|\,\overline{\chi}(x) < y\} \cap \overline{X}_<)$ для любого $y \leq \frac{w}{2}$. Таким образом, $\rho(X_<) > \rho(\overline{X}_<)$. Но тогда $\rho(\overline{X}_\geq) > \rho(X_\geq)$, и из свойств функции $\phi(x)$ в области $\overline{X}_\geq$ (см. рис. 1.10) имеем

$$\int\limits_{\overline{X}_\geq}(\overline{\phi}(x) - \frac{w}{2})dx \geq \int\limits_{X_\geq}(\phi(x) - \frac{w}{2})dx.$$

Таким образом,

$$tw = \int\limits_{x_0}^{x_0+q}\phi(x)dx = \int\limits_{x_0}^{x_0+q}\chi(x)dx + 2\int\limits_{X_\geq}(\phi(x) - \frac{w}{2})dx <$$

$$< \int\limits_{x_0}^{x_0+q}\overline{\chi}(x)dx + 2\int\limits_{\overline{X}_\geq}(\overline{\phi}(x) - \frac{w}{2})dx = \int\limits_{x_0}^{x_0+q}\overline{\phi}(x)dx = tw.$$

Итак, предположение $\int\limits_{x_0}^{x_0+q}\chi(x)dx < \int\limits_{x_0}^{x_0+q}\overline{\chi}(x)dx$ привело нас к противоречию. Таким образом, лемма 1.6 доказана. $\square$

Из лемм 1.5, 1.6 вытекает, что

$$k_2{}^*(t) = \begin{cases} tw, & t \leq \lfloor\frac{w}{2}\rfloor \\ \lfloor\frac{w^2}{2}\rfloor, & t > \lfloor\frac{w}{2}\rfloor \end{cases},$$

т.е. график зависимости $k_2{}^*(t)$ выглядит следующим образом (рис. 1.11).

Рис. 1.11: *график зависимости* $k_2{}^*(t)$.

Лемма 1.7. *Для любого* $2 \le t \le \lfloor q/2 \rfloor$ *справедливо*
а) $w \ge 4 \Rightarrow k_2{}^*(t) - k_2{}^*(1) \ge w$.
б) $w = 3 \Rightarrow k_2{}^*(t) - k_2{}^*(1) = 1$.
в) $w \le 2 \Rightarrow k_2{}^*(t) - k_2{}^*(1) = 0$.

Доказательство. а) $w \ge 4$. Поскольку $k_2{}^*(t)$ не убывает,

$$k_2{}^*(t) - k_2{}^*(1) \ge k_2{}^*(2) - k_2{}^*(1) = 2w - w = w.$$

б) $w = 3$. Тогда $\lceil \frac{w}{2} \rceil = 2$, и

$$k_2{}^*(t) - k_2{}^*(1) = k_2{}^*(2) - k_2{}^*(1) = \left\lfloor \frac{3^2}{2} \right\rfloor - 3 = 1.$$

в) $w = 1$ или $w = 2$. Тогда $\lceil \frac{w}{2} \rceil = 1$, откуда $k_2{}^*(2) = k_2{}^*(1)$. □

Лемма 1.8. *Для любого* $2 \le t \le \lfloor q/2 \rfloor$ *справедливо* $k_3{}^*(1) - k_3{}^*(t) \le 1$.

Доказательство. Рассмотрим часть **3**. При любом одноэлементном множестве T в каждой строке находится не более одного элемента из T, а значит, в ней можно обойтись не более чем одним джокером. В некоторых строках оптимальной для части **3** расстановки при $t > 1$ может не быть джокеров; мы укажем допустимую расстановку при $t = 1$, в которой джокер стоит ровно в одной из таких строк. Тем самым лемма будет доказана.

Пусть значение $k_3{}^*(t)$ $(t > 1)$ достигается при множестве T'. Если j-ая строка не содержит джокеров при соответствующей расстановке, то $Z_{T'}(w, j) = 1 \ldots 1$ или $Z_{T'}(w, j) = 0 \ldots 0$. Выберем в $Z_{T'}$ какую-нибудь единицу, входящую в подслово 01 или 10. Среди строк, в которые попадает q-класс, соответствующий этой единице, не более чем в одной нет джокеров. Для $t = 1$ возьмём множество T, состоящее из выбранного q-класса. Соответствующая ему расстановка является искомой. □

Пример 1.4. Для бланка $B(40, 11, 8)$ $k_3{}^*(1) = 1$, $k_3{}^*(3) = 0$ (достигается при $T = \{0, 2, 5\}$, см. рис. 1.12a)). Более того, заметим, что при указанном T минимальное количество джокеров в частях **3** и **4** в сумме равно 0, в то время как при $t = 1$ в частях **3** и **4** в совокупности нужно разместить 2 джокера.

1	4	7	2
2	5	0	3
3	6	1	4
4	7	2	5
5	0	3	6
6	1	4	7
7	2	5	0
0	3	6	
1	4	7	
2	5	0	
3	6	1	

a)

1	4	7
2	5	0
3	6	1
4	7	**2**
5	0	**3**
6	1	4
7	2	5
0	3	6
1	4	7
2	5	0
3	6	1

b)

Рис. 1.12: *бланки* $B(40, 11, 8)$ *и* $B(33, 11, 8)$.

Доказательство предложения 1.1. Для любого фиксированного бланка $B(L, p, q)$ и любого $t > 1$ выполняется

$$k^*(t) \geq \lfloor L/pq \rfloor \cdot k_1{}^*(t) + k^*_{2,1}(t) + \ldots + k^*_{2,\lfloor p/q \rfloor}(t) + k_3{}^*(t),$$

где $k^*_{2,1}(t), \ldots, k^*_{2,\lfloor p/q \rfloor}(t)$ — количества джокеров в оптимальных при t расстановках для соответствующих частей **2**. Обозначим сумму в правой части через $\sigma(t)$. Мы построим в бланке специальную расстановку (назовём её *оптимизированной*), количество джокеров в которой не превосходит $\sigma(t)$. Тем самым, с учётом леммы 1.3, предложение будет доказано.

Любую специальную расстановку в бланке можно построить по следующему алгоритму:

1) выбрать некоторый q-класс в качестве единственного элемента T;

2) заменить джокерами все вхождения выбранного элемента в бланк;

3) выбрать в бланке некоторую строку;

4) в выбранной строке удалить все джокеры, поставленные ранее, и поставить джокеры во всех оставшихся позициях.

Относительно расстановки, полученной на шаге 2), заметим, что её пересечение с любой частью **2** бланка является расстановкой для этой части, оптимальной при $t = 1$, а её пересечение с частью **4** бланка содержит не более одного джокера. Пересечение итоговой специальной расстановки с любой частью **1** бланка также является расстановкой для этой части, оптимальной при $t = 1$.

Оптимизированную расстановку построим таким образом: q-класс выберем так, чтобы минимизировать число джокеров в части **3**, а строку — так, чтобы увеличение числа джокеров на шаге 4) в бланке было минимальным. Отметим, что выбор в каждом из случаев может быть не единственным, т.е. для данного бланка может существовать несколько оптимизированных расстановок; зафиксируем одну из них. Количество джокеров в построенной расстановке обозначим через k.

Количество джокеров в каждой части **1** оптимизированной расстановки равно ${k_1}^*(1)$. Количество джокеров в хвосте оптимизированной расстановки есть $k^*_{2,1}(1) + \ldots + k^*_{2,\lfloor p/q \rfloor}(1) + {k_3}^*(1)$ плюс количество джокеров, поставленных на шаге 2) в части **4** (обозначим его через k_4), плюс минимальная разность между количеством элементов из S и элементов из T в одной строке в хвосте бланка (обозначим её через δ).

Заметим, что специальная расстановка, не являющаяся оптимизированной, не может содержать менее k джокеров: в ней может быть на джокер меньше только в части **4**, но обязательно имеется дополнительный джокер в части **3** и/или за счёт δ.

Нам надо доказать, что

$$k - \sigma(t) =$$

$$= \underbrace{\left\lfloor \frac{L}{pq} \right\rfloor ({k_1}^*(1) - {k_1}^*(t)) + (k^*_{2,1}(1) - k^*_{2,1}(t)) + \ldots + (k^*_{2,\lfloor p/q \rfloor}(1) - k^*_{2,\lfloor p/q \rfloor}(t))}_{\Delta_1} +$$

$$+ \underbrace{({k_3}^*(1) - {k_3}^*(t)) + k_4 + \delta}_{\Delta_2} \leq 0.$$

Величина Δ_1 есть сумма неположительных слагаемых согласно леммам 1.4 и 1.7. Поэтому мы вначале оценим величину Δ_2. Если в хвосте бланка нет элементов из T, то ${k_3}^*(1) = k_4 = \delta = 0$, т.е. $\Delta_2 \leq 0$ и требуемое неравенство доказано. Предположим, что в хвосте есть элементы из T, а значит, на шаге 2) построения специальной расстановки в хвосте были поставлены джокеры. По построению $k_4 \leq 1$, а по лемме 1.8 $({k_3}^*(1) - {k_3}^*(t)) \leq 1$. Оценим δ. В хвосте бланка есть «длинные» строки, состоящие из $(\lceil (L \bmod pq)/p \rceil)$ позиций, и «короткие», из $(\lfloor (L \bmod pq)/p \rfloor)$ позиций. Если на шаге 2) был поставлен джокер в «короткой» строке, то $\delta = \lfloor (L \bmod pq/p) \rfloor - 2$, и, таким образом, $\Delta_2 \leq$

$\lfloor (L \bmod pq)/p \rfloor$. Если же на шаге 2) в «коротких» строках не появилось джокеров, то $\delta = \lfloor (L \bmod pq)/p \rfloor - 1$; но в этом случае $k_3{}^*(1) = 0$ (часть **3** состоит только из «коротких» строк). Следовательно, и в этом случае $\Delta_2 \leq \lfloor (L \bmod pq)/p \rfloor$.

Теперь оценим Δ_1, используя условие $L \geq 3p + q$. Если в бланке есть часть **1**, то по лемме 1.4

$$\Delta_1 \leq k_1{}^*(1) - k_1{}^*(t) \leq 2 - p.$$

Поскольку $\Delta_2 \leq q - 1$, получаем, что требуемое условие $\Delta_1 + \Delta_2 \leq 0$ выполнено.

Если же в бланке нет части **1**, то в нём должна присутствовать часть **2** ширины не менее 4. Тогда по лемме 1.7

$$\Delta_1 \leq k^*_{2,1}(1) - k^*_{2,1}(t) \leq -w,$$

где $w = \lceil (L \bmod pq)/p \rceil$. Условие $\Delta_1 + \Delta_2 \leq 0$ снова выполнено. Предложение доказано. □

Замечание 1.6. *Оценка для L, приводимая в условии предложения 1.1, может быть незначительно понижена. Однако, точная оценка сильно зависит от соотношения между конкретными p и q. Отметим здесь, что при $p > 2q$ условие $L \geq 2(p + q)$ является достаточным для существования специальной оптимальной расстановки (такая длина бланка обеспечивает наличие либо части* **1**, *либо части* **2** *ширины ≥ 4, либо двух частей* **2** *ширины 3; в последнем случае применение леммы 1.7 также позволяет доказать предложение). В то же время, в бланке $B(33, 11, 8)$ (см. рис. 1.12б)) любая специальная расстановка содержит не менее 5 джокеров, а оптимальная (при $T = \{0, 2, 5\}$) — 4 джокера, поставленные вместо номеров, выделенных жирным шрифтом; т.е. условие $L \geq 3p$ в общем случае не является достаточным.*

Следующее предложение даёт решение обратной задачи. Через L' обозначим длину хвоста бланка $B(L, p, q)$, равную $L \bmod pq$.

Предложение 1.2. *Для произвольного бланка $B(L, p, q)$ такого, что $q = 3$, $L \geq 2p - 2$ или $q \geq 4$, $L \geq 3p + q$, выполняется*

a) если $L' < q$, то $k^* = (p+q-2) \cdot \left\lfloor \dfrac{L}{pq} \right\rfloor$;

b) если $q \leq L' < p$, то $k^* = (p+q-2) \cdot \left\lfloor \dfrac{L}{pq} \right\rfloor + \left\lfloor \dfrac{L'}{q} \right\rfloor - 1$;

c) если $L' \geq p$, то

$$(p+q-2)\left\lfloor\frac{L}{pq}\right\rfloor+\left\lfloor\frac{L'}{q}\right\rfloor+\left\lfloor\frac{L'}{p}\right\rfloor-2 \leq k^* \leq (p+q-2)\left\lfloor\frac{L}{pq}\right\rfloor+\left\lfloor\frac{L'(p+q)}{pq}\right\rfloor-2.$$

Доказательство. Согласно предложению 1.1 и следствию 1.1, достаточно подсчитать джокеры в оптимальной специальной расстановке; сделаем это в предположении, что такая расстановка построена по алгоритму, указанному в доказательстве предложения 1.1. Каждая часть **1** бланка содержит $p+q-2$ джокеров при любой специальной расстановке. Минимальное количество джокеров во всём бланке, таким образом, достигается при минимуме джокеров в его хвосте.

В случае a) выберем элемент множества T так, чтобы он не встречался в хвосте бланка, а строку, в которой джокерами заменяются элементы из S — чтобы она не пересекала хвост бланка (и то, и другое возможно, поскольку хвост содержит менее q символов). Тогда в хвосте не будет ни одного джокера, и мы получим указанную в случае a) оценку.

Поскольку всякий номер q-класса встречается в бланке ровно 1 раз в каждых последовательных q позициях, в случае b) выберем элемент множества T так, что на шаге 2) построения расстановки в хвосте будет поставлено $\left\lfloor\frac{L'}{q}\right\rfloor$ джокеров. Все они находятся в разных строках (каждая строка пересекается с хвостом бланка не более, чем по одному символу). На шаге 3) выберем строку, содержащую джокер в хвосте, после чего на шаге 4) количество джокеров в хвосте уменьшится на 1. Это даёт нам требуемую оценку.

В случае c) заметим, что оценки снизу и сверху для k^*, в зависимости от параметров бланка, либо равны, либо различаются на 1. Первое слагаемое в обеих оценках есть количество джокеров в теле бланка, одинаковое для всех специальных расстановок. Аналогично случаю b), выберем элемент из T так, что в хвосте бланка на шаге 2) ставится $\lfloor L'/q\rfloor$ джокеров. Далее, на шаге 3) выберем строку, в которой в хвосте имеется джокер (тогда количество джокеров, добавляемых на шаге 4), будет меньше, чем если выбрать строку без джокеров в хвосте). При этом на шаге 4) в хвост бланка будет добавлено $\lfloor L'/p\rfloor - 2$ или $\lfloor L'/p\rfloor - 1$ джокеров в зависимости от того, «короткая» или «длинная» строка выбрана. В итоге, если найдётся номер q-класса, который встречается в хвосте бланка только $\lfloor L'/q\rfloor$ раз и при этом хотя бы один раз — в «короткой» строке, то k^* совпадает со значением левой части; если же всякий такой номер встречается только в «длинных» строках, то k^* больше на единицу. Докажем, что в последнем случае

$$\left\lfloor\frac{L'(p+q)}{pq}\right\rfloor = \left\lfloor\frac{L'}{q}\right\rfloor + \left\lfloor\frac{L'}{p}\right\rfloor + 1;$$

тем самым будет завершено доказательство случая c) и всего предложения.

Пусть $L' \bmod q = r$. Тогда $q - r$ символов встречаются в хвосте «редко», т.е. $\lfloor L'/q \rfloor$ раз. Поскольку эти символы встречаются только в «длинных» строках, то количество «коротких» строк не превосходит r (первые символы нижних r строк в хвосте бланка, очевидно, различны). С другой стороны, $L' \bmod p$ равно количеству «длинных» строк, т.е. $p - r$. (Заметим, что случай $r = 0$ невозможен, так как при этом число L', меньшее pq по определению, оказывается общим кратным взаимно простых p и q.) В итоге, получаем

$$\left\lfloor\frac{L'(p+q)}{pq}\right\rfloor = \left\lfloor\frac{L'}{q} + \frac{L'}{p}\right\rfloor = \left\lfloor\frac{L'}{q}\right\rfloor + \left\lfloor\frac{L'}{p}\right\rfloor + \left\lfloor\frac{L' \bmod q}{q} + \frac{L' \bmod p}{p}\right\rfloor =$$

$$= \left\lfloor\frac{L'}{q}\right\rfloor + \left\lfloor\frac{L'}{p}\right\rfloor + \left\lfloor\frac{r}{q} + \frac{p-r}{p}\right\rfloor = \left\lfloor\frac{L'}{q}\right\rfloor + \left\lfloor\frac{L'}{p}\right\rfloor + 1.$$

□

§5 Решение исходной задачи

В данном параграфе из решения обратной задачи мы получаем решение исходной задачи.

Доказательство теоремы 1.4. Прежде всего, отметим, что функция $k^* = k^*(L, p, q)$, которую мы умеем вычислять (предложение 1.2), и длина взаимодействия $L(k, p, q)$ связаны следующим соотношением:

$$L(k, p, q) = \min\{L | k^*(L, p, q) = k + 1\}. \tag{1.1}$$

Поскольку утверждение предложения 1.2 справедливо при некоторых ограничениях на L, необходимо проверить, что эти ограничения выполняются при имеющихся в теореме 1.4 условиях на k. Для удобства изложения, эта проверка будет проведена в конце доказательства.

Доказательство теоремы 1.4 начнём с доказательства того, что

$$L(k, p, q) = \frac{pq}{p+q-2} \cdot k + \Delta(k, p, q),$$

где функция $\Delta(k,p,q)$ является периодической по k с периодом $p+q-2$. Покажем, что

$$k^*(L+pq,p,q) = k^*(L,p,q) + (p+q-2).$$

В самом деле, бланки $B(L+pq,p,q)$ и $B(L,p,q)$ имеют одинаковые хвосты, а значит, в их хвостах при оптимальной специальной расстановке размещается одно и то же количество джокеров. Тело первого из бланков содержит на один блок больше — в этом блоке будет размещено $(p+q-2)$ джокеров, что и требовалось. Отсюда следует, что

$$\min\{L \mid k^*(L,p,q) = k+1\} = \min\{L \mid k^*(L+pq,p,q) = k+p+q-1\}.$$

Из (1.1) получаем

$$L(k+p+q-2,p,q) = L(k,p,q) + pq,$$

откуда

$$\Delta(k+p+q-2,p,q) = L(k+p+q-2,p,q) - \frac{pq}{p+q-2}(k+p+q-2) =$$

$$= L(k,p,q) + pq - \frac{pq}{p+q-2}k - pq = L(k,p,q) - \frac{pq}{p+q-2}k = \Delta(k,p,q).$$

В дальнейшем мы обозначаем через L' длину хвоста бланка, а через k' — количество джокеров в хвосте, т.е. предполагаем, что для некоторого $\beta \in \mathbb{Z}$ выполняется

$$\begin{array}{rclr} L & = & \beta pq + L', & L' < pq \\ k & = & \beta(p+q-2) + k', & k' < p+q-2. \end{array}$$

Перейдём к доказательству условий для функции $\Delta(k,p,q)$. Вначале докажем часть пункта 3) условия теоремы — оценку снизу для функции $\Delta(k,p,q)$.

$$\Delta(k,p,q) \geq \frac{pq}{p+q-2}.$$

С учётом (1.1), достаточно показать, что в хвосте бланка длиной $L' = \left\lfloor \frac{pq(k'+1)}{p+q-2} - 1 \right\rfloor$ оптимальная расстановка содержит не более k' джокеров. Согласно пункту 3) предложения 1.2, эта расстановка содержит в хвосте не более $\left\lfloor \frac{L'(p+q)}{pq} \right\rfloor - 2$ джокеров. Осталось заметить, что

$$\left\lfloor \frac{L'(p+q)}{pq} \right\rfloor - 2 = k' + 1 + \left\lfloor \frac{2k'+1}{p+q-2} \right\rfloor - 2 \leq k'.$$

Оценка снизу доказана.

Пункт 1). Вначале докажем, что $\max(\Delta(k,p,q)) < 4q-1$. Для этого докажем, что данное неравенство справедливо в трёх частных случаях, после чего покажем, что при любом значении k применим один из этих частных случаев. Параметры w, α и γ являются натуральными числами (параметр w соответствует ширине хвоста в бланке, т.е. $w = \lceil L'/p \rceil$; параметр α соответствует количеству полных сегментов длины q в хвосте, т.е. $\alpha = \lfloor L'/q \rfloor$; параметр $\gamma = wp - \alpha q = p\lceil L'/p \rceil - q\lfloor L'/q \rfloor$).

1. Пусть $L' = \alpha q$, а $w \geq 2$ таково, что $wp = \alpha q + \gamma$, $\gamma < p$. Тогда $k' \geq \alpha + (w-1) - 2 = \alpha + w - 3$ по предложению 1.2. Из (1.1) следует, что для любого β

$$L(\beta(p+q-2)+\alpha+w-4, p, q) \leq \beta pq + \alpha q.$$

Тогда

$$\Delta(\beta(p+q-2)+\alpha+w-4, p, q) \leq \beta pq + \alpha q - \frac{pq}{p+q-2}(\beta(p+q-2)+\alpha+w-4) =$$

$$= \alpha q - \frac{pq}{p+q-2}(\alpha+w-4) = \alpha q - \frac{(q\alpha+\gamma)q}{q\alpha+(\gamma+(q-2)w)}(\alpha+w-4) =$$

$$= \alpha q - q\frac{q\alpha^2+(qw-4q+\gamma)\alpha+(w-4)\gamma}{q\alpha+(\gamma+(q-2)w)} = q\frac{(4q-2w)\alpha-(w-4)\gamma}{q\alpha+(\gamma+(q-2)w)} =$$

$$= 4q - 2w - \frac{(4q-2w)((q-2)w+\gamma)+(w-4)q\gamma}{q\alpha+(\gamma+(q-2)w)}.$$

Знаменатель последней дроби положителен; докажем, что положителен и числитель. Числитель является квадратным трёхчленом относительно q, после раскрытия скобок приводящимся к виду

$$4wq^2 + (\gamma w - 8w - 2w^2)q + (4w^2 - 2\gamma w).$$

Его дискриминант после упрощений оказывается равен

$$D = w^2(2w - (\gamma+8))^2,$$

что позволяет вычислить корни $q_1 = (2w-\gamma)/4$ и $q_2 = 2$. Поскольку $q_1 < q$ по определению w и γ, $q_2 < q$ по условию теоремы и коэффициент при q^2 положителен, то квадратный трёхчлен при допустимых значениях q положителен, а значит, положительна и вся дробь. Таким образом,

$$\Delta(\beta(p+q-2)+\alpha+w-4, p, q) < (4q-2w) \leq 4(q-1). \tag{1.2}$$

2. Пусть $L' = wp$, $w \geq 2$, а α таково, что $wp = \alpha q + \gamma$, $\gamma < q$. Тогда $k' \geq \alpha + w - 2$ по предложению 1.2. Воспользовавшись (1.1), получаем для любого β

$$L(\beta(p+q-2)+\alpha+w-3, p, q) \leq \beta pq + wp,$$

и производим аналогичные выкладки для Δ:

$$\Delta(\beta(p+q-2)+\alpha+w-3, p, q) \leq \beta pq + wp - \frac{pq}{p+q-2}(\beta(p+q-2)+\alpha+w-3) =$$

$$= \alpha q + \gamma - \frac{pq}{p+q-2}(\alpha+w-3) = \alpha q + \gamma - \frac{(q\alpha+\gamma)q}{q\alpha + (\gamma + (q-2)w)}(\alpha+w-3) =$$

$$= \alpha q + \gamma - q\frac{q\alpha^2 + (qw-3q+\gamma)\alpha + (w-3)\gamma}{q\alpha + (\gamma + (q-2)w)} = \gamma + q\frac{(3q-2w)\alpha - (w-3)\gamma}{q\alpha + (\gamma + (q-2)w)} =$$

$$= \gamma + 3q - 2w - \frac{(3q-2w)((q-2)w+\gamma) + (w-3)q\gamma}{q\alpha + (\gamma + (q-2)w)}.$$

Снова знаменатель последней дроби положителен, а числитель является квадратным трёхчленом относительно q; после раскрытия скобок он приводится к виду

$$3wq^2 + (\gamma w - 6w - 2w^2)q + (4w^2 - 2\gamma w).$$

Вычисляя дискриминант, получаем

$$D = w^2(2w - (\gamma + 6))^2,$$

отсюда находим корни $q_1 = (2w - \gamma)/3$ и $q_2 = 2$. Как и в случае 1, при допустимых значениях q квадратный трёхчлен положителен, а значит, положительна и вся дробь. Таким образом,

$$\Delta(\beta(p+q-2)+\alpha+w-3, p, q) < (\gamma + 3q - 2w) \leq 4(q-1). \qquad (1.3)$$

3. Пусть $L' = \alpha q$ и $p = \alpha q + \gamma$ (т.е. $w = 1$). В этом случае в хвост бланка необходимо поставить $(\alpha-1)$ джокеров согласно пункту 2) предложения 1.2. Снова пользуясь (1.1), для любого β имеем

$$L(\beta(p+q-2)+\alpha-2, p, q) \leq \beta pq + \alpha q,$$

и проводим вычисления для Δ:

$$\Delta(\beta(p+q-2)+\alpha-2, p, q) \leq \beta pq + \alpha q - \frac{pq}{p+q-2}(\beta(p+q-2) + \alpha - 2) =$$

$$= \alpha q - \frac{pq}{p+q-2}(\alpha-2) = \alpha q - \frac{(q\alpha+\gamma)q}{q\alpha+(q+\gamma-2)}(\alpha-2) =$$
$$= \alpha q - q\frac{q\alpha^2+(\gamma-2q)\alpha-2\gamma}{q\alpha+(q+\gamma-2)} = q\frac{(3q-2)\alpha+2\gamma}{q\alpha+(q+\gamma-2)} =$$
$$= 3q-2-\frac{(3q-2)(q+\gamma-2)-2q\gamma}{q\alpha+(q+\gamma-2)}.$$

Знаменатель последней дроби положителен, числитель приводится к виду $3q^2+(\gamma-8)q-2\gamma+4$ с корнями $q_1=(2-\gamma)/3$, $q_2=2$. Снова при всех допустимых значениях q дробь положительна, откуда

$$\Delta(\beta(p+q-2)+\alpha-2,p,q) < 3q-2 < 4(q-1). \tag{1.4}$$

Теперь возьмём произвольное $k=\beta(p+q-2)+k'$, $k'<(p+q-2)$. Докажем, что $\Delta(k,p,q)<4(q-1)$. Если $k'+2<p/q$, положим $\alpha=k'+2$, $\gamma=p-\alpha q$ и применим случай 3, получая неравенство (1.4).

Пусть $k'+2>p/q$. Предположим, что найдётся такое α, что $k'+4=\alpha+\left\lceil\frac{\alpha q}{p}\right\rceil$. Тогда, положив $w=\left\lceil\frac{\alpha q}{p}\right\rceil$ и $\gamma=wp-\alpha q$, мы можем применить случай 1 и вывести неравенство (1.2).

Предположим теперь, что указанного выше значения α не существует. Поскольку значения выражения $\alpha+\left\lceil\frac{\alpha q}{p}\right\rceil$ с ростом α на единицу увеличивается не более, чем на 2, то существует такое α, что

$$\alpha+\left\lceil\frac{\alpha q}{p}\right\rceil = k'+3, \tag{1.5}$$

$$(\alpha+1)+\left\lceil\frac{(\alpha+1)q}{p}\right\rceil = k'+5. \tag{1.6}$$

Положим $w=\left\lceil\frac{\alpha q}{p}\right\rceil$ и $\gamma=wp-\alpha q$ и убедимся в применимости случая 2. В самом деле, если $w=\left\lceil\frac{\alpha q}{p}\right\rceil=1$, получаем $k'+2=\alpha<p/q$, противоречие; отсюда $w\geq 2$. Из (1.5) и (1.6) следует, что $\left\lceil\frac{(\alpha+1)q}{p}\right\rceil>\left\lceil\frac{\alpha q}{p}\right\rceil$, откуда $\gamma<q$. Итак, случай 2 применим. Получаем неравенство (1.3).

Теперь докажем, что $\max(\Delta(k,p,q))\geq 2q$. Пусть $k=\beta(p+q-2)$. Из предложения 1.2 получаем $L(k,p,q)=\beta pq+2q$ и $\Delta(k,p,q)=\beta pq+2q-\beta pq=2q$.

Пункт 2). Произведём выкладки, аналогичные приведённым выше. Пусть $p=\alpha q+1$ и $L'=2\alpha q$. Тогда всякий q-класс встречается в хвосте ровно 2α раз, т.е. на шаге 2) построения специальной расстановки в хвосте будет поставлено 2α джокеров. В качестве T можно выбрать класс, встречающийся в строке длины 1, после чего на шаге 3) выбрать эту строку. Тогда на шаге 4) количество джокеров уменьшится на 1. В итоге, $k' = 2\alpha - 1$. Если же $L' = 2\alpha q - 1$, то выберем $T = \{0\}$ — этот класс встречается в хвосте $2\alpha - 1$ раз, в том числе в одноэлементной строке (а именно, в предпоследней, в силу выбора p). Таким образом, мы можем ограничиться $2\alpha - 2$ джокерами. Из (1.1) следует, что для любого β

$$L(\beta(p+q-2) + (2\alpha-2), p, q) = \beta pq + 2\alpha q.$$

Тогда

$$\Delta(\beta(p+q-2)+2\alpha-2, p, q) = \beta pq + 2\alpha q - \frac{pq}{p+q-2}(\beta(p+q-2)+2\alpha-2) =$$

$$= 2\alpha q - \frac{pq}{p+q-2}(2\alpha-2) = 2\alpha q - \frac{(q\alpha+1)q}{q\alpha+1+q-2}(2\alpha-2) =$$

$$= 2\alpha q - 2q\frac{q\alpha^2-(q-1)\alpha-1}{q\alpha+(q-1)} = 2q\frac{2(q-1)\alpha+1}{q\alpha+(q-1)} = 4(q-1) - \frac{4(q-1)^2-2q}{q\alpha+(q-1)}.$$

Последняя дробь положительна, а при увеличении α она стремится к нулю. Таким образом, при любом заданном $\varepsilon > 0$ и заданном q можно указанным способом выбрать p и k так, чтобы

$$4(q-1) - \varepsilon < \Delta(k, p, q) < 4(q-1).$$

Пункт 3). Из предложения 1.2 следует, что $k^*(\beta pq - 1, p, q) = \beta(p+q-2) - 1$ и $k^*(\beta pq, p, q) = \beta(p+q-2)$. Из (1.1) получаем, что если $k = \beta(p+q-2) - 1$, то $L(k, p, q) = \beta pq$. Тогда $\Delta(k, p, q) = pq/(p+q-2)$ и, с учётом нижней оценки для $\Delta(k, p, q)$, пункт 3) доказан.

Мы доказали все пункты теоремы. Осталось доказать правомерность применения предложения 1.2, т.е. выполнение приведённых в нём ограничений на L при заданных в теореме ограничениях на k. Пусть вначале $q = 3$.

$$k^*(2p-2, p, 3) \le \left\lfloor \frac{(2p-2)(p+3)}{3p} \right\rfloor - 2 = \left\lfloor \frac{2p+1}{3} - \frac{2}{p} \right\rfloor - 1 = \left\lfloor \frac{2p}{3} \right\rfloor - 1.$$

Это означает, что

$$L(\left\lfloor \frac{2p}{3} \right\rfloor - 1, p, q) \ge 2p - 2,$$

что и требовалось. Если же $q \ge 4$, то

$$k^*(3p+q, p, q) \le \left\lfloor \frac{(3p+q)(p+q)}{pq} \right\rfloor - 2 \le \left\lfloor \frac{3p}{q} \right\rfloor + 3.$$

Отсюда получаем требуемое условие

$$L(\left\lfloor \frac{3p}{q} \right\rfloor + 3, p, q) \geq 3p + q.$$

Доказательство теоремы завершено. $\square$

Предложение 1.3. *Для фиксированных периодов p, q можно определить точный вид функции $\Delta(k, p, q)$ за время, зависящее только от p и q.*

Доказательство. Из теоремы 1.4 следует, что функция $\Delta(k, p, q)$ является периодической по k с периодом $p + q - 2$. Это означает, что достаточно определить $\Delta(k, p, q)$ для всех целых $k \in [k^*, k^* + p + q - 2)$, где $k^* = \lfloor 2p/3 \rfloor - 1$ при q=3 и $k^* = \lfloor 3p/q \rfloor + 3$ при $q \geq 4$, а для этого достаточно определить для данных значений k значения $L(k, p, q)$. Для данного k значение $L(k, p, q)$ на единицу больше, чем длина оптимальной расстановки. Из доказательства теоремы 1.4 следует, что при указанных значениях k существует оптимальная специальная расстановка. Таким образом, для определения для данного k значения $L(k, p, q)$ достаточно определить максимум из длин всех допустимых специальных расстановок. Специальных расстановок для данного k всего q штук (в зависимости от выбора единственного элемента множества T). Таким образом, время, необходимое для определения точного вида функции $\Delta(k, p, q)$, зависит только от p и q.

$\square$

Глава II

Статистические закономерности взаимодействия периодов

Зафиксируем взаимно простые периоды p, q. В данной главе рассматриваются расстановки, длина которых меньше длины взаимодействия. В этом случае выполнение свойства взаимодействия периодов зависит от расположения джокеров, и расстановки одинаковой длины с одинаковым количеством джокеров могут иметь различную размерность. Напомним используемые обозначения. Мы обозначаем через $M(r, L, k)$ множество всех расстановок длины L, содержащих k джокеров и имеющих размерность r. Если какой-то параметр заменен знаком $\forall$, то это означает, что рассматривается объединение таких множеств по всем возможным значениям параметра. Обозначим через $P(r, L, k)$ долю расстановок длины L, содержащих k джокеров и имеющих размерность r, т.е. отношение мощности множества $M(r, L, k)$ к мощности множества $M(\forall, L, k)$. Основным результатом данной главы является эффективный алгоритм для определения $P(r, L, k)$. Время работы «наивного» алгоритма экспоненциально зависит от длины слова; время работы нашего алгоритма полиномиально зависит от L и p и экспоненциально — от параметра q.

§6 Идея алгоритма

В данном параграфе излагается основная идея алгоритма. Алгоритм основан на разбиении расстановок из $M(r, L, k)$ на классы и вычислении их количества в каждом классе.

Для того, чтобы разбить расстановки на классы, сопоставим каждой расстановке G расстановку длины pq следующим образом. Пусть $l = \lceil |G|/pq \rceil$. Дополним расстановку G нулями до длины lpq. Тогда ее можно представить как конкатенацию расстановок $G_i, i = 1, \dots l$, где $|G_i| = pq$. Применим к словам $G_i, i = 1, \dots l$ последовательно логическое ИЛИ. Полученную в результате расстановку длины pq назовем *характеристической* расстановкой.

Пример 2.1. Рассмотрим частичное слово с периодами 4,5 и соответствующую ему расстановку.

$$
\begin{array}{c}
aa\diamond\diamond\diamond\diamond\diamond ba\diamond\diamond\diamond\diamond\diamond\diamond\diamond\diamond\diamond a\diamond\diamond\diamond\diamond\diamond\diamond\diamond\diamond\diamond baa\diamond\diamond \\
110000011000000000100000000011100
\end{array}
$$

Характеристическая расстановка получается из исходной расстановки следующим образом:
1. Дополним исходную расстановку нулями до длины, кратной pq, и представим получившееся слово как произведение слов G_1 и G_2, $|G_1| = |G_2| = pq$:

$$
\underbrace{11000001100000000010}_{G_1}\underbrace{0000000011100\ \ 0000000}_{G_2}.
$$

2. К словам G_1 и G_2 применим последовательно логическое ИЛИ, получая характеристическую расстановку:

$$
\begin{array}{rl}
\vee\, G_1 = & 11000001100000000010 \\
G_2 = & \underline{00000001110000000000} \\
 & 11000001110000000010 \text{ — характеристическая расстановка}
\end{array}
$$

Предложение 2.1. *Размерность расстановки совпадает с размерностью ее характеристической расстановки.*

Доказательство. Рассмотрим p- и q-классы расстановки и характеристической расстановки. Напомним, что если две позиции из разных q-классов, занятые буквами, принадлежат одному p-классу (т.е. «связаны» при помощи периода p), то в этих q-классах содержится одна и та же буква. Если же любая пара позиций из данных q-классов, принадлежащая одному p-классу, содержит хотя бы один джокер, то эти q-классы могут содержать различные буквы.

Таким образом, размерность расстановки зависит от того, какие q-классы связаны между собой при помощи периода p. Заметим, что в характеристической расстановке единица присутствует в позиции, принадлежащей данным q- и p-классу тогда и только тогда, когда в исходной расстановке единица присутствует хотя бы в одной из позиций, принадлежащих тем же q- и p-классу. Поэтому в характеристической расстановке q- и p-классы связаны между собой так же, как в исходной, а значит размерности исходной и характеристической расстановок совпадают. $\square$

Пример 2.2. Продолжим рассмотрение примера 2.1. В соответствующих характеристической расстановке 1100000111000000010 словах множество позиций, в которых стоят буквы, состоит из 6 элементов: $D = \{1, 2, 8, 9, 10, 19\}$. Будем нумеровать 4-классы (5-классы) остатком от деления на 4 (соответственно, на 5) номеров позиций, принадлежащих этому классу. Первая и девятая позиции принадлежат первому 4-классу, девятая и девятнадцатая — первому 5-классу, а значит, в этих позициях должна стоять одна и та же буква. Вторая и десятая позиции принадлежат второму 4-классу. Больше связей с помощью периодов между позициями из D нет. Значит, соответствующие характеристической расстановке частичные слова могут содержать не более 3 различных букв (например, $ab\diamond\diamond\diamond\diamond\diamond cab\diamond\diamond\diamond\diamond\diamond\diamond\diamond\diamond a\diamond$). Поэтому размерность характеристической (а значит, и исходной) расстановки равна 3.

Поставим в соответствие каждой расстановке ее характеристическую расстановку и разобьем расстановки на соответствующие классы. Тогда для определения $P(r, L, k)$ необходимо найти все характеристические расстановки размерности r, а затем для каждой такой характеристической расстановки определить количество соответствующих ей расстановок из $M(r, L, k)$. Характеристические расстановки можно находить, например, перебором (вычислительная сложность такого перебора зависит только от p и q, и не зависит от длины L, которая может быть сколь угодно большой по сравнению с периодами). Тем не менее, на практике и такой перебор достаточно затруднителен. Поэтому далее будет рассмотрен эффективный способ подсчета характеристических расстановок заданной размерности.

§7 Количество расстановок с заданной характеристической расстановкой

Во данном параграфе рассматривается задача вычисления мощности класса расстановок, которым соответствует данная характеристическая расстановка. Из решения этой задачи будет следовать, что можно не перечислять все характеристические расстановки, а разбить их на классы так, что характеристическим расстановкам из одного класса соответствует одно и то же число расстановок, а затем подсчитать число характеристических расстановок в каждом классе. Возможны два варианта интересующей нас задачи:

1. Найти количество расстановок из множества $M(\forall, L, \forall)$ с данной характеристической расстановкой.

2. Найти количество расстановок из множества $M(\forall, L, k)$ с данной характеристической расстановкой.

Начнем с решения первого варианта задачи. Известно, что в позициях исходной расстановки, соответствующих позициям нулей в характеристической расстановке, должны находиться нули, в остальных позициях могут стоять и нули, и единицы, но каждой единице в характеристической расстановке должна соответствовать хотя бы одна единица в исходной расстановке.

Рассмотрим случай $L = lpq$. Тогда каждой единице в характеристической расстановке H соответствуют l позиций в исходной расстановке. В каждой из этих позиций может находиться 0 или 1, что дает 2^l комбинаций, однако во всех позициях нули находиться не могут, поэтому каждой единице характеристической расстановки соответствуют $2^l - 1$ комбинаций. Пусть $h_1 = |H|_1$ и $h_0 = |H|_0$. Таким образом, количество расстановок длины $L = lpq$, которым соответствует данная характеристическая расстановка H, равно

$$(2^l - 1)^{h_1}.$$

В случае $L = lpq + b, 0 < b < pq$ обозначим через H' префикс слова H длины b. Тогда количество расстановок длины $L = lpq + b$, которым соответствует данная характеристическая расстановка H, равно

$$(2^{l+1} - 1)^{|H'|_1} + (2^l - 1)^{h_1 - |H'|_1}.$$

Таким образом, решение первой задачи можно получить с помощью указанных выше формул за время $O(pq)$, требуемое для вычисления h_0 и h_1.

Во втором варианте задачи рассматриваются не все расстановки, которым соответствует данная характеристическая расстановка H, а только те, которые содержат ровно k нулей.

Пусть $l = \lfloor L/pq \rfloor, k' = k - lh_0$. Обозначим через $d(k', h_1)$ количество расстановок из $M(r, L, k)$, соответствующих данной характеристической расстановке H. При фиксированном l эта величина зависит только от k' и количества единиц в H (и не зависит от периодов p, q).

Обозначим через $K_{i,j}$ множество всех представлений числа i в виде суммы j неотрицательных слагаемых, каждое из которых не превосходит $l - 1$, а через z_{nm} — m-е слагаемое из n-го представления.

Предложение 2.2. *Если $L = lpq$, то*

$$d(k', h_1) = \sum_{n=1}^{|K_{k',h_1}|} \prod_{m=1}^{h_1} C_l^{z_{nm}}.$$

Доказательство. В позициях, соответствующих нулям в характеристической расстановке H, в исходной расстановке также должны стоять нули. Таких позиций в исходной расстановке lh_0. Оставшиеся $k' = k - lh_0$ нулей нужно расставить в позициях, соответствующих единицам в H. Для этого рассмотрим все (неупорядоченные) представления числа k' в виде суммы h_1 неотрицательных целых слагаемых, такие что каждое слагаемое не превосходит $l-1$. Тогда i-е слагаемое в представлении соответствует i-й по счету единице в характеристической расстановке и задает количество нулей, которые необходимо поставить в исходной расстановке в позициях, соответствующих позиции этой единицы. Таких позиций l штук, поэтому количество возможных расстановок нулей, соответствующих слагаемому, равному z, составляет C_l^z. Отсюда получаем искомую формулу.

$$d(k', h_1) = \sum_{n=1}^{|K_{k',h_1}|} \prod_{m=1}^{h_1} C_l^{z_{nm}}.$$

□

Предложение 2.3. *Если $L = lpq + g, 0 < g < pq$, то*

$$d(k', h_1) = \sum_{n} \prod_{m=1}^{s} C_{l+1}^{z_{nm}} \prod_{m=s+1}^{h_1} C_l^{z_{nm}},$$

где s — количество единиц, находящихся в первых g позициях характеристической расстановки, а суммирование производится по всем представлениям числа k' в виде суммы h_1 слагаемых, первые s из которых не превосходят l, а остальные не превосходят $l-1$.

Доказательство. В случае $L = lpq + g, 0 < g < pq$ необходимо учитывать, что единицам, принадлежащим первым g позициям характеристической расстановки, соответствуют $l+1$ позиций исходной расстановки. Обозначим через s количество единиц, находящихся в первых g позициях характеристической расстановки. Тогда для заданной характеристической расстановки H количество соответствующих ей расстановок $M(\forall, L, k)$ равно

$$d(k', h_1) = \sum_{n} \prod_{m=1}^{s} C_{l+1}^{z_{nm}} \prod_{m=s+1}^{h_1} C_l^{z_{nm}}$$

где суммирование производится по всем представлениям числа k' в виде суммы h_1 слагаемых, первые s из которых не превосходят l, а остальные не превосходят $l-1$. □

В следующем параграфе будет представлен алгоритм вычисления $d(i, j)$, время работы которого полиномиально зависит от p и экспоненциально — от q.

Итак, в случае $L = lpq$ для определения количества расстановок из $M(r, L, k)$ достаточно знать для всех $x = 1, \dots, pq$ количество расстановок (r, pq, x). В случае $L = lpq + b, 0 < b < pq$ достаточно знать для всех $x = 1, \dots, pq$ и для всех $y = 1, \dots, x$ количество расстановок (r, pq, x) таких, что в первых b позициях находятся y единиц.

§8 Построение вспомогательных таблиц

Данный параграф посвящен описанию алгоритмов построения вспомогательных таблиц B и D с использованием динамического программирования.

При вычислении $P(r, L, k)$ для различных r и k можно построить один раз таблицы $B_{q \times pq}$, где $b(i, j)$ равно количеству характеристических расстановок с размерностью i и количеством единиц j, и $D_{k' \times pq}$, элементы $d(i, j)$ которой определены выше. Тогда

$$P(r, L, k) = \frac{\sum_{i=0}^{pq} b(r, i) d(k - l(pq - i), i)}{C_L^k}, \qquad (2.1)$$

т.е. при наличии этих таблиц временна́я сложность вычисления $P(r, L, k)$ линейно зависит от pq. В этом параграфе рассмотрены алгоритмы построения таблиц B и D и временна́я сложность этих алгоритмов.

В дальнейшем для упрощения изложения будем рассматривать случай $L = lpq$. Запишем характеристическую расстановку H в виде бланка. Напомним, что строки бланка соответствуют p-классам. Переставим элементы в строках так, чтобы элементы одного q-класса находились в одной колонке. Вместо номера q-класса запишем в ячейку, соответствующую позиции n, элемент $H(n)$. Таким образом, каждый элемент $H(n)$ записывается в ячейке (i, j), где $i = n \bmod p, j = n \bmod q$. Все буквы в позициях, принадлежащих одной строке или одному столбцу бланка, должны быть одинаковыми.

Приведем полученную таблицу перестановками строк и столбцов к клеточно-диагональному виду с максимальным возможным количеством клеток (мы будем называть эти клетки *блоками*); ограничение на перестановки таково: порядок следования строк (столбцов) внутри блоков должен быть таким же, как в исходном бланке. Далее переставим блоки по убыванию высоты, а одинаковые по высоте по убыванию ширины. Полученную таблицу будем называть *фактор-бланком*.

Пример 2.3. Вернемся к рассмотрению примера 2.1. Представим характеристическую расстановку 11000001110000000010 в виде таблицы. Первый символ расстановки будет записан в клетке (1,1), второй — (2,2), третий — (3,3), четвертый — (4,0), пятый — (0,1), и т.д. Результат приведен на рис.2.1a). При перестановке строк и столбцов так, как указано на рис.2.1b), получается фактор-бланк.

	0	1	2	3
0	0	0	1	0
1	0	1	0	0
2	0	0	1	0
3	1	0	0	0
4	0	1	0	1

a)

	1	3	2	0
1	1	0	0	0
4	1	1	0	0
0	0	0	1	0
2	0	0	1	0
3	0	0	0	1

b)

Рис. 2.1: Представление расстановки 11000001110000000010 в виде бланка (a)) и соответствующий фактор-бланк (b)).

Предложение 2.4. *Количество блоков в фактор-бланке совпадает с размерностью исходной расстановки.*

Доказательство. Две единицы принадлежат одному блоку тогда и только тогда, когда их можно соединить ломаной такой, что она начинается в клетке первой единицы, заканчивается в клетке второй единицы, и в каждой клетке излома стоит единица. В противном случае перестановкой строк и столбцов этот блок можно было бы привести к клеточно-диагональному виду с как минимум двумя клетками, что противоречит максимальности количества блоков в фактор-бланке.

Если две единицы фактор-бланка можно соединить ломаной, построенной указанным выше способом, то это означает, что в соответствующем данной расстановке частичном слове находящиеся в позициях этих единиц буквы равны, так как эта ломаная соединяет между собой позиции, принадлежащие одной и той же строке или столбцу, т.е. равные по модулю p или q. В противном случае в соответствующих позициях могут находиться разные буквы. Это означает, что в соответствующем данной расстановке частичном слове в позициях из одного блока должны находиться одинаковые буквы, в позициях из разных блоков могут находиться разные буквы. Поэтому количество блоков в фактор-бланке совпадает с размерностью исходной расстановки. □

Приступим к построению таблицы B. Все бланки размера $i \times j$ с n единицами (их C_{ij}^n) можно разделить на (1) бланки, фактор-бланк которых содержит более одного блока (обозначим количество таких бланков через $s_m(i,j,n)$, где m — количество блоков в фактор-бланке), (2) блоки размера $i' \times j'$ с n единицами, где $i' \leq i$, $j' \leq j$, $i'j' < ij$, дополненные пустыми строками или столбцами до размера $i \times j$ (обозначим количество таких бланков через $t_{i',j'}(i,j,n)$), (3) блоки размера $i \times j$ с n единицами (обозначим количество таких бланков через $e(i,j,n)$). Каждый элемент таблицы B вычисляется независимо от остальных по данным вспомогательных таблиц E, $S_m(m = 2,\ldots,q)$ и $T_{i',j'}(i' = 1,\ldots,q, j' = 1,\ldots,p)$. Размер каждой из этих таблиц $p \times q \times pq$.

Для построения таблицы E используем метод динамического программирования. Заметим, что $e(i,j,n) = e(j,i,n)$, поэтому достаточно построить только ту часть таблицы E, где $i \leq j$. Начнем с минимального блока размера 1×1, содержащего одну единицу. При построении таблицы E переход от элемента к элементу осуществляется увеличением на единицу первого справа индекса, для которого это возможно.

Бланки вида (1) получаются из фактор-бланка, содержащего несколько блоков, с помощью перестановки строк и столбцов с сохранением порядка строк и столбцов внутри блоков, т.е. для каждой строки и столбца бланка достаточно указать номер соответствующего блока. При такой перестановке строк и столбцов одинаковые бланки при разных перестановках могут получиться только в случае наличия одинаковых блоков. Для оптимизации будем использовать дополнительные таблицы F_m, где элемент $f_m(i,j,n)$ равен количеству бланков длины ij с n единицами вида (1) без пустых строк и столбцов, фактор-бланк которых содержит m блоков. Пусть фактор-бланк состоит из блоков $B_1,\ldots,B_m, m > 1$, каждый блок встречается соответственно $n_1,\ldots n_m$ раз, размеры этих блоков соответственно равны $x_1 \times y_1,\ldots,x_m \times y_m$, $\sum_{g=1}^m x_g n_g = i, \sum_{g=1}^m y_g n_g = j$. Тогда количество соответствующих этому фактор-бланку бланков равно $C_i^{x_1 n_1,\ldots,x_m n_m} C_j^{y_1 n_1,\ldots,y_m n_m}$. Таким образом, для определения $f_m(i,j,n), m > 1$ необходимо рассмотреть все возможные комбинации m блоков, сумма размеров которых равна $i \times j$, а сумма единиц равна n. Для каждой из этих комбинаций блоков можно определить количество соответствующих ему бланков без пустых строк и столбцов по указанной выше формуле. Все бланки длины ij с n единицами вида (1) получаются из бланков длины $i'j'$, $i'j' \leq ij$ с n единицами с помощью добавления пустых строк и/или столбцов, поэтому

$$s_m(i,j,n) = \sum_{i'=1}^{i} \sum_{j'=1}^{j} f_m(i',j',n) C_i^{i'} C_j^{j'}, m = 2,\ldots,q.$$

При построении таблицы F_m для $m = 2, \ldots, q$ происходит перебор всех возможных комбинаций параметров m блоков (т.е. высоты, ширины и количества единиц). Каждая такая комбинация рассматривается не более одного раза. Количество блоков, имеющих заданные параметры, определяется по уже построенной части таблицы E (элемент $e(x, y, n)$ равен количеству блоков размера $x \times y$ с n единицами). Поэтому количество необходимых операций при этом ограничено произведением количеств разбиений (т.е. представлений в виде суммы неупорядоченных натуральных слагаемых) чисел p, q и pq, причем эти разбиения могут содержать не более q частей. Таким образом, таблицы $F_m, m = 2, \ldots, q$ можно построить за время $O((pq)^{2q})$. Величина $s_m(i, j, n)$ вычисляется по данным таблицы F_m за полиномиальное от p, q время.

Бланки вида (2) получаются добавлением пустых строк и столбцов к блокам с n единицами, размер которых меньше $i \times j$. Количество этих блоков известно, т.к. в таблице E заполнены все элементы вплоть до элемента $e(i, j, n)$. Из блока размера $x \times y$ с n единицами получается $C_i^x C_j^y$ бланков размера $i \times j$ вида (2). Таким образом,

$$t_{x,y}(i, j, n) = e(x, y, n) C_i^x C_j^y.$$

Вычисление количества бланков вида (2) также производится за полиномиальное от p, q время.

Итак, для определения очередного элемента таблицы E (т.е. количества бланков вида (3)) необходимо найти разность C_{ij}^n и количества бланков вида (1) и (2):

$$e(i, j, n) = C_{ij}^n - \sum_{m=2}^{q} s_m(i, j, n) - \sum_{i'=1}^{i-1} \sum_{j'=1}^{j-1} t_{i',j'}(i, j, n) -$$

$$- \sum_{i'=1}^{i-1} t_{i',j}(i, j, n) - \sum_{j'=1}^{j-1} t_{i,j'}(i, j, n).$$

Элементы таблицы B получаются при вычислении последних pq элементов таблицы E (для всех $n = 1, \ldots, pq$ элементы $b(m, n), m = 1, \ldots, q$ получаются при вычислении $e(p, q, n)$) следующим образом:

$$b(1, n) = e(p, q, n) + \sum_{i'=1}^{p} \sum_{j'=1}^{q} t_{i',j'}(p, q, n) - t_{p,q}(p, q, n),$$

$$b(m, n) = s_m(p, q, n), m = 2, \ldots, q.$$

Таким образом, элементы $b(1,n)$ равны сумме количества бланков вида (1) и (2). Элементы $b(m,n), m = 2,\dots,q$ равны количеству бланков вида (3), фактор-бланк которых состоит ровно из m блоков. Временна́я сложность выполнения указанных выше операций для построения таблицы E полиномиально зависит от p и экспоненциально — от q. Таким образом, выполняется следующее предложение.

Предложение 2.5. *Временна́я сложность построения таблицы B полиномиально зависит от p и экспоненциально — от q.*

Для построения таблицы D также используем метод динамического программирования. Напомним, что элемент $d(i,j)$ равен количеству расстановок $(\forall, L, l(pq-i)+j)$, характеристической расстановкой которых является данная расстановка H, содержащая i единиц, причем эта величина не зависит от выбора такой H. Заметим, что $d(0,j) = 1, j = 1,\dots,pq$, поскольку имеется ровно одно представление 0 в виде любого количества неотрицательных слагаемых, $C_l^0 = 1, l = 1,\dots,pq$. Положим $d(0,0) = 1$ (этот элемент соответствует «пустому месту»), $d(i,0) = 0, i > 0$. Обозначим через $K_{i,j}$ множество всех представлений числа i в виде суммы j неотрицательных слагаемых, каждое из которых не превосходит $l-1$, а через z_{nm} — m-е слагаемое из n-го представления. По определению

$$d(i,j) = \sum_{n=1}^{|K_{ij}|} \prod_{m=1}^{j} C_l^{z_{nm}},$$

где суммирование производится по всем представлениям из $K_{i,j}$. Каждому представлению из $K_{i,j}$, в котором последнее слагаемое равно x, можно поставить во взаимно однозначное соответствие представление из $K_{i-x,j-1}$ (первое представление получается из второго с помощью «дописывания» справа слагаемого x). Для первого представления величина $\prod_{m=1}^{j} C_l^{z_{nm}}$ равна произведению той же величины для второго представления на C_l^x. Разобьем представления (i,j) на l классов в соответствии с величиной последнего слагаемого представления (от 0 до $l-1$). Тогда

$$d(i,j) = \sum_{m=\min\{0,i-l+1\}}^{i} d(m,j-1) C_l^{(i-m)}.$$

Предложение 2.6. *Временна́я сложность построения таблицы D равна $O(L^2)$.*

Доказательство. Для определения каждого элемента таблицы D необходимо просмотреть в худшем случае l ее элементов. Высота таблицы равна k, ширина равна pq. Таким образом, временная сложность построения таблицы D равна $O(kpql) = O(L^2)$. $\square$

Итак, построив таблицы B и D, мы можем вычислить $P(r, L, k)$ за время $O(pq)$ по формуле 2.1.

Алгоритм 2.1. (Вычисляет $P(r, L, k)$).
Шаг 1. Вычислить вспомогательную таблицу $B_{q\times pq}$, элементами которой являются $b(i, j)$.
Шаг 2. Вычислить вспомогательную таблицу $D_{k'\times pq}$, элементами которой являются $d(i, j)$.
Шаг 3. Вычислить $P(r, L, k)$ по формуле 2.1.

Таким образом, общее время работы нашего алгоритма полиномиально зависит от L и экспоненциально — от параметра q. Заметим, что с одними и теми же таблицами B,D можно вычислить $P(r, L, k)$ для всех $k' \leq k$. Более того, при вычислении $P(r, L', k')$ для $L' \neq L$ требуется перестраивать только таблицу D.

§9 Анализ полученных результатов

Данный параграф посвящен анализу эмпирических данных, полученных с помощью программы, реализующуй алгоритм 2.1. Эта программа использовалась для получения данных о поведении $P(r, L, k)$ при больших (по сравнению с периодами) длинах (эти данные невозможно было получить с помощью «наивного» алгоритма, т.к. перебор при этом был слишком велик). В частности, исследовалась зависимость доли неунарных расстановок (т.е. расстановок, размерность которых больше единицы) длины L с k джокерами от плотности джокеров в слове (т.е. от величины k/L). Все графики данной зависимости очень похожи (см. примеры на рис.2.2, 2.3). Вертикальная линия на обоих графиках соответствует $\frac{k}{L} = \frac{p+q-2}{pq}$ — оценке длины взаимодействия из теоремы 1.4.

При исследовании полученных графиков были выявлены следующие закономерности.

Закономерность 1. Доля неунарных расстановок близка к нулю на большом отрезке графика «справа» от длины взаимодействия.

Закономерность 2. Плотность джокеров, при которой начинается заметный рост доли неунарных расстановок, при увеличении длины также увеличивается. Таким образом, если зафиксировать $0 < \varepsilon < 1$, то при увеличении длины количество джокеров, при котором доля неунарных

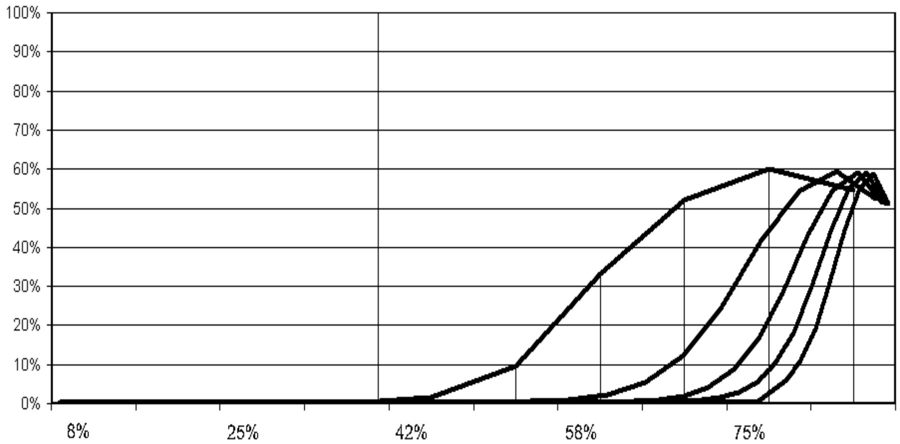


Рис. 2.2: *Зависимость доли неунарных расстановок от плотности джокеров в случае $q = 3, p = 4, L = 12, 24, 36, 48, 60$ (слева направо).*

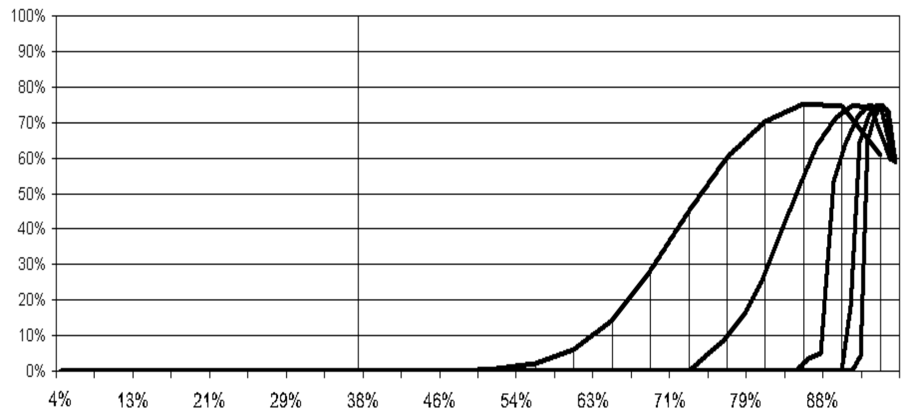


Рис. 2.3: *Зависимость доли неунарных расстановок от плотности джокеров в случае $q = 3, p = 8, L = 24, 48, 72, 96, 120$ (слева направо).*

расстановок начинает превышать ε, увеличивается не пропорционально увеличению длины, а на бо́льшую величину.

Закономерность 3. Плотность джокеров, при которой достигается максимум доли неунарных расстановок, при увеличении длины также увеличивается. Более того, в рассмотренных случаях максимум доли неунарных расстановок достигался при $L - k = const \leq q$ (т.е. одном и том же количестве единиц в расстановке).

Глава III

Свойство взаимодействия локальных периодов

Данная глава посвящена изучению свойства взаимодействия локальных периодов. Напомним, что длину взаимодействия локальных периодов можно указать только для частичных слов, не содержащих *специальных* подпоследовательностей. В бесконечном частичном слове вероятность наличия специальных подпоследовательностей равна единице (если джокеры распределены равномерно, т.е. джокер с равной ненулевой вероятностью может находиться в любой позиции). Кроме того, выполнение свойства взаимодействия локальных периодов (в обобщенной формулировке) также тесно связано с наличием в частичном слове специальных подпоследовательностей. Таким образом, данная глава посвящена изучению свойств специальных подпоследовательностей и их связи с различными вариантами обобщенного свойства взаимодействия локальных периодов. Прежде всего, необходимо найти необходимые и достаточные условия наличия в слове специальных подпоследовательностей. Эта задача рассматривается в §12. В этом параграфе также приводится описание полиномиального алгоритма с конечной памятью для проверки этих условий.

§10 Используемая техника

Напомним понятия, определенные во введении. Для данного частичного слова W рассмотрим множество частичных слов с теми же периодами и той же областью определения. Среди этих слов выберем слово U с максимальным количеством различных букв. Множество всех позиций слова U, содержащих одну и ту же букву, будем называть *доменом*

W. Количество доменов слова W (т.е. количество различных букв в U) назовем *размерностью* слова W и обозначим через $r(W)$. Из определений следует, что количество букв в частичном слове не превышает его размерность, а две позиции могут содержать различные буквы только если эти позиции принадлежат разным доменам. Обобщенное свойство взаимодействия периодов состоит в выполнении некоторых условий для множества доменов частичного слова.

Для изучения свойства взаимодействия локальных периодов используем граф локальной периодичности. Определим его аналогично определению графа периодичности в §1. Пусть частичное слово U имеет локальные периоды p, q. Будем соотносить с этим словом бинарное отношение R на множестве $D(U)$. Отношение состоит из всех пар (s, t), где $|s - t| = p$ или $|s - t| = q$. Отношение R является симметричным, поэтому пару $(D(U), R)$ можно рассматривать как неориентированный граф. Этот граф мы будем называть *графом локальной периодичности.* Отношение R^* (рефлексивно-транзитивное замыкание R) является отношением эквивалентности. В U на эквивалентных позициях обязательно стоят одинаковые буквы. Отношению эквивалентности R^* соответствует некоторое разбиение множества $D(U)$. Количество классов этого разбиения равно количеству компонент связности соответствующего графа и равно размерности слова U. Таким образом, множество доменов частичного слова совпадает с множеством компонент связности графа локальной периодичности.

Пример 3.1. Пусть $q = 3, p = 5$, $U = a_1\diamond a_3\diamond a_5\diamond a_7a_8a_9$. Граф обычного слова с такими периодами и граф локальной периодичности U представлены на рис.3.1. Размерность U равна 4 (компоненты связности графа локальной периодичности $\{1\}$, $\{7\}$, $\{9\}$ и $\{3, 5, 8\}$).

В дальнейшем мы будем рассматривать частичные Z-слова. Полученные результаты можно распространить и на конечные частичные слова, если рассматривать конечное частичное слово как частичное Z-слово с той же областью определения.

Граф локальной периодичности обычного Z-слова с двумя периодами можно изобразить на бесконечном цилиндре двумя системами спиралей (см. рис.3.2). Граф локальной периодичности частичного Z-слова с теми же периодами получается из этого графа удалением вершин, соответствующих позициям джокеров.

Нас интересует множество доменов частичного слова. Как уже отмечалось ранее, это множество зависит только от периодов частичного

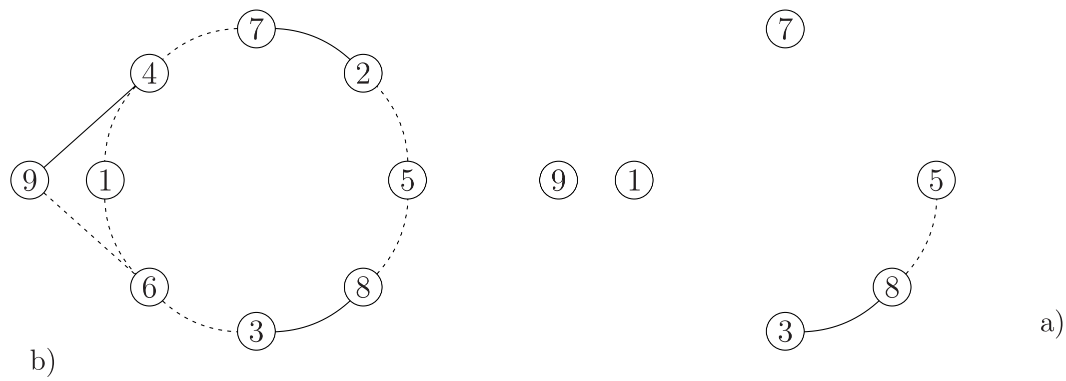


Рис. 3.1: Графы локальной периодичности обычного и частичного слова. Сплошные линии обозначают связь при помощи периода p, штриховые — при помощи q.

слова и расположения в нем джокеров. Поэтому в дальнейшем, как и в предыдущей главе, мы вместо слов будем рассматривать расстановки.

Для удобства граф локальной периодичности частичного Z-слова W будем изображать на плоскости. Для этого возьмем граф локальной периодичности обычного Z-слова и «разрежем» цилиндр вдоль. Полученный граф представим в виде таблицы: множество ячеек соответствует множеству вершин, ячейки находятся в таблице рядом (в одном ряду в соседних колонках или в одной колонке в соседних рядах) только тогда, когда в исходном графе соответствующие вершины соединены ребром.

В ячейке (x, y) запишем 0, если $W(xp + yq) = 0$, и 1 – в противном случае. Получим бесконечную в одну сторону таблицу ширины p. Теперь «восстановим» разорванные связи: добавим слева и справа по еще одной такой же таблице, слева – со сдвигом вниз на q строк, справа – со сдвигом вверх на q строк. Продолжим процесс добавления для только что добавленных таблиц, и т.д. Полученную бесконечную двумерную таблицу мы будем называть *решеткой*.

Выберем в расстановке некоторую начальную позицию (для удобства будем считать, что это позиция 0).

Будем называть *блоком* прямоугольную часть решетки размера $p \times q$,

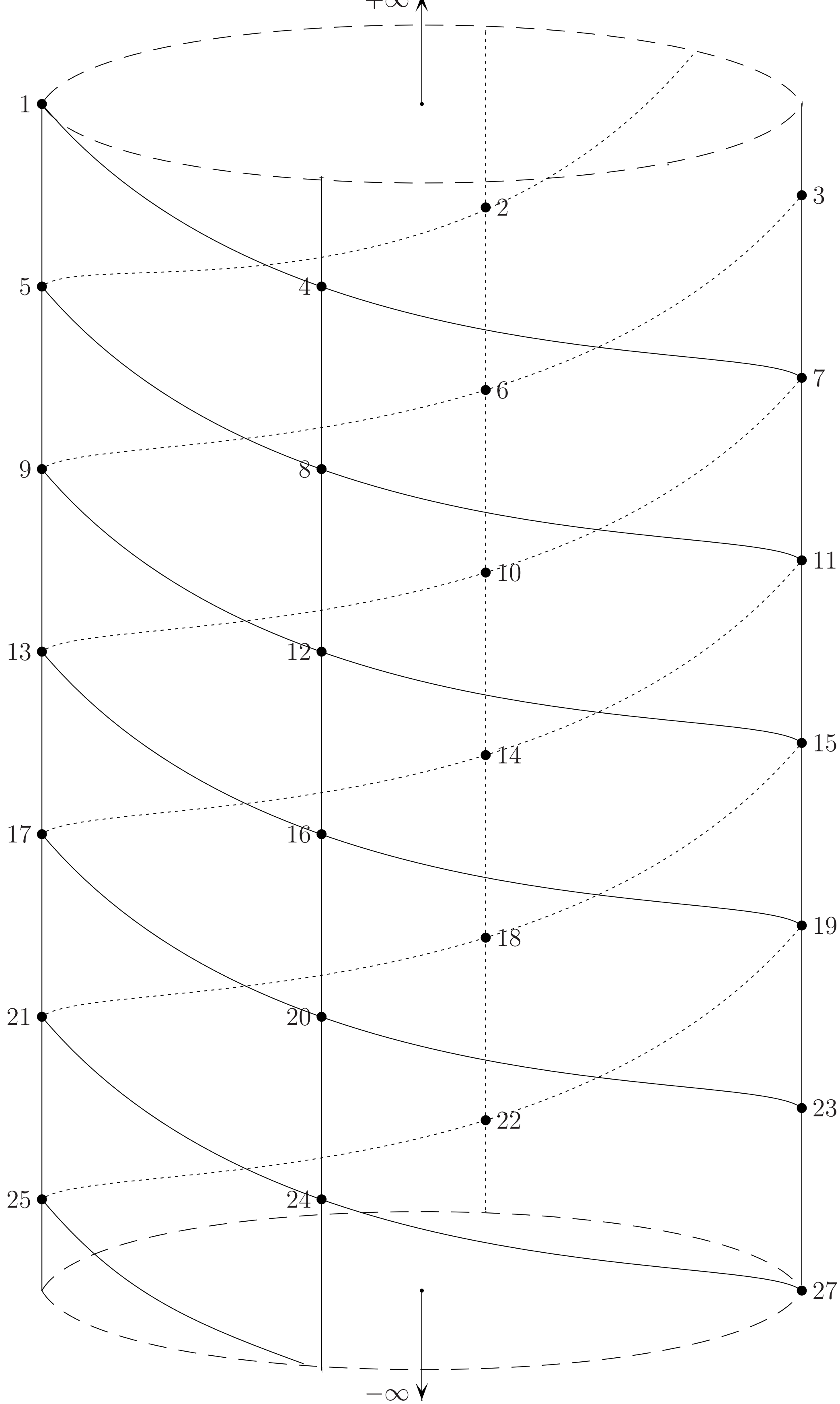


Рис. 3.2: Фрагмент графа локальной периодичности обычного бесконечного слова с периодами $q = 3, p = 4$.

т.е. множество позиций расстановки $\{n : n = apq + xp + yq, a \in \mathbb{Z}, x = 0, \ldots, q-1, y = 0, \ldots, p-1\}$.

Заметим, что по построению решетки одной позиции расстановки соответствует бесконечно много позиций решетки. Рассмотрим позицию x. Существует единственное разложение $x = apq+bp+cq, b = 0, \ldots, q-1, c = 0, \ldots, p-1$. Определим $a(x) = a, b(x) = b, c(x) = c$. Тогда множество ячеек решетки, соответствующих данной позиции x, можно определить следующим образом: $((a(x)-n)q + b(x), c(x) + np), n \in \mathbb{Z}$.

Пример 3.2. В следующей таблице приведена часть решетки бесконечной расстановки с периодами $5, 7$, соответствующая следующей подпоследовательности этой расстановки:
...0????0?0??1?1?00?0?1011?1101111101000101010111111011010100111...
Выделенные цветом позиции соответствуют одной и той же позиции исходной расстановки.

	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
0	0	0	1	0	0	1	1	0	0	1	1	1	1	0	0	0	1	1	1	1	1
1	0	1	0	1	1	0	1	0	1	0	0	0	1	1	1	1	1	0	0	1	0
2	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1
3	1	0	1	0	1	1	1	0	0	1	1	0	1	0	1	1	1	1	0	1	1
4	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	0
5	0	0	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1
6	0	1	0	0	0	1	1	1	1	1	0	0	1	0	1	1	1	1	0	1	1
7	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	1
8	0	0	1	1	0	1	0	1	1	1	1	0	1	1	0	0	1	1	1	0	1
9	0	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1
10	0	0	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	0	1
11	1	1	1	0	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1
12	1	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1
13	1	1	1	1	0	1	1	0	0	1	1	1	0	1	0	0	1	1	1	1	0
14	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0	0	1	1	1	0

В следующей таблице в ячейке (x, y) приведено значение $xp+yq$, т.е. в каждой ячейке указан номер позиции, значение из которой отображается в данной ячейке в решетке.

	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
0	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
1	7	12	17	22	27	32	37	42	47	52	57	62	67	72	77	82	87	92	97	102	107
2	14	19	24	29	34	39	44	49	54	59	64	69	74	79	84	89	94	99	104	109	114
3	21	26	31	36	41	46	51	56	61	66	71	76	81	86	91	96	101	106	111	116	121
4	28	33	38	43	48	53	58	63	68	73	78	83	88	93	98	103	108	113	118	123	128
5	35	40	45	50	55	60	65	70	75	80	85	90	95	100	105	115	120	125	130	135	140
6	42	47	52	57	62	67	72	77	82	87	92	97	102	107	112	117	122	127	132	137	142
7	49	54	59	64	69	74	79	84	89	94	99	104	109	114	119	124	129	134	139	144	149
8	56	61	66	71	76	81	86	91	96	101	106	111	116	121	126	131	136	141	146	151	156
9	63	68	73	78	83	88	93	98	103	108	113	118	123	128	133	138	143	148	153	158	163
10	70	75	80	85	90	95	100	105	115	120	125	130	135	140	145	150	155	160	165	170	175
11	77	82	87	92	97	102	107	112	117	122	127	132	137	142	147	152	157	162	167	172	177
12	84	89	94	99	104	109	114	119	124	129	134	139	144	149	154	159	164	169	174	179	184
13	91	96	101	106	111	116	121	126	131	136	141	146	151	156	161	166	171	176	181	186	191
14	98	103	108	113	118	123	128	133	138	143	148	153	158	163	168	173	178	183	188	193	198

§11 Разрезы

Пусть J – некоторое множество позиций. Обозначим через $P(J)$ Z-расстановку $W:(W(x)=0)\Leftrightarrow(x\in J)$. Будем называть *разрезом* расстановки W минимальное по включению множество позиций джокеров T такое, что расстановка $P(T)$ содержит не менее двух доменов. Таким образом, разрезы – это вершинные разрезы графа периодичности обычного Z-слова с периодами p,q.

Пример 3.3. В следующих четырех таблицах приведена одна и та же часть решетки некоторой расстановки. Красным цветом выделены несколько разрезов.

0	0	1	0	0	0	0	1
0	0	0	0	0	0	1	0
0	1	1	1	1	1	0	0
0	1	0	0	0	1	0	0
0	1	0	1	0	1	0	0
0	1	1	0	1	1	0	0
0	0	0	0	0	0	0	0

0	0	1	0	0	0	0	1
0	0	0	0	0	0	1	0
0	1	1	1	1	1	0	0
0	1	0	0	0	1	0	0
0	1	0	1	0	1	0	0
0	1	1	0	1	1	0	0
0	0	0	0	0	0	0	0

$$
\begin{array}{|cccccccc|}
\hline
0 & {\color{red}0} & 1 & {\color{red}0} & 0 & 0 & {\color{red}0} & 1 \\
{\color{red}0} & 0 & {\color{red}0} & 0 & {\color{red}0} & {\color{red}0} & 1 & {\color{red}0} \\
{\color{red}0} & 1 & 1 & 1 & 1 & 1 & 0 & {\color{red}0} \\
{\color{red}0} & 1 & 0 & 0 & 0 & 1 & 0 & {\color{red}0} \\
{\color{red}0} & 1 & 0 & 1 & 0 & 1 & {\color{red}0} & 0 \\
{\color{red}0} & 1 & 1 & 0 & 1 & 1 & 0 & {\color{red}0} \\
0 & {\color{red}0} & {\color{red}0} & {\color{red}0} & {\color{red}0} & {\color{red}0} & {\color{red}0} & 0 \\
\hline
\end{array}
\qquad
\begin{array}{|cccccccc|}
\hline
0 & 0 & 1 & 0 & 0 & 0 & {\color{red}0} & 1 \\
0 & 0 & 0 & 0 & 0 & {\color{red}0} & 1 & {\color{red}0} \\
0 & 1 & 1 & 1 & 1 & 1 & {\color{red}0} & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\end{array}
$$

В данном разделе мы рассмотрим свойства разрезов. Эти свойства будут использованы в §12 для конструирования алгоритма проверки наличия в расстановке специальной подпоследовательности.

Сначала введем необходимые определения. На множестве позиций расстановки определим следующие отношения. Отношение ρ («иметь одинаковые значения»):

$$x \rho y \Leftrightarrow W(x) = W(y).$$

Отношение ϕ («быть соседом»):

$$x \phi y \Leftrightarrow (x - y = ap + bq, a, b \in \{-1, 0, 1\}).$$

Отношение ψ («быть близким соседом»):

$$x \psi y \Leftrightarrow (|x - y| = q) \vee (|x - y| = p).$$

Таким образом, $\psi \subset \phi$. Отношения ϕ и ψ симметричны. Рассмотрим также отношения $\phi_1 = \phi \cap \rho$ и $\psi_1 = \psi \cap \rho$. Эти отношения также симметричны. Рефлексивно-транзитивные замыкания двух последних отношений будут являться отношениями эквивалентности. Назовем отношение ϕ_1^* *отношением связанности*, а отношение ψ_1^* – *отношением сильной связанности*.

Определим линейный порядок на множестве позиций как лексикографический порядок троек $(a(x), b(x), c(x))$, т.е.

$$x \prec y \Leftrightarrow (a(x) < a(y)) \vee$$

$$\vee((a(x) = a(y)) \wedge (b(x) < b(y))) \vee$$

$$\vee((a(x) = a(y)) \wedge (b(x) = b(y)) \wedge (c(x) < c(y)))$$

В дальнейшем сравнение позиций будет производиться в порядке $\prec$. Наибольшую позицию разреза будем называть *замыкающей*.

Путь – это последовательность позиций расстановки $x_1, \ldots, x_n$ такая, что для $i = 1, \ldots, n-1$ выполняется $x_i \psi x_{i+1}$. *Конечная цепочка* – это последовательность позиций джокеров расстановки $x_1, \ldots, x_n$ такая, что $x_i \phi_1 x_{i+1}, i = 1, \ldots, n-1$, $x_i \phi x_j \Leftrightarrow |j - i| = 1$ и если $x_i = x_j$, то либо $i = j$, либо $j = n$ и $i = 1$. *Бесконечная цепочка* — это последовательность позиций джокеров расстановки $x_i, i \in \mathbb{Z}$ такая, что $x_i \phi_1 x_{i+1}, i \in \mathbb{Z}$, $x_i \phi x_j \Leftrightarrow |j - i| = 1$ и если $x_i = x_j$, то $i = j$. Бесконечную цепочку будем называть Z-цепочкой (ω-цепочкой), если это множество позиций составляет Z-слово (соответственно, ω-слово). Конечную цепочку будем называть *замкнутой*, если $x_1 = x_n$, $n > 4$. Таким образом, в замкнутой цепочке соседями могут быть только две последовательных позиции. Будем говорить, что две цепочки не пересекаются, если никакие две позиции из разных цепочек не являются соседями.

Предложение 3.1. *Для любой позиции x разреза T в расстановке $P(T)$ есть два близких соседа позиции x, принадлежащие различным доменам.*

Доказательство. По определению разреза при замене в любой из позиций расстановки $P(T)$ джокера на букву получившаяся расстановка содержит один домен. Таким образом, при замене в позиции x джокера на букву в расстановке появляется путь, соединяющий позиции из разных доменов, проходящий через x. Это означает, что два близких соседа x принадлежат различным доменам. $\square$

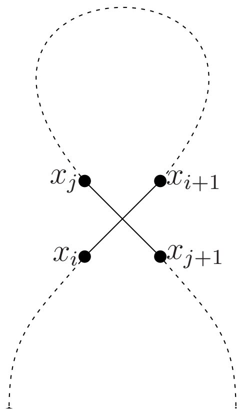
Предложение 3.2. *Пусть T – разрез. Для любой позиции $x \in T$ можно указать позиции джокеров $y, z \in T, y \neq z$, являющиеся соседями x.*

Доказательство. Рассмотрим соседей позиции x. Если среди соседей x присутствует менее двух позиций джокеров, то все соседи x сильно связаны, что противоречит предложению 3.1. $\square$

Из предложения 3.2 следует, что любой разрез представляет собой объединение бесконечных и конечных замкнутых цепочек. Таким образом, представляет интерес изучение свойств таких цепочек. На основе этих свойств мы приведем классификация разрезов.

Пусть граф локальной периодичности частичного Z-слова изображен на поверхности бесконечного цилиндра, как на рис. 3.2. Соединим ломаной последовательные позиции каждой цепочки.

Предложение 3.3. *На поверхности цилиндра каждой цепочке соответствует линия без самопересечений.*



Доказательство. Предположим, что линия, построенная по цепочке, имеет самопересечение.

Позиции x_i и x_{i+1} (а также x_j и x_{j+1}) являются соседями. Т.к. расстояние между x_i и x_j меньше, чем расстояние между x_i и x_{i+1}, то позиции x_i и x_j тоже являются соседями, что противоречит определению цепочки, т.к. $|i - j| > 1$. □

Для классификации разрезов воспользуемся следующими фундаментальными результатами о топологии поверхности цилиндра (см., например, [37]):

1. Поверхность бесконечного цилиндра гомеоморфна сфере (т.е. ее свойства, касающиеся связности и «разрезаемости», вытекают из соответствующих свойств сферы). При этом некоторые две точки сферы соответствуют бесконечно удаленным точкам цилиндра. Будем считать, что эти точки — полюса сферы.

2. Сфера односвязна (теорема Пуанкаре). В частности, кривые, разрезающие сферу на две области — это в точности кривые, гомеоморфные окружности (т.е. замкнутые кривые без самопересечений).

Существует четыре варианта расположения окружности на сфере относительно полюсов (см. рис. 3.3).

a. Линия не проходит через полюса, полюса принадлежат разным областям (см. рис. 3.3 а). На поверхности цилиндра этой линии соответствует конечная замкнутая линия, которая делит поверхность цилиндра на две бесконечные области (т.е. «разрезает цилиндр поперёк»). Если линия такого типа построена на цилиндре по цепочке, то это — замкнутая цепочка; такую цепочку будем называть *преградой*.

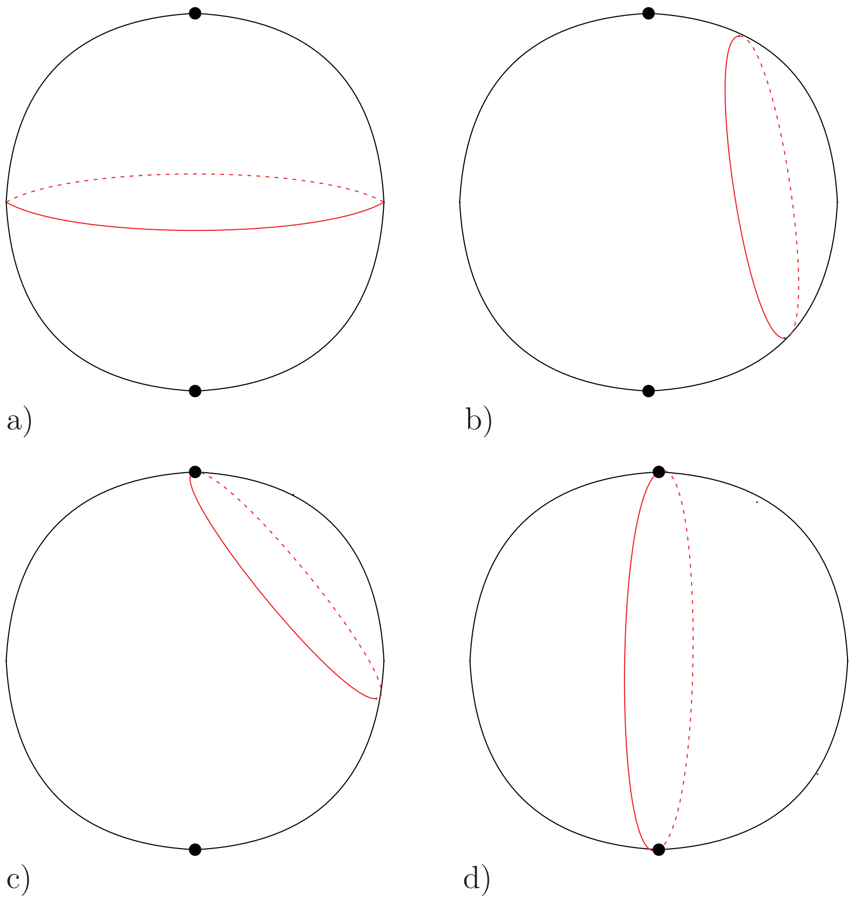


Рис. 3.3: Варианты расположения замкнутой линии без самопересечений на поверхности сферы.

b. Линия не проходит через полюса, оба полюса принадлежат одной области (см. рис. 3.3 b). На поверхности цилиндра этой линии соответствует конечная замкнутая линия, которая делит поверхность цилиндра на бесконечную и конечную область. Если линия такого типа построена на цилиндре по цепочке, то это — замкнутая цепочка; такую цепочку будем называть *ловушкой*.

c. Линия проходит через только один из полюсов (см. рис. 3.3 c). Если линия такого типа построена на цилиндре по цепочке, то это — ω цепочка.

d. Линия проходит через оба полюса (см. рис. 3.3 d). Если линия такого типа построена на цилиндре по цепочкам, то это — две непересекающихся Z-цепочки.

Предложение 3.4. *1. Замкнутая цепочка является разрезом.*

2. ω-цепочка является разрезом.

3. Объединение двух непересекающихся Z-цепочек является разрезом.

4. Других разрезов нет.

Доказательство. Пусть множество T — одно из множеств позиций джокеров, указанных в условии предложения. Из рассмотренного выше следует, что линия, построенная по цепочкам множества T, выделяет на поверхности цилиндра две непрерывных замкнутых области. Докажем, что внутри каждой из этих областей есть хотя бы одна вершина графа (а следовательно, позиция буквы).

1. Пусть множество T является замкнутой цепочкой. Рассмотрим два возможных случая расположения замкнутой цепочки на цилиндре. Во-первых, замкнутая цепочка может являться преградой, т.е. выделять на поверхности цилиндра две бесконечных замкнутых области (см. случай a). В таком случае внутри каждой из этих областей есть хотя бы одна вершина графа. Во-вторых, замкнутая цепочка может являться ловушкой, т.е. выделять на поверхности цилиндра конечную и бесконечную область (см. случай b). Докажем, что внутри конечной области есть хотя бы одна вершина графа. Пусть x и y — наименьшая и наибольшая позиции цепочки T, т.е. цепочка представима в виде $x = x_1, x_2, \ldots, x_k = y, x_{k+1}, \ldots, x_{n-1}, x_n = x$. Любая позиция из цепочки лежит в решетке либо ниже x, либо в той же строке и правее; либо выше y, либо в той же строке и левее. Т.к. $n > 4$, то если позиции x и y лежат в одной и той же строке или в соседних строках, то для некоторых позиций из T нарушится условие $x_i \phi x_j \Leftrightarrow |j - i| = 1$. Таким образом, существует строка решетки такая, что позиция x находится выше ее, а позиция y — ниже. Т.к. y не является соседом x, выполняется $y \neq x_2, y \neq x_{n-1}$. Позиции x_2 и x_{n-1} различны и являются соседями x, но не являются соседями друг друга. В силу минимальности x хотя бы одна из этих позиций лежит в следующей строке решетки. Пусть x_2 находится в решетке левее x_{n-1}. Тогда все возможные варианты расположения позиций x_2 и x_{n-1} относительно позиции x изображены на рис. 3.4.

   Позиция, которая является правым соседом позиции x_2, не может принадлежать цепочке, а значит, является позицией буквы. Эта позиция принадлежит рассматриваемой конечной области. Таким

	x	x_{n-1}
x_2		
$\dots$	$\dots$	$\dots$
		y

	x	x_{n-1}
	x_2	
$\dots$	$\dots$	$\dots$
		y

	x	
x_2		x_{n-1}
$\dots$	$\dots$	$\dots$
		y

Рис. 3.4: Возможные варианты расположения позиций x_2 и x_{n-1}.

образом, внутри каждой из выделяемых замкнутой цепочкой областей есть хотя бы одна вершина графа.

2. Пусть множество T является ω-цепочкой (см. случай c). Тогда начиная с некоторой строки решетки, в любой строке присутствует не менее двух позиций из T, не являющихся соседями. Т.к. они не являются соседями, то с обеих сторон между ними есть позиции буквы. Таким образом, внутри каждой из выделяемых ω-цепочкой областей есть хотя бы одна вершина графа.

3. Пусть множество T является объединением двух непересекающихся Z-цепочек (см. случай d). В каждой строке решетки присутствует не менее одной позиции из каждой Z-цепочки. Поскольку цепочки не пересекаются, эти позиции не могут быть соседями. Т.к. они не являются соседями, то с обеих сторон между ними есть позиции буквы. Таким образом, внутри каждой из выделяемых двумя непересекающимися Z-цепочками областей есть хотя бы одна вершина графа.

Любой путь из позиции одной области в позицию другой области пройдет через позицию из множества T. Это означает, что в расстановке $P(T)$ два домена, т.е. T содержит разрез. При замене в одной из позиций множества T джокера на букву появится путь, соединяющий позиции из разных областей и не проходящий через позиции из T, т.е. размерность T уменьшится. Это означает, что множество T совпадает с содержащимся в нем разрезом.

$\square$

Пример 3.4. Пусть $q = 3, p = 4$. Множество позиций $T = \{9, 10, 11, 12\}$ составляет замкнутую цепочку — преграду. В следующей таблице приведен фрагмент решетки расстановки $P(T)$, на рис. 3.5 — фрагмент графа локальной периодичности этой расстановки.

	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
-6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
-5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
-4	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
-3	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1
-2	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
-1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1
1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
3	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
4	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Пример 3.5. Пусть $q = 3, p = 4$. Множество позиций $T = \{8, 9, 13, 15, 19, 20\}$ составляет замкнутую цепочку — ловушку. В следующей таблице приведен фрагмент решетки расстановки $P(T)$, на рис. 3.6 — фрагмент графа локальной периодичности этой расстановки.

	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
-6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
-5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
-4	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
-3	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1
-2	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1
-1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1
0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1
1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1
2	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1
3	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1
4	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1
5	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1

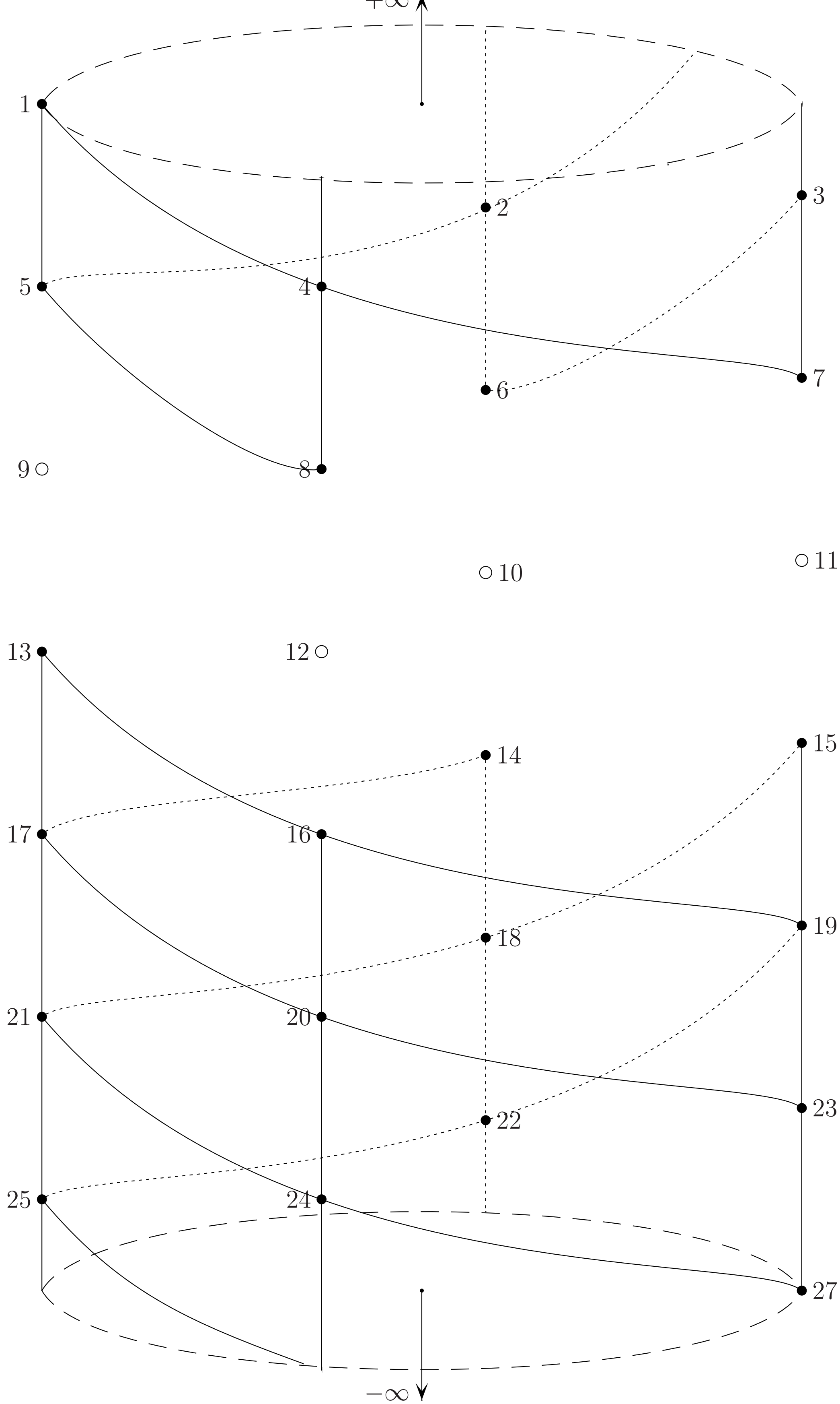


Рис. 3.5: Фрагмент графа локальной периодичности расстановки $P(T)$, $T = \{9, 10, 11, 12\}$ с периодами $q = 3, p = 4$.

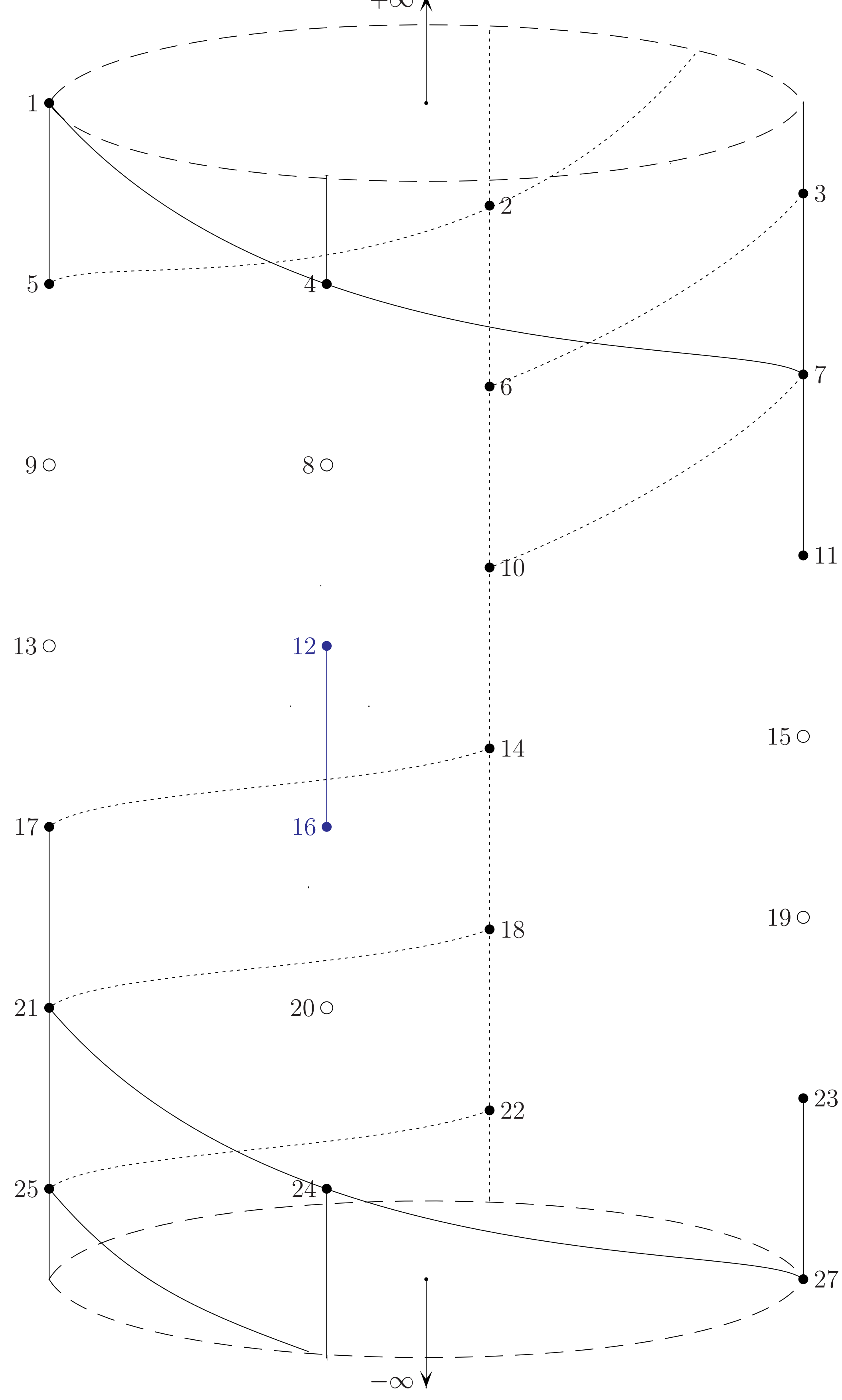


Рис. 3.6: Фрагмент графа локальной периодичности расстановки $P(T)$, $T = \{8, 9, 13, 15, 19, 20\}$ с периодами $q = 3, p = 4$.

Пример 3.6. Пусть $q = 3, p = 4$. Множество позиций $T = \{11\} \cup \{8 + np, n \in \mathbb{N}\} \cup \{14 + np, n \in \mathbb{N}\}$ составляет бесконечный разрез, являющийся ω-цепочкой. В следующей таблице приведен фрагмент решетки расстановки $P(T)$, на рис. 3.7 — фрагмент графа локальной периодичности этой расстановки.

	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
-6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-4	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
-3	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
-2	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
-1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	1
0	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	1
1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	1
2	1	1	1	1	1	0	1	1	0	1	0	1	0	1	0	1
3	1	1	1	1	0	1	0	1	0	1	0	1	0	1	0	1
4	1	1	1	1	0	1	0	1	0	1	0	1	0	1	0	1
5	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1

Пример 3.7. Пусть $q = 3, p = 4$. Бесконечное множество позиций $T = \{np, n \in \mathbb{Z}\} \cup \{2 + np, n \in \mathbb{Z}\}$ составляет бесконечный разрез, являющийся объединением двух Z-цепочек. В следующей таблице приведен фрагмент решетки расстановки $P(T)$, на рис. 3.8 — фрагмент графа локальной периодичности этой расстановки.

	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
-6	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
-5	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
-4	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
-3	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
-2	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
-1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
2	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
3	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
4	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
5	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

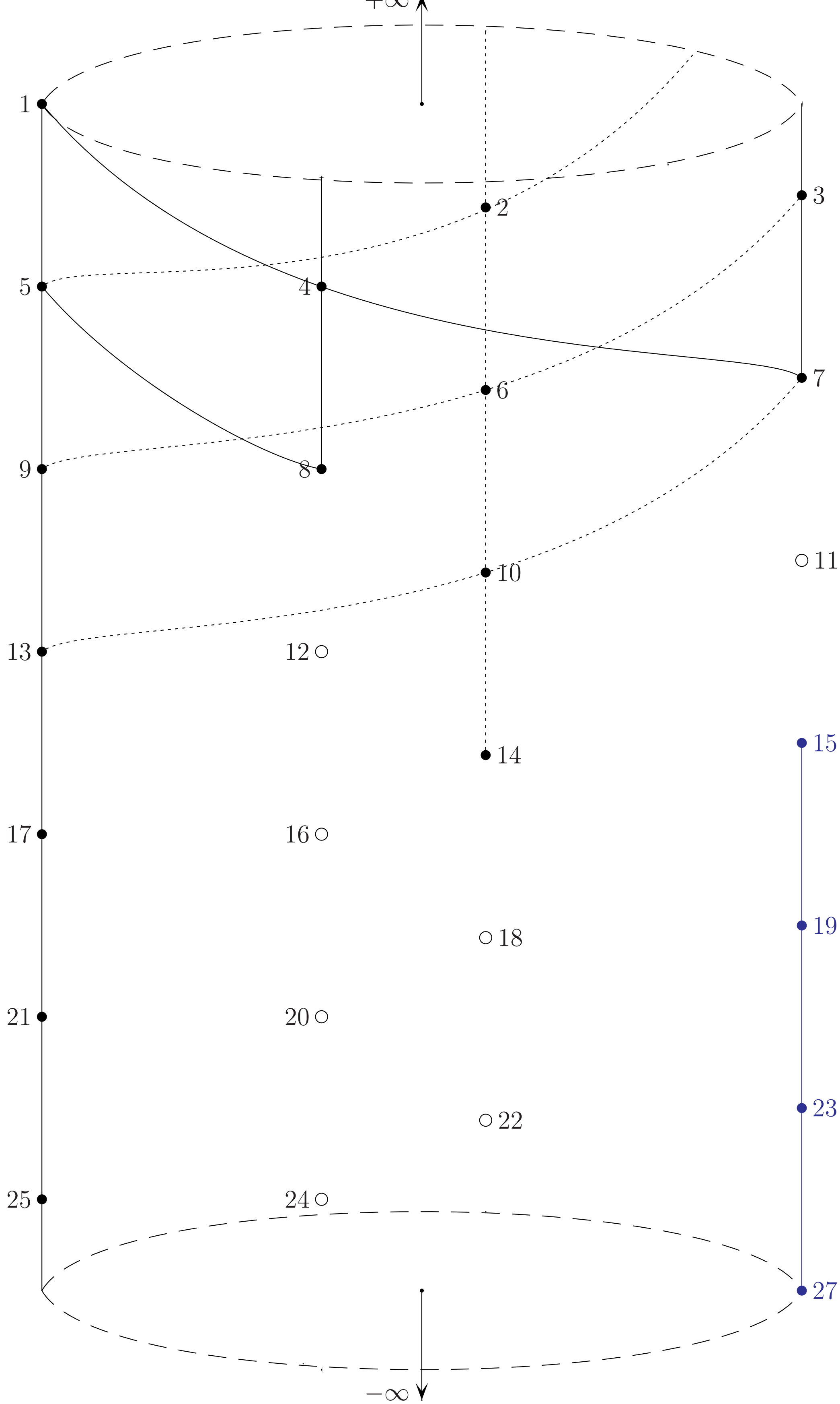


Рис. 3.7: Фрагмент графа локальной периодичности расстановки $P(T)$, $T = \{11\} \cup \{8 + np, n \in \mathbb{N}\} \cup \{14 + np, n \in \mathbb{N}\}$ с периодами $q = 3, p = 4$.

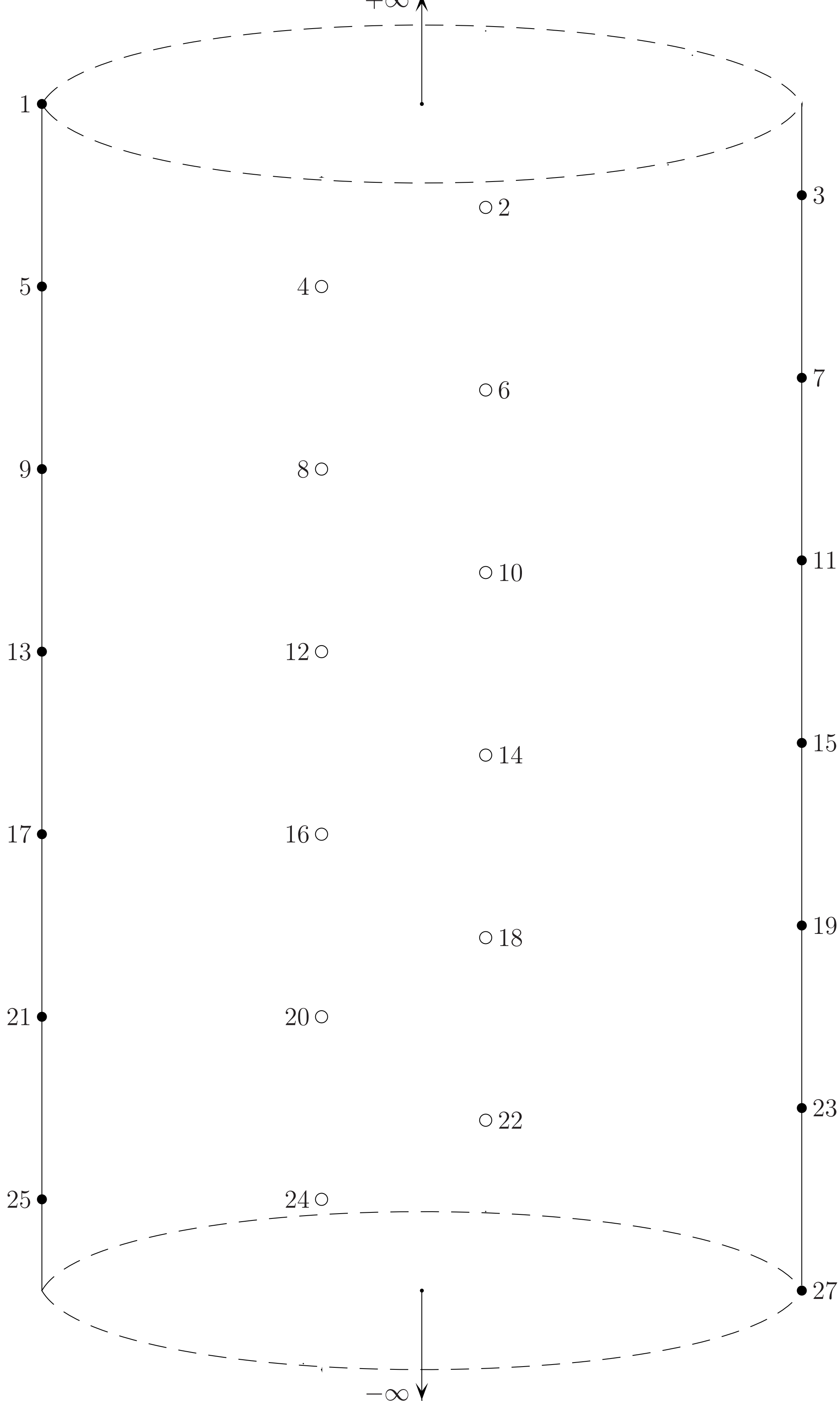


Рис. 3.8: Фрагмент графа локальной периодичности расстановки $P(T)$, $T = \{np, n \in \mathbb{Z}\} \cup \{2 + np, n \in \mathbb{Z}\}$ с периодами $q = 3, p = 4$.

Пусть T — разрез расстановки W. Рассмотрим произвольные позиции x, y из области определения исходной расстановки W. Если в расстановке $P(T)$ позиции x, y принадлежат различным доменам, то и в исходной расстановке они принадлежат различным доменам. Таким образом, разрезу расстановки W соответствует разбиение множества доменов расстановки W на два класса. При этом один из классов может быть пустым (оба класса могут быть пусты только если $D(W) = \emptyset$).

Будем считать, что множество номеров позиций букв в Z-расстановке не ограничено. Напомним, что *специальные* подпоследовательности — это конечные подпоследовательности такого вида, что для расстановок, содержащих такие подпоследовательности, не выполняется свойство взаимодействия периодов (в классической формулировке), т.е. нельзя указать такую плотность джокеров, что размерность любой расстановки, плотность джокеров в которой не превосходит указанную, обязательно равна единице.

Из определений следует, что любой конечный разрез является специальной подпоследовательностью, а любая специальная подпоследовательность содержит хотя бы один конечный разрез.

Таким образом, для проверки наличия в расстановке специальной подпоследовательности достаточно проверить, содержит ли данная расстановка конечный разрез. В следующем разделе мы приведем соответствующий алгоритм.

§12 Алгоритм проверки наличия в расстановке специальной подпоследовательности

В данном разделе представлен алгоритм проверки наличия в расстановке специальной подпоследовательности. На вход алгоритма подается некоторое конечное подслово Z-расстановки. Алгоритм осуществляет проверку наличия в этом подслове конечного разреза. Время работы данного алгоритма линейно зависит от длины рассматриваемой расстановки, а объем используемой памяти постоянен (если считать периоды p, q параметрами).

Для каждой позиции x определим *метку* следующим образом

$$Label(x) = \begin{cases} \min_{\prec}\{y : y\phi_1 x_1 \phi_1 x_2 \phi_1 \ldots \phi_1 x_n \phi_1 x, \\ \qquad\qquad\qquad y, x_1, \ldots, x_n \prec x\}, & W(x) = 0 \\ \qquad\qquad\qquad\qquad \infty, & W(x) = 1. \end{cases}$$

По предложению 3.2 все позиции разреза связаны между собой, поэтому все позиции разреза будут иметь одинаковую метку.

Предложение 3.5. *Позиция x, замыкающая разрез, имеет двух соседей y, z с одинаковой конечной меткой. При этом $y \prec x, z \prec x$.*

Доказательство. Согласно предложению 3.4, конечный разрез является замкнутой цепочкой. Таким образом, у позиции x есть 2 соседа y, z, принадлежащих разрезу, причем все остальные позиции разреза представляют собой цепочку, соединяющую y и z. Поскольку x является наибольшей позицией разреза, все остальные позиции разреза имеют меньшие номера. Это означает, что y и z имеют одинаковые конечные метки. □

В этом алгоритме позиции частичного слова рассматриваются в порядке $\prec$. Для вычисления метки текущей позиции используются вспомогательные массивы B размера p и D размера q, в которых хранятся метки нескольких уже просмотренных позиций.

Алгоритм 3.1. (Проверяет наличие в расстановке W специальных подпоследовательностей.)
Шаг 1. Произвести начальное заполнение вспомогательных массивов B размера p и D размера q. В качестве начальных данных в массивы записывается ∞. Объявить текущей позицией наименьшую позицию W.
Шаг 2. Проверить для текущей позиции выполнение условий из предложения 3.5. Если эти условия выполняются, то объявить текущую позицию замыкающей границу, вернуть значение «Да» и закончить работу.
Шаг 3. Вычислить метку текущей позиции на основании данных вспомогательных массивов B, содержащего сведения о метках предыдущей строки и D, содержащего сведения о метках правой колонки предыдущего блока. Метка текущей позиции устанавливается равной минимуму из меток ее соседей с меньшими номерами. При необходимости обновить данные в массивах B и D. Если текущая позиция является наибольшей позицией W, то вернуть значение «Нет» и закончить работу. Иначе объявить текущей позицией следующую и вернуться на Шаг 2.

Пример 3.8. Пример работы алгоритма 3.1 приведен в следующей таблице. В ячейках, соответствующих позициям джокеров, указана их метка на момент окончания просмотра данной строки. Остальные ячейки заняты буквами, их метки ∞. Желтым цветом выделены позиции, принадлежащие разрезу, красным цветом — позиция, замыкающая разрез.

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
1	1,1			1,4		1,6							1,4	1,4
2				1,4	1,4				7,2		7,4			
3						1,4	**1,4**	1,4		1,4			8,6	8,6
4	4,1			1,4	1,4	**1,4**		**1,4**		1,4	1,4		8,6	
5				1,4		**1,4**		**1,4**				1,4		1,4
6						1,4	**1,4**							
7		7,2		7,4										
8	1,4		1,4			8,6	8,6							
9	**1,4**		1,4	1,4		8,6								
10	**1,4**													

В ходе работы алгоритма ячейки просматриваются построчно: (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (1,7), (2,1), (2,2), и т.д. Во время просмотра позиции (10,1) массив B содержит значения ((1,4), ∞, ∞, ∞, (1,4), (1,4), ∞, (8,6)) — это метки позиций предыдущей строки, а массив D — значения ((1,4), ∞, (8,6), ∞, ∞). Для позиции (10,1) выполняется условие из предложения 3.5 — два ее соседа с меньшими номерами ((6,7) и (9,1)) имеют одинаковую конечную метку (1,4).

Теорема 3.1. *Время работы алгоритма 3.1 линейно зависит от длины исследуемой расстановки. Используемый объем памяти линейно зависит от периодов p, q.*

Доказательство. На шаге 1 начальное заполнение вспомогательных массивов B размера p и D размера q занимает время $O(p+q)$. В ходе дальнейшей работы алгоритма каждая позиция просматривается только один раз, и для нее выполняются шаги 2-3. На шаге 2 для проверки выполнения условий необходимо просмотреть не более пяти соседей текущей позиции, что требует постоянного количества операций. На шаге 3 для определения метки текущей позиции достаточно знать метки соседей текущей позиции с меньшими номерами (таких соседей также не более пяти). Эти позиции уже просмотрены, а значит, их метки уже определены. Для текущей позиции метка устанавливается равной минимуму из меток ее соседей с меньшими номерами. Это также требует постоянного количества операций.

Таким образом, выполнение шагов 2-3 требует постоянного количества операций, а количество выполнений шагов 2-3 линейно зависит от длины просматриваемой расстановки. Это означает, что время работы алгоритма линейно зависит от длины просматриваемой расстановки.

Для хранения меток достаточно использовать массив размера p, содержащий сведения о метках предыдущей строки, и массив размера q, содержащий сведения о метках правой колонки предыдущего блока, т.к. по свойствам решетки позиции этой колонки являются левыми соседями позиций из левой колонки текущего блока. Таким образом, используемый объем памяти линейно зависит от периодов p, q. $\square$

§13 Модификации алгоритма

В данном параграфе мы рассмотрим две модификации алгоритма , Первая модификация предназначена для определиения количества доменов в конечной расстановке. Вторая модификация позволяет искать разрезы определенного типа — ловушки или преграды.

Границей домена D назовем множество позиций джокеров, являющихся близкими соседями позиций из домена D. Наибольшую (в смысле порядка $\prec$) позицию границы будем называть *замыкающей*. Граница может быть как конечной, так и бесконечной. Граница может являться разрезом; может содержать разрез, но не совпадать с ним; может не содержать разрез.

Пример 3.9. Продолжим рассмотрение части решетки расстановки из примера 3.3. В следующих таблицах отмечены позиции границ, синим — позиции соотвествующих доменов.

0	0	1	0	0	0	0	1
0	0	0	0	0	0	1	0
0	1	1	1	1	1	0	0
0	1	0	0	0	1	0	0
0	1	0	1	0	1	0	0
0	1	1	0	1	1	0	0
0	0	0	0	0	0	0	0

0	0	1	0	0	0	0	1
0	0	0	0	0	0	1	0
0	1	1	1	1	1	0	0
0	1	0	0	0	1	0	0
0	1	0	1	0	1	0	0
0	1	1	0	1	1	0	0
0	0	0	0	0	0	0	0

Граница, выделенная в левой таблице, является разрезом. Граница, выделенная в правой таблице, содержит два разреза и не является их объединением. Кроме того, последняя граница содержит в себе первую.

В следующей таблице приведена часть решетки Z-расстановки размерности 1. Все позиции букв этой расстановки принадлежат ее единственному домену, все позиции джокеров принадлежат бесконечной границе этого домена. Эта граница не содержит разрез.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1
1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1
2	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1
3	1	0	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	1
4	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1
5	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1
6	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1
7	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0
8	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1
9	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
10	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
11	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1
12	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1
13	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
14	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1

В этом разделе мы будем рассматривать конечные расстановки. Особенность рассмотрения конечных расстановок в том, что начало и конец расстановки также участвуют в «ограничении» доменов. Для удобства в дальнейшем будем предполагать, что к расстановке добавлено по p джокеров в начало и в конец (от этого размерность расстановки не изменится). Легко проверить, что множество из p последовательных джокеров является разрезом.

Предложение 3.6. *Любая позиция конечного домена меньше, чем замыкающая позиция границы этого домена.*

Доказательство. Наибольшая позиция домена обязательно имеет близкого соседа — позицию джокера с большим номером. Эта позиция принадлежит границе домена, а значит, не больше, чем замыкающая позиция домена. □

Предложение 3.7. *Граница любого домена конечной расстановки, содержит хотя бы один разрез. Позиция, замыкающая границу такого домена, является замыкающей позиций хотя бы одного из содержащихся в границе разрезов.*

Доказательство. Пусть n — позиция, замыкающая границу T домена D расстановки W. Если позиция n является последней позицией расстановки W, то она замыкает разрез, состоящий из p последних джокеров. Если позиция n не является последней позицией расстановки W, то в расстановке $P(T)$ позиция m, следующая за n в порядке $\prec$, не принадлежит домену D (по предложению 3.6) и содержит букву (т.к. не принадлежит T). Близкими соседями любой позиции из домена D являются позиции из D или джокеры из границы T, а значит, любой путь из любой позиции из D в позицию $m+1$ проходит через позицию границы T. Это означает, что в расстановке $P(T)$ позиции из D и позиция m принадлежат различным доменам, т.е. T содержит разрез.

При замене в позиции m джокера на букву появится путь из позиции из домена D в позицию n, не принадлежащую D. Это означает, что позиция m, замыкающая границу, принадлежит некоторому разрезу. $\square$

Пусть позиция x замыкает границу домена. Обозначим соседей позиции x следующим образом:

NW	N	NE
W	■	E
SW	S	SE

(закрашенная клетка в центре соответствует позиции x).

Предложение 3.8. *Для того, чтобы позиция замыкала границу домена, необходимо и достаточно, чтобы для соседей данной позиции выполнялось одно из следующих условий:*

1. $c(x) > 0$, *в позиции N находится буква, позиция NE и одна из позиций W, NW имеют одинаковую конечную метку;*
2. $c(x) = 0$,
   (a) *в позиции W находится буква, позиция SW и одна из позиций NW, N, NE имеют одинаковую конечную метку;*
   (b) *в позиции N находится буква, позиция NE и одна из позиций SW, W, NW имеют одинаковую конечную метку.*

Доказательство. $\Rightarrow$ Пусть позиция x замыкает границу. Согласно предложению 3.5, среди соседей x с меньшим номером имеются две позиции джокеров с одинаковой конечной меткой. Среди близких соседей x с меньшим номером имеется позиция буквы. Рассмотрим возможные случаи.

1. $c(x) > 0$. В этом случае все рассматриваемые позиции принадлежат одному блоку. В ходе работы алгоритма рассматриваемые позиции просматриваются в следующем порядке: NW, N, NE, W, x, E, SW, S, SE. Соседи, меньшие x – это позиции W, NW, N, NE; близкие соседи, меньшие x – позиции W, N.

   (a) Предположим, что буква находится в позиции N.

      i. Пусть в позиции W находится буква. Тогда позиции джокеров с одинаковой меткой – это позиции NW, NE.

      ii. Пусть в позиции 2 находится джокер. Тогда позиция NE должна принадлежать разрезу (в противном случае в расстановке $P(J \cup x)$ все близкие соседи позиции x сильно связаны, что противоречит наличию в этой расстановке разреза). Тогда позиции джокеров с одинаковой меткой – это позиции NE и одна из позиций W, NW.

   (b) Предположим, что в позиции N находится джокер, тогда буква находится в позиции W. В расстановке $P(J \cup x)$ все близкие соседи позиции x сильно связаны, что противоречит наличию в этой расстановке разреза.

   Итак, в этом случае в позиции N обязательно должна присутствовать буква, а позиции джокеров с одинаковой меткой – это позиции NE и одна из позиций W, NW.

2. $c(x) = 0$. В этом случае позиции NW, W, SW принадлежат одному блоку, а все остальные – следующему блоку. В ходе работы алгоритма рассматриваемые позиции просматриваются в следующем порядке: NW, W, SW, N, NE, x, E, S, SE. Соседи, меньшие x – это позиции SW, W, NW, N, NE, близкие соседи, меньшие x – позиции W, N.

   (a) Предположим, что буква находится в позиции W. Тогда (по соображениям, аналогичным рассмотренным выше) позиции джокеров с одинаковой меткой – это позиции SW и одна из позиций NW, N, NE.

   (b) Предположим, что буква находится в позиции N. Тогда позиции джокеров с одинаковой меткой – это позиции NE и одна из позиций SW, W, NW.

$\Leftarrow$ Пусть для позиции x выполняется одно из указанных в теореме условий. Т.к. среди близких соседей x с меньшим номером имеется позиция буквы из некоторого домена D, то позиция x принадлежит границе домена D. Т.к. два меньших соседа позиции x имеют одинаковую конечную метку, то существует цепочка, соединяющая эти позиции и не проходящая через x. Добавление к этой цепочке позиции x делает ее замкнутой. Таким образом, x замыкает разрез, внутри которого находится домен D. Любая позиция, большая x, не быть близким соседом позиции из D, а значит, x — наибольшая позиция границы домена D.

$\square$

Алгоритм 3.1 можно модифицировать для поиска позиций, замыкающих границы доменов (и определения количества доменов). Для этого Шаг 2 необходимо изменить следующим образом:

Шаг 2. Проверить для текущей позиции выполнение условий из предложения 3.8. Если эти условия выполняются, то объявить текущую позицию замыкающей границу.

По позиции, замыкающей границу, можно также за время, линейно зависящее от размера домена (т.е. от длины расстановки) определить множество позиций соответствующего домена и границу этого домена, но для хранения этих результатов потребуется объем памяти, линейно зависящий от длины расстановки.

Пример 3.10. Пример работы модифицированного алгоритма 3.1 приведен в следующей таблице. В ячейках, соответствующих позициям джокеров, указана их метка на момент окончания просмотра данной строки. Желтым и синим цветом выделены позиции, принадлежащие границам доменов, красным цветом — позиции, замыкающие границы.

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
1	1,1		1,3	1,4		1,6							1,4	**1,4**
2				1,4	1,4				**7,2**		7,4			
3						1,4	**1,4**	**1,4**		**1,4**			8,6	**8,6**
4	4,1			1,4	1,4	**1,4**		**1, 4**		1,4	**1,4**		**8,6**	
5				1,4		**1,4**		**1,4**				**1,4**		**1,4**
6						1,4	**1, 4**							1,4
7		**7,2**		7,4				**1,4**	1,4	1,4		12,5		
8	**1,4**		**1,4**			8,6	**8, 6**	**1,4**				12,5		
9	**1,4**		1,4	**1,4**		**8, 6**			**1,4**					
10	**1,4**				**1,4**		**1,4**			**1,4**			15,6	
11							1,4	**1,4**	**1,4**		**1,4**	1,4		
12	**1,4**	1,4	1,4		12,5				**1,4**			**1,4**		
13	**1,4**				12,5				1,4	**1,4**	**1,4**	1,4		
14		**1,4**							1,4	**1,4**		**1,4**		
15			**1,4**			15,6		1,4		1,4	**1,4**	1,4	1,4	
16	**1,4**	**1,4**		**1,4**	1,4			1,4				1,4		
17		**1,4**			**1,4**			1,4		1,4		1,4		
18		1,4	**1,4**	**1,4**	1,4							1,4	1,4	
19		1,4	**1,4**		**1,4**				24,2	24,2		1,4		
20	1,4		1,4	**1,4**	1,4	1,4						1,4		25,7

При построении решетки мы «разрезали» граф локальной периодичности по вертикали. Одно из отличий преграды от ловушки состоит в том, что каждая вертикаль пересекает преграду нечетное число раз, а ловушку — четное число раз. Используем это отличие для модификации алгоритма 3.1 для поиска преград.

Введем еще два вспомогательных массива B' размера p и D' размера q. В этих массивах будет храниться информация о количестве пересечений вертикали, по которой был «разрезан» цилиндр, для позиции, метка которой хранится в той же позиции массива B или D. В начале массивы B' и D' заполняются значением 0. В ходе работы алгоритма при присваивании метке позиции x метки позиции y количество пересечений вертикали для позиции x устанавливается равным соответствующему количеству для позиции y, если позиции x и y принадлежат одному блоку, и на единицу больше, если позиции x и y принадлежат разным блокам. Если на некотором шаге обнаружена позиция, замыкающая разрез, это означает, что у данной позиции есть два соседа y, z с меньшими номерами, имеющие одинаковую конечную метку. Количество пересечений вертикали для разреза равно сумме соответствующих количеств для позиций y, z, если x, y, z принадлежат одному блоку; на единицу больше, если x и одна из позиций y, z принадлежат разным блокам; больше на два, если x и каждая из позиций y, z принадлежат разным блокам. Таким образом, если полученное количество нечетно, то замыкаемый данной позицией разрез является преградой.

Index

Литература

[1] *Thue A.* Über unendliche Zeichenreihen // *Norske Vid. Selsk. Skr. I Math-Nat. Kl.* — 1906. — Vol. 7. — Pp. 1–22.

[2] *Lothaire M.* Combinatorics on Words. — Addison-Wesley, 1983.

[3] *Lothaire M.* Algebraic combinatorics on Words. — Cambridge University Press, 2002.

[4] *Lothaire M.* Applied combinatorics on Words. — Cambridge University Press, 2005.

[5] *Choffrut C., Karhumäki J.* Combinatorics of words // Handbook of formal languages / Ed. by G.Rozenberg, A.Salomaa. — Springer, Berlin, 1997. — Vol. 1.

[6] *Berstel J., Boasson L.* Partial words and a theorem of Fine and Wilf // *Theor. Comp. Sci.* — 1999. — Vol. 218. — Pp. 135–141.

[7] *Head T., Păun G., Pixton D.* Language theory and molecular genetics // Handbook of formal languages / Ed. by G.Rozenberg, A.Salomaa. — Springer, Berlin, 1997. — Vol. 2.

[8] *Blanchet-Sadri F., Duncan S.* Partial words and the critical factorization theorem // *Journal of Combinatorial Theory, Series A.* — 2005. — Vol. 109. — Pp. 221–245.

[9] *Césari Y., Vincent M.* Une caractérisation des mots périodiques // *C. R. Acad. Sci. Paris.* — 1978. — Vol. 286. — Pp. 1175–1177.

[10] *Duval J.-P.* Une caractérisation des fonctions périodiques // *C. R. Acad Sci. Paris.* — 1979. — Vol. 289. — Pp. 185–187.

[11] *Blanchet-Sadri F., Wetzler N. D.* Partial words and the critical factorization theorem revisited. — в печати.

[12] *Guibas L., Odlyzko A.* Periods in strings // *Journal of Combinatorial Theory, Series A.* — 1981. — Vol. 30. — Pp. 19–42.

[13] *Halava V., Harju T., Ilie L.* Periods and binary words // *Journal of Combinatorial Theory, Series A.* — 2000. — Vol. 89. — Pp. 298–303.

[14] *Blanchet-Sadri F., Chriscoe A.* Local periods and binary partial words: an algorithm // *Theor. Comp. Sci.* — 2004. — Vol. 314. — Pp. 401–419.

[15] *Blanchet-Sadri F., Chen C.-T.* Periods and Binary Partial Words with Two Holes. — в печати.

[16] *Blanchet-Sadri F., Shirey B.* Periods, Partial Words, and a Result of Guibas and Odlyzko. — в печати.

[17] *Blanchet-Sadri F.* Primitive partial words // *Discrete Applied Mathematics.* — 2005. — Vol. 148. — Pp. 195–213.

[18] *Blanchet-Sadri F., Anavekar A. R.* Testing Primitivity on Partial Words. — в печати.

[19] *Blanchet-Sadri F.* Codes, orderings and partial words // *Theor. Comp. Sci.* — 2004. — Vol. 329. — Pp. 177–202.

[20] *Blanchet-Sadri F., Moorefield M.* Pcodes of Partial Words. — в печати.

[21] *Leupold P.* Languages of partial words - how to obtain them and what properties they have // *Grammars.* — 2004. — Vol. 7. — Pp. 179–192.

[22] *Fine N., Wilf H.* Uniqueness theorem for periodic functions // *Proc. Amer. Math. Soc.* — 1965. — Vol. 16. — Pp. 109–114.

[23] *Constantinescu S., Ilie L.* Generalised Fine and Wilf's theorem for arbitrary number of periods // *Theor. Comput. Sci.* — 2005. — Vol. 339, no. 1. — Pp. 49–60.

[24] *Castelli M. G., Mignosi F., Restivo A.* Fine and Wilf's theorem for three periods and a generalization of Sturmian words. // *Theor. Comput. Sci.* — 1999. — Vol. 218, no. 1. — Pp. 83–94.

[25] *Justin J.* On a paper by Castelli, Mignosi, Restivo. // *Theoret. Inform. Appl.* — 2000. — Vol. 34. — Pp. 373–377.

[26] *Simpson R. J., Tijdeman R.* Multi-dimensional versions of a theorem of Fine and Wilf and a formula of Sylvester. // *Proc. Amer. Math. Soc.* — 2003. — Vol. 131. — Pp. 1661–1671.

[27] *Tijdeman R., Zamboni L.* Fine and Wilf words for any periods. // *Proc. Amer. Math. Soc.* — 2003. — Vol. 131. — Pp. 1661–1671.

[28] *Constantinescu S., Ilie L.* Fine and Wilf's theorem for abelian periods // *Bulletin of EATCS.* — в печати.

[29] *Mignosi F., Shallit J., Wang M.* Variations on a theorem of Fine & Wilf. // MFCS. — 2001. — Pp. 512–523.

[30] *Carpi A., de Luca A.* Semiperiodic words and root-conjugacy // *Theor. Comput. Sci.* — 2003. — Vol. 292, no. 1. — Pp. 111–130.

[31] *Mignosi F., Restivo A., Silva P. V.* On Fine and Wilf's theorem for bidimensional words // *Theor. Comput. Sci.* — 2003. — Vol. 292, no. 1. — Pp. 245–262.

[32] *Restivo A., Silva P. V.* Periodicity vectors for labelled trees // *Discrete Appl. Math.* — 2003. — Vol. 126, no. 2-3. — Pp. 241–260.

[33] A periodicity theorem for trees. / D. Giammarresi, S. Mantaci, F. Mignosi, A. Restivo // IFIP Congress (1). — 1994. — Pp. 473–478.

[34] Periodicities on trees. / D. Giammarresi, S. Mantaci, F. Mignosi, A. Restivo // *Theor. Comput. Sci.* — 1998. — Vol. 205, no. 1-2. — Pp. 145–181.

[35] *Blanchet-Sadri F., Hegstrom R. A.* Partial words and a theorem of Fine and Wilf revisited // *Theor. Comp. Sci.* — 2002. — Vol. 270. — Pp. 401–419.

[36] *Blanchet-Sadri F.* Periodicity on partial words // *Comput. Math. Appl.* — 2004. — Vol. 47. — Pp. 71–82.

[37] *Постников М. М.* Лекции по геометрии. Часть IV : Дифференциальная геометрия. — М., Наука, 1988.

[38] *Shur A. M., Konovalova Yu. V.* On the periods of partial words // *Lect. Notes Comp. Sci.* — 2001. — Vol. 2136. — Pp. 657–665.

[39] *Коновалова Ю. В., Шур А. М.* Периодические частичные слова // Российская конф.“Дискретный анализ и исследование операций”: Тез. докл. — Новосибирск, Институт математики СО РАН, 2002. — P. 130.

[40] *Shur A. M., Gamzova Yu. V.* Periods' interaction property for partial words // Proceedings of WORDS'03, 4th International Conference on Combinatorics on Words. — 2003. — Pp. 75–82.

[41] *Шур А. М., Гамзова Ю. В.* Частичные слова и свойство взаимодействия периодов // *Изв. РАН. Серия матем.* — 2004. — Т. 68, № 2. — С. 199–222.

[42] *Гамзова Ю. В.* Статистические закономерности взаимодействия периодов частичных слов // Российская конф."Дискретный анализ и исследование операций": Тез. докл. — Новосибирск, Институт математики СО РАН, 2004. — С. 82.

[43] *Гамзова Ю. В.* Статистические закономерности взаимодействия периодов частичных слов // *Дискретный анализ и исследование операций. Серия 1.* — 2004. — Т. 11, № 4. — С. 20–35.

[44] *Gamzova Yu. V.* Infinite walks on the set of integers with gaps // Международная алгебраическая конф.: Тез.докл. — Екатеринбург, УрГУ, 2005. — С. 195–196.

[45] *Гамзова Ю. В.* Локально периодические бесконечные частичные слова // *Изв. УрГУ. Серия компьютерные науки и информационные технологии, вып.1.* — 2006. — Т. 43. — С. 5–21.